March 2018

WORKING PAPER SERIES

2018-ACF-03

Searching the Factor Zoo

Soosung Hwang
College of Economics, Sungkyunkwan University, Seoul, Korea

Alexandre Rubesam
IÉSEG School of Management (LEM-CNRS - UMR 9221)

IÉSEG School of Management
Lille Catholic University
3, rue de la Digue
F-59000 Lille
www.ieseg.fr

Searching the Factor Zoo

This version: March 2018

Soosung Hwang[a], Alexandre Rubesam[b]

[a]*College of Economics, Sungkyunkwan University, Seoul, Korea*
[b]*IÉSEG School of Management (LEM-CNRS 9221), Paris, France*

Abstract

Hundreds of factors have been proposed to explain asset returns during the past two decades, a situation which Cochrane (2011) has dubbed "a zoo of new factors". In this paper, we develop a Bayesian approach to explore the space of possible linear factor models in the "zoo". We conduct an extensive search for promising models using a set of 83 candidate factors based on the literature and applying the methodology to thousands of individual stocks. Despite the large number of factors that have been proposed, our results show that (i) only a handful of factors appear to explain the returns on individual stocks; (ii) from these, the only factor that is consistently selected over time is the market excess return; and (iii) other factors which are selected during certain periods are not those in widely used multi-factor models.

Keywords: Factor selection, Bayesian variable selection, Seemingly Unrelated Regressions

Email addresses: shwang@skku.edu (Soosung Hwang), a.rubesam@ieseg.fr (Alexandre Rubesam)

1. Introduction

Which factors should enter a linear factor model, and what kind of fundamental, pervasive, non-diversifiable risks do they represent? This is a crucial question that has haunted researchers for a long time. As more data have become available, and computational costs have decreased, the number of proposed factors to explain asset returns has increased significantly. For example, Harvey *et al.* (2016) document more than three hundred factors that have been proposed in the literature[1], most within the last two decades, a situation which Cochrane (2011) has dubbed "a zoo of new factors".

However, it is doubtful that all of these factors really matter in asset pricing; it is more likely that some of them are redundant, or proxies for the same kind of fundamental risk, whilst many (or even most) may just be a product of data mining[2]. The huge number of factors that have been identified in empirical studies is a challenge for both practitioners and academics, in particular, considering that earlier empirical studies suggested five to six factors[3], and that most prominent and widely used models such as the ones proposed by Fama & French (1992) (henceforth, FF3), Carhart (1997), Pastor & Stambaugh (2003), Chen & Zhang (2010) (henceforth, CZ), Hou *et al.* (2015a) (henceforth, HXZ) and Fama & French (2015) (henceforth, FF5) have five or less factors. Some recent studies, such as Harvey & Liu (2016), Green *et al.* (2017), Yan & Zheng (2017), and Feng *et al.* (2017), have investigated large numbers of factors proposed in the literature in order to identify independent information about average stock returns.

In this study we develop a Bayesian approach to explore the space of possible linear factor models and identify the most promising models to explain asset returns. We propose an estimation method for the posterior probabilities of models, *i.e.* sets of factors, rather than testing individual factors with respect to pre-specified models such as the Fama-French five factor model. With so many candidate factors within the factor "zoo", the number of possible models is astronomical, making model comparison a challenging task. For example, the total number of models is over 1 billion with 30 factors because the number of possi-

[1] They also provide a taxonomy of these factors, refer to their Table 1. Also see Green *et al.* (2017) and McLean & Pontiff (2016), which summarize hundreds of factors proposed in the literature.

[2] See Chordia *et al.* (2017) and Kewei *et al.* (2017).

[3] Roll & Ross (1980), Chen *et al.* (1986), Connor & Korajczyk (1988), and Lehmann & Modest (1988).

ble models with K factors is 2^K. We develop a novel method that evaluates the space of all possible models, which would be computationally prohibitive in the conventional framework even for moderate number of factors.

For this purpose, we introduce a Bayesian variable selection method that explores the model space, *i.e.* posterior probabilities of the most promising models. For the K candidate predictor variables, a vector of the joint posterior distribution of $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)'$ of dummy variable γ_j can be defined to represent whether the $j - th$ predictor should be included in the model. A non-hierarchical model is employed, where the prior distribution of the regression coefficients (the factor sensitivities or βs in a multi-factor asset pricing model) is independent from that of γ_js. The introduction of $\boldsymbol{\gamma}$ makes it possible to explore the space of possible models, even with large numbers of factors. Our contribution in terms of methodology can be summarized as follows. First, we propose a simple approach by specifying independent priors for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, extending the univariate regression model proposed by Kuo & Mallick (1998) to the multivariate seemingly unrelated regressions (SUR) model. Second, we derive a sequential algorithm to estimate the regression coefficients (factor sensitivities) of each response variable (each asset) using the Gibbs sampler[4]. This provides an efficient way to estimate the model even for larger numbers of test assets and factors, which allows the application of the method to sets of large numbers of individual stocks, instead of portfolios[5].

Our Bayesian approach overcomes the multiple comparison problem raised by Harvey *et al.* (2016) and others. When a large number of signals are tested to investigate cross-sectional asset returns, some signals will appear to be statistically significant by random chance even if they have no genuine predictive ability. Our procedure allows us to simultaneously assess the most promising models within the space of all possible models, instead of statistical inference based on a "single" test, and therefore, all individual signals are evaluated together as (argued) in Sullivan *et al.* (1999, 2001). The Bayesian framework differs from the frequentist perspective of Harvey *et al.* (2016) who propose a t-statistic greater than 3 for any new factor. Our approach can be applied to thousands of individual assets together with hundreds of potential factors, and thus does not need to reduce dimensions of test assets by forming portfolios (Lo & Mackinlay, 1990; Ferson

[4]See Kim & Nelson (1999) for a review of Gibbs sampling estimation in Econometric models.

[5]In terms of methodology, our approach is mostly related to the literature on variable selection in multivariate regression models, of which the SUR model is a special case, see Brown *et al.* (1998), Smith & Kohn (2000), Hall *et al.* (2002), Wang (2010), Ando (2011), Ouysse & Kohn (2010), and Puelz *et al.* (2017).

et al. , 1999; Berk, 2000). The complex cross-sectional dependencies can be considered in this framework as all possible combinations of factors are evaluated. Multi-collinearity problems that become serious when the number of independent variables increase, can be avoided because our approach is to select the best possible models from the posterior probability of $\boldsymbol{\gamma}$ [6].

In the empirical tests, we consider a set of 83 candidate asset pricing factors that have been proposed in the literature. In addition to the market factor (excess market return), we compute 82 tradable factors by sorting stocks into value-weighted decile portfolios based on various firm characteristics and variables that have been proposed in the literature, and calculating the return difference between the top and bottom decile portfolios. We apply our Bayesian variable selection methodology to all available stocks in different sub-sample periods from 1980 to 2016. We also consider 20 different sets of portfolios of stocks, comprising over 300 individual portfolios.

Our empirical results with individual stocks suggest that only a small number of factors (5 to 6) are important to explain the individual stock returns. The only factor that is consistently selected over time is the market excess return. Moreover, the other factors that are selected in this study are not those that have been widely used in the literature, i.e, factors such as those in the FF3, FF5, CZ and HXZ models, but include short-term reversal, change in 6-month momentum, change in number of analysts following stocks, and industry concentration. These results are robust to different specifications of the priors about the factor sensitivities.

In comparison with some recent studies such as Green *et al.* (2017) and Barillas & Shanken (2017), our results show a smaller number of relevant factors. For example, Green *et al.* (2017) use a set of 94 firm characteristics in Fama-MacBeth regressions and show that 12 characteristics are important to explain returns on stocks over the period 1985-2014[7]. Barillas & Shanken (2017) find evidence supporting a six-factor model including the the market return, investment,

[6] Although Green *et al.* (2017) and Feng *et al.* (2017) evaluate the effects of the multi-collinearity problem carefully, this problem does not disappear in the conventional regression with the large number of independent variables that are possibly cross-correlated.

[7] The 12 characteristics identified in the study are book-to-market, cash, change in the number of analysts, earnings announcement return, one-month momentum, change in six-month momentum, number of consecutive quarters with earnings higher than the same quarter a year ago, annual R&D to market cap, return volatility, share turnover, volatility of share turnover, and zero trading days. The authors also find that this number reduces to only 2 (industry-adjusted change in employees and number of earnings increases) since 2003, with the returns to hedge portfolios that attempt to exploit this predictability becoming insignificant.

profitability, size, book-to-market, and momentum factor. Other recent studies with a large number of candidate factors, such as Harvey *et al.* (2016) and Feng *et al.* (2017), have also found that the market factor is the most important one, with a possible role for profitability and investment.

The main difference from these results lies in how these factors are selected. The factors in our study are selected from a larger set of factors, using an approach that considers all factors simultaneously. Therefore, our results do not suffer from the multiple comparison problem (Harvey *et al.* , 2016). On the other hand, most other studies in the literature identify additional factors relative to arbitrary sets of factors, or even without considering other possible factors. When the entire set of factors is searched for the best model, a set of a smaller number of factors is required.

One interesting result is that the intercept term is not selected to explain returns on individual stocks, indicating that the models with high posterior probabilities explain well the returns of individual stocks. Therefore, our method can be an alternative to the traditional asset pricing tests such as the Fama & MacBeth (1973) procedure or the Gibbons *et al.* (1989) test. These tests suffer from problems that reduce their power, and require grouping of individual stocks into portfolios, which introduces biases if the variable used to sort stocks into portfolios is related to the factors in the model. In contrast, our procedure considers many factors simultaneously and can naturally be applied when the number of assets is larger than the number of time-series observations.

Our work also differs markedly from previous studies that apply a Bayesian approach to select asset pricing factors, such as Ericsson & Karlsson (2003), Ouysse & Kohn (2010), Puelz *et al.* (2017) and Barillas & Shanken (2017). These studies have focused on a smaller number of candidate factors, with a relatively small number of portfolios as test assets. Although the Bayesian approach proposed by Barillas & Shanken (2017) is designed to test individual asset pricing models, the number of candidate factors is limited due to its computational costs. In contrast, our methodology allows us to explore a larger model space with many possible factors, using thousands of individual stocks simultaneously, therefore bypassing the problem of using as test assets portfolios that may be related to the factors by construction. In fact, when we apply our methodology to 20 different sets of portfolios (300 portfolios in total), we find a very strong dependence between the portfolio formation criteria and the posterior probability of factors. For example, when portfolios are formed on firm characteristics, models with the factors formed on these characteristics are selected with high posterior probability[8].

The rest of this paper is organized as follows. We introduce the model and briefly discuss the estimation method in Section 2. The explanation of the data set and factor construction follow in Section 3. Section 4 provides the main empirical results of the paper, as well as robustness tests and comparison with previous studies. Section 5 concludes. The Bayesian estimation of the SUR model is reviewed in Appendix A. Appendix B contains detailed explanations of the variable selection model and its estimation. Appendix C provides the full list of firm characteristics used in this study, and the associated references.

2. Methodology

Consider N assets and K predictor variables (factors) over T periods. The factor model is a multivariate linear regression with N equations:

$$\mathbf{r}_i = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{e}_i, \qquad i = 1, \ldots, N \tag{1}$$

where, for each asset i, $\mathbf{r}_i$ is the $T \times 1$ vector of excess returns, $\mathbf{X}$ is the matrix of factors with dimension $T \times K$, $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,K})'$ is a vector of unknown regression coefficients (factor sensitivities), and $\mathbf{e}_i$ is a $T \times 1$ vector of disturbances[9]. If the error terms are contemporaneously cross-correlated, the system of regressions above is a special case of the Seemingly Unrelated Regressions (SUR) model, where the predictor variables are the same for all equations[10].

The system can be stacked in a single equation $\tilde{\mathbf{r}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}}$ in the following

[8]This is related to the concerns expressed by Lo & Mackinlay (1990), Ferson *et al.* (1999), Berk (2000), Roll (1977) and Lewellen *et al.* (2010) in the context of bias in asset pricing tests using portfolios related to the factors. A similar conclusion is reached by Harvey & Liu (2016). They argue that dispersion in portfolios is largely driven by a few portfolios that are dominated by small stocks, which leads asset pricing tests to identify factors that can explain these extreme portfolios.

[9]To avoid ambiguity, throughout this article we use the subscripts i and j for assets and predictor variables, respectively.

[10]The SUR model, introduced by Zellner (1962), consists of N regression equations, each with T observations, which are linked solely through the covariance structure of error terms at each observation, *i.e.* errors are contemporaneously correlated but not autocorrelated.

way:

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_N \end{bmatrix}, \tag{2}$$

where $\tilde{\mathbf{e}} = (\mathbf{e}_1' \quad \mathbf{e}_2' \quad \ldots \quad \mathbf{e}_N')'$, and $\mathbb{E}(\tilde{\mathbf{e}}\tilde{\mathbf{e}}') = \mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I_T}$.

Bayesian inference in the SUR model can be carried out in a relatively straightforward manner, see for example Giles (2003). Since our variable selection procedure will rely on a Markov Chain Monte Carlo (MCMC) approach using the Gibbs sampler, we start by reviewing the estimation of the SUR through this approach. Suppose $\tilde{\mathbf{e}} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I_T})$ and the following prior distributions for $\tilde{\boldsymbol{\beta}}$ and $\mathbf{\Sigma}$:

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &\sim N(\mathbf{b}_0, \mathbf{B}_0) \\ \mathbf{\Sigma} &\sim IW(\nu_0, \mathbf{\Phi}_0), \end{aligned} \tag{3}$$

where $IW(\nu_0, \mathbf{\Phi}_0)$ denotes the inverted-Wishart distribution with ν_0 degrees of freedom and parameter matrix $\mathbf{\Phi}_0$. With these choices, it can be shown that the conditional posterior distributions required for the Gibbs sampler are as follows[11]:

$$\begin{aligned} \tilde{\boldsymbol{\beta}}|\mathbf{\Sigma}, \mathbf{r} &\sim N(\mathbf{b}_1, \mathbf{B}_1) \\ \mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \mathbf{r} &\sim IW(\nu_1, \mathbf{\Phi}_1), \end{aligned} \tag{4}$$

where

$$\begin{aligned} \mathbf{b}_1 &= (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}\tilde{\mathbf{X}})^{-1}(\mathbf{B}_0\mathbf{b}_0 + \tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}\tilde{\mathbf{r}}) \\ \mathbf{B}_1 &= (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}'\mathbf{\Omega}^{-1}\tilde{\mathbf{X}})^{-1} \\ \nu_1 &= \nu_0 + T, \quad \mathbf{\Phi}_1 = \mathbf{\Phi}_0 + \mathbf{S}. \end{aligned}$$

In the above, $\mathbf{S}$ is the matrix of cross-products of the residuals, that is, if $\mathbf{E} = [\mathbf{e}_1 \ldots \mathbf{e}_N]$, then $\mathbf{S} = \mathbf{E}'\mathbf{E}$. We also note that $\mathbf{\Omega}^{-1} = \mathbf{\Sigma}^{-1} \otimes \mathbf{I_T}$.

The approach above may be computationally intensive if the number of equations is large, since it requires multiplication and inversion of large matrices. For example, $\tilde{\mathbf{X}}$ has dimension $NT \times NK$ and $\mathbf{\Omega}^{-1}$ has dimension $NT \times NT$. By

[11]The full derivation of all the conditional distributions required for the Gibbs sampler estimation is provided in Appendix A.

sampling each $\boldsymbol{\beta}_i$ conditionally on the remaining $\boldsymbol{\beta}_j$, $j \neq i$ and $\boldsymbol{\Sigma}$, we derive an alternative and quicker approach for a large panel. Let $\tilde{\boldsymbol{\beta}}_{-i}$ denote the full vector $\tilde{\boldsymbol{\beta}}$ omitting $\boldsymbol{\beta}_i$. Assume that

$$\boldsymbol{\beta}_i | \tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\Sigma} \sim N(\mathbf{b}_{0,i}, \mathbf{B}_{0,i}).$$

Then, $\boldsymbol{\beta}_i | \tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\Sigma}, \mathbf{r} \sim N(\mathbf{b}_{1,i}, \mathbf{B}_{1,i})$, with

$$\begin{aligned}\mathbf{b}_{1,i} &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_{0,i}\mathbf{b}_{0,i} + \sigma^{ii}\mathbf{X}'\mathbf{r}_i^*) \\ \mathbf{B}_1 &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}'\mathbf{X})^{-1},\end{aligned}$$

where σ^{ii} denotes the (i,i) element of $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{r}_i^*$ is suitably defined based on a partition of the systems of equations, see Appendix A.2. Note that the expressions above depend only on the smaller matrices $\mathbf{X}$ and $\boldsymbol{\Sigma}$. In the Gibbs sampler, each $\boldsymbol{\beta}_i$ can be generated in random order.

2.1. Bayesian Variable Selection in the SUR Model

There is a vast literature focusing on Bayesian variable selection in linear models with a single response variable, see for example George & McCulloch (1993, 1997); Kuo & Mallick (1998); Dellaportas *et al.* (1999); Hans *et al.* (2007); Clyde & George (2004); O'Hara & Sillanpää (2009). For a single regression equation, Bayesian variable selection is typically done by first introducing a vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$ of dummy variables, where if $\gamma_j = 1$, the $j-th$ predictor is included in the model, and conducting inference on the posterior distribution of $\boldsymbol{\gamma}$. Since the vector of K dummy variables indicates 2^K possible models, comparison of all possible models becomes computationally infeasible for even moderate numbers of regressors. In this case, MCMC methods provide a fast way to obtain consistent estimates of model probabilities.

Variable selection in the multivariate regression models (of which the SUR model is a special case) has been the subject of a number of studies, mostly focusing on generalizations of the hierarchical Bayesian model of George & McCulloch (1993)[12], which defines the distribution of $\boldsymbol{\beta}$ conditionally on $\boldsymbol{\gamma}$. This is done by specifying a "slab and spike" mixture distribution which places a spiked prior on

[12]One of the first examples of this approach is Brown *et al.* (1998). Smith & Kohn (2000) introduced a Bayesian hierarchical model which considers variable selection by explicitly allowing the possibility that some coefficients are equal to zero. Hall *et al.* (2002) consider a hierarchical Bayesian model related to Smith & Kohn (2000) to choose style factors in models for global stock returns. Wang (2010) also follows the hierarchical setup of George & McCulloch (1993),

zero for $\beta_j|\gamma_j = 0$ and a slab or flat prior on $\beta_j|\gamma_j = 1$. One disadvantage of this approach is that it often requires data-dependent tuning of the hyper-parameters. In this study, we assume *a priori* independence between β_j and γ_j which requires no tuning and extend the univariate regression model proposed by Kuo & Mallick (1998) to the case of the SUR model with common regressors.

We generalize the method proposed by Kuo & Mallick (1998) to the SUR model as follows. Let $\mathbf{X}_\gamma$ represent the matrix $\mathbf{X}$ where each column has been multiplied by the corresponding γ_j. Then we can write the model with variable selection as $\mathbf{r}_i = \mathbf{X}_\gamma \boldsymbol{\beta}_i + \mathbf{e}_i, i = 1, \ldots, N$, or stacking the N equations as before,

$$\tilde{\mathbf{r}} = \tilde{\mathbf{X}}_\gamma \tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}},$$

where $\tilde{\mathbf{X}}_\gamma$ is defined analogously as before. Equivalently, we define a new variable $\boldsymbol{\theta}_i = \boldsymbol{\beta}_i \odot \boldsymbol{\gamma}$, where $\odot$ represents element-wise multiplication. The system can then be represented by $\tilde{\mathbf{r}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\theta}} + \tilde{\mathbf{e}}$. Analysis of the posterior distribution of $\tilde{\boldsymbol{\theta}}$ would be useful to understand which variables are important for each equation.

To derive the conditional distributions required for the Gibbs sampler, we need to specify the prior distribution for $\boldsymbol{\gamma}$. We follow Kuo & Mallick (1998) and set independent priors as $\gamma_j \sim B(1, \pi_j), j = 1, \ldots, K$. Therefore, the prior distribution of $\boldsymbol{\gamma}$ is given by

$$f(\boldsymbol{\gamma}) = \prod_{j=1}^{K} \pi_j^{\gamma_j}(1 - \pi_j)^{1-\gamma_j}.$$

Note that, conditional on a known value of $\boldsymbol{\gamma}$, the model reduces to a SUR with the corresponding predictors for which $\gamma_j = 1$. Therefore, using the same prior distributions for $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$ in Equation (3), the conditional distributions for $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$ are those given in equation (4), with $\tilde{\mathbf{X}}$ replaced by $\tilde{\mathbf{X}}_\gamma$. Thus we have

$$\begin{aligned} \tilde{\boldsymbol{\beta}}|\boldsymbol{\gamma}, \boldsymbol{\Sigma}, \mathbf{r} &\sim N(\mathbf{b}_1, \mathbf{B}_1) \\ \boldsymbol{\Sigma}|\boldsymbol{\gamma}, \tilde{\boldsymbol{\beta}}, \mathbf{r} &\sim IW(\nu_1, \boldsymbol{\Phi}_1), \end{aligned} \tag{5}$$

considering structured covariance matrices within the context of normal graphical models. Ando (2011) proposes a Direct Monte Carlo method estimation for a hierarchical model related to Smith & Kohn (2000). Ouysse & Kohn (2010) apply a model related to Brown *et al.* (1998) to simultaneously make inferences on asset pricing factors and estimate factor risk premia. Puelz *et al.* (2017) consider the case of treating the variables as random, and propose strategies for model summarization.

where

$$\begin{aligned}\mathbf{b}_1 &= (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}_\gamma'\mathbf{\Omega}^{-1}\tilde{\mathbf{X}}_\gamma)^{-1}(\mathbf{B}_0\mathbf{b}_0 + \tilde{\mathbf{X}}_\gamma'\mathbf{\Omega}^{-1}\tilde{\mathbf{r}})\\ \mathbf{B}_1 &= (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}_\gamma'\mathbf{\Omega}^{-1}\tilde{\mathbf{X}}_\gamma)^{-1}\\ \nu_1 &= \nu_0 + T, \quad \mathbf{\Phi}_1 = \mathbf{\Phi}_0 + \mathbf{S}.\end{aligned}$$

As before, if the number of equations is large, we can sample each $\boldsymbol{\beta}_i, i = 1, \ldots, N$ in turn from $\boldsymbol{\beta}_i|\tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\gamma}, \mathbf{\Sigma}, \tilde{\mathbf{r}} \sim N(\mathbf{b}_{1,i}, \mathbf{B}_{1,i})$, where

$$\begin{aligned}\mathbf{b}_{1,i} &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}(\mathbf{B}_{0,i}\mathbf{b}_{0,i} + \sigma^{ii}\mathbf{X}_\gamma'\mathbf{r}_i^*)\\ \mathbf{B}_1 &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}.\end{aligned}$$

To generate $\boldsymbol{\gamma}$, we use the Gibbs sampler to generate each value of $\boldsymbol{\gamma}$ as in Kuo & Mallick (1998). The relevant conditional posterior probability of $\gamma_j = 1$ for the SUR model is given by

$$P(\gamma_j = 1|\boldsymbol{\gamma}_{-j}, \tilde{\boldsymbol{\beta}}, \mathbf{\Sigma}, \tilde{\mathbf{r}}) = \left(1 + \frac{1-\pi_j}{\pi_j}\exp(-0.5\,\mathrm{Tr}(\mathbf{\Sigma}^{-1}(\mathbf{S}_\gamma^1 - \mathbf{S}_\gamma^0))\right)^{-1}, \qquad (6)$$

where $\mathbf{S}_\gamma^1$ and $\mathbf{S}_\gamma^0$ represent the matrices of residuals when $\gamma_j = 1$ and $\gamma_j = 0$, respectively. Each γ_j can be generated, preferably in random order, using the expression above[13].

2.2. Prior Distributions

The most important prior distribution is the one for $\tilde{\boldsymbol{\beta}}$. As discussed by O'Hara & Sillanpää (2009), the MCMC algorithm might not mix well in the $\boldsymbol{\gamma}$ space if the prior for $\tilde{\boldsymbol{\beta}}$ is too vague. The reason for this is that, when a particular $\gamma_j = 0$, the $\beta_{ij}, i = 1, \ldots, N$ are sampled from the full prior conditional distribution. In this case, it may be difficult for the model to transition between $\gamma_j = 0$ and $\gamma_j = 1$, since the generated β_{ij} will be unlikely to be in the region where θ_{ij} has higher posterior probability.

We propose a few choices for the priors on $\tilde{\boldsymbol{\beta}}$. The first is $\tilde{\boldsymbol{\beta}} \sim N(\mathbf{0}, c\mathbf{I})$. This choice reflects a complete lack of knowledge about the predictors, both in terms of which predictors should enter the model as well as regarding the dependence structure of the regression coefficients. A second possibility is $\tilde{\boldsymbol{\beta}} \sim$

[13] An alternative approach is to apply a Metropolis-within-Gibbs step of the type suggested by Brown *et al.* (2002), see also George & McCulloch (1997) .

$N(\mathbf{0}, c(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1})$, which makes the prior covariance structure equal to the design covariance structure, as suggested by Zellner (1971). A final possibility is to center each $\boldsymbol{\beta}_i$ around their OLS or maximum likelihood estimate, *i.e.* $\boldsymbol{\beta}_i \sim N((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{r}_i, c_i(\mathbf{X}'\mathbf{X})^{-1})$. All of these choices can be made less informative by increasing c. Note that the first component of each β_i is for the alpha of each regression. The intercept is included as a factor because there is no guarantee that the factors we test in this study can fully explain individual stock returns.

The standard choice for the prior for $\boldsymbol{\Sigma}$ is to set $\nu_0 = N$ and $\boldsymbol{\Phi}_0 = \mathbf{I}$. Another possibility is to choose the parameters so that the prior variance will be equal to a given number, which may come from our knowledge of the problem. For the prior of π_j, the prior probability that predictor j is included in the model, we choose an equal probability of $\frac{1}{2}$ for all factors. This prior reflects the lack of knowledge about the inclusion of the predictors, and implies that any model, regardless of its possible number of combinations, has an equal prior probability of $\frac{1}{2^K}$. Prior information regarding predictors that researchers include in the model can be incorporated by letting $\pi_j = 1$. For example, if we want to reflect a prior belief that the market factor should always be included, we can set the corresponding π_j equal to 1.

2.3. *Comparison with other variable selection models*

The method we propose has two main differences compared to other approaches. The first one is that we do not follow the hierarchical structure as in Brown *et al.* (1998); Smith & Kohn (2000); Ouysse & Kohn (2010); Wang (2010); Ando (2011); Puelz *et al.* (2017). Our non-hierarchical structure results in a simple method for variable selection in SUR models, with the advantage that it does not require complex tuning of the hyper-parameters.

It is possible to make inference about which variables matter for each asset (equation) by summarizing the posterior distribution of $\boldsymbol{\theta}_i = \boldsymbol{\beta}_i \odot \boldsymbol{\gamma}$. Thus, by focusing on finding a single set of predictors for the N equations, we identify common pricing factors in the multi-factor models. Other studies such as Wang (2010) and Puelz *et al.* (2017) propose methods that can identify different sets of predictors for each equation.

3. Data and Factor Construction

3.1. *Factor Returns and Their Statistical Properties*

For the candidate factors, we use 82 firm characteristics that have been tested by Green *et al.* (2017) for the sample period from January 1980 to December

2016, a total of 37 years (444 months)[14]. Factor returns are calculated by the difference between the value-weighted returns on the highest and lowest decile portfolios[15]. We use all available U.S. common stocks from the CRSP and Compustat databases for the calculation of factor returns. As in Green *et al.* (2017), characteristics are updated on a monthly basis using the available accounting information[16]. In addition to these 82 factors, we also consider the excess market return[17]. Since we also consider the intercept as a factor for the purposes of our Bayesian variable selection procedure, the total number of factors is 84. Due to differences in data availability, different factors are available for different subsample periods.

Table 1 reports basic descriptive statistics for the factors used in this study. For each factor, we calculate and report statistics using all the available stocks. We also report Dependent False Discovery Rate (DFDR) p-values using the method of Benjamini & Yekutieli (2001), which takes into account the fact that multiple tests are being run simultaneously. The factors with a DFDR p-value less than 0.05 are shown in bold, and the corresponding t-statistic includes an asterisk.

The average returns on the factors based on well-known characteristics such as mve (market cap), bm (book-to-market ratio) and mom12m (12-month momentum) are in line with numbers reported on the literature. Despite the differences in the factor return calculation and the sample period, the average factor returns are similar to those of Green *et al.* (2017). It is noteworthy that only 6 of the 83 factors are significant when the DFDR p-values are considered, despite the fact that 27 factors have t-statistics higher than 2.0 in absolute value, reflecting the much higher burden of significance when multiple testing is taken into account. These factors are acc (working capital accruals), chcsho (change in shares outstanding), chinv (change in inventory), invest (capital expenditures and inventory), nanalyst

[14]We thank Jeremiah Green for making his SAS code available online. The firm characteristics that have too many missing values or whose deciles are not meaningful are excluded. The excluded characteristics are convind (convertible debt indicator), divi (dividend initiation), divo (dividend omission), dy (dividend yield), ipo (new equity issue), nincr (number of earnings increases), rd (R&D increase), rd_mve (R&D to market capitalization), rd_sale, secured (secured debt), securedind (secured debt indicator) and sin (Sin stocks). The exclusion of these 12 firm characteristics, however, does not affect our main results because the intercept is not selected for the explanation of individual stocks.

[15]We apply the same procedure for all characteristics, and thus average return difference may be positive (*e.g.* book-to-market ratio) or negative (*e.g.* market value of equity).

[16]The details are described in pg. 4398 of Green *et al.* (2017).

[17]The excess market return is taken from Kenneth French's data library.

(number of analysts covering stock), and sfe (scaled earnings forecast). On the other hand, some characteristics which are significant in their univariate regressions are not significant in multiple testing, although most have large t-statistics. This is the case of agr (asset growth, t-stat = -2.9), chatoia (industry-adjusted change in turnover, t-stat = 2.85), ear (earnings announcement return, t-stat = 2.93), egr (growth in common shareholder equity, t-stat = -2.84), grcapx (growth in capital expenditures, t-stat -2.70), grltnoa (growth in long term net operating assets, -3.20), pchsalepchnvt (change in sales - change in inventory, t-stat = 1.58), and sue (unexpected quarterly earnings, t-stat = 2.57).

[Table 1 about here.]

Table 2 reports statistics of the absolute pairwise correlations between the factors for the period during which all 83 factors (except the intercept) are available, a total of 114 months from July 2007 to December 2016. There are 3403 total pairwise correlations. The median absolute correlation is 0.179, and 90% of all absolute correlations are below 0.498. We also report the 10 largest absolute correlations. Some factors are highly correlated, and 4 correlations are higher than 0.90. Figure 1 plots the distribution of the absolute correlations. When these factors are used all together in the conventional regression equation, multi-collinearity problems would arise despite a weak cross-correlations between individual firm characteristics as in Green *et al.* (2017).

[Table 2 about here.]

[Figure 1 about here.]

3.2. Test Assets

The Bayesian variable selection method is applied to thousands of common stocks listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and NASDAQ. We exclude financial stocks (Standard Industrial Classification code from 6000 to 6999) because their accounting practices and variables are not compatible with those of the other sectors. We also exclude a large number of microcap stocks whose market size is less than the bottom 20th percentile of market cap of NYSE stocks as well as penny stocks whose prices are less than US$1 at the beginning of test periods. When these stocks are included, our results would be affected too much by the market microstructure biases and thin trading by these stocks whose value is less than 3 percent of the total market value. For robustness purposes, we later investigate factor selection using micro-cap stocks only.

When considering individual stocks over long sample period, there is a serious survivorship bias. Therefore, we consider shorter sub-sample periods to minimize survivorship bias and also capture time variation in factor selection. When deciding the length of the sub-sample periods, we need to consider a trade-off between precision (using more data to conduct inference on factor selection) and the potential for survivorship bias. We also face a natural limit given the large number of candidate factors *i.e.* the number of months in each sub-sample period should be larger than that of factors.

Therefore, two different approaches have been used to balance these concerns. First, we divide our sample period into three sub-sample periods of 144 (January 1980 to December 1991), 144 (January 1992 to December 2003) and 156 (January 2004 to December 2016) months and apply the Bayesian variable selection method using all available factors in each sub-sample period[18]. This approach allows us to study a larger set of candidate factors, with some reduction in survivorship bias. The second approach consists of 5 shorter sub-sample periods, with the first 4 containing 90 months each, and the last containing 84 months. For these shorter sub-sample periods (not larger than 90 monthly observations), we restrict the number of candidate factors into 55 including the market excess return and the intercept. These are the factors that are significant in any of the regressions in Green *et al.* (2017).

To compare our results with those of previous studies and to assess the performance of our method, we also consider an extensive set of portfolios formed by sorting stocks according to different criteria[19]. The portfolio return data are obtained from Kenneth French's data library[20]. We consider portfolios formed on univariate and bivariate sorts, as well as industry classification. The total number

[18]These two break points are chosen considering the importance of research on firm characteristics (e.g., Fama & French (1992) and Jegadeesh & Titman (1993)) and the structural breaks in the performance of firm characteristics in cross-sectional asset returns (Green *et al.* (2017)).

[19]Testing portfolios is motivated by the dependence we observe between the portfolio formation criteria and the selected factors. Lewellen, Nagel and Shanken (2007) argue that it is problematic to use portfolios formed on firm characteristics to test asset pricing models, because these portfolios will have a tight factor structure by construction. In this case, the idiosyncratic component of the model will be very small, and the factors will appear to be statistically significant cross-sectionally.

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

of portfolios considered is over 300[21].

4. Selection of Asset Pricing Factors

We first discuss our main empirical findings for individual stocks,followed by the results of portfolios. These results have been obtained using an empirical Bayes prior for the factor sensitivities, *i.e.* we center each $\boldsymbol{\beta}_i$ around their OLS estimate by setting $\boldsymbol{\beta}_i \sim N((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{r}_i, c_i(\mathbf{X}'\mathbf{X})^{-1})$, with $c_i = 1$. We consider an equal prior probability for each factor: $\pi_j = \pi = 0.5$. The results for individual stocks are based on 10,000 iterations of the MCMC algorithm, while those for portfolios are based on 50,000 iterations. At the end of this section we test the robustness of our results with respect to these choices.

4.1. Individual Stocks as Test Assets

We apply our Bayesian variable selection methodology to individual stocks for various sub-sample periods using all available factors in the sub-sample periods in Section 3.2. These results are free from the biases inherent in using portfolios formed on characteristics which may be related to the factors we study, as discussed by Lo & Mackinlay (1990), Ferson *et al.* (1999), Berk (2000), Lewellen *et al.* (2010) and others.

4.1.1. Results using three sub-sample periods and the full set of factors

The results for the three sub-sample periods are reported in Table 3: January 1980 to December 1991, January 1992 to December 2003, and January 2004 to December 2016. The total number of non-microcap stocks in each period are 807, 893 and 967, respectively, while the number of candidate factors are 75, 81 and 83, respectively.

[Table 3 about here.]

For all sub-sample periods, models with less than 5 factors are generally selected by our Bayesian variable selection procedure. This is quite surprising, as

[21] The univariate sort portfolios considered are those formed on size, book-to-market, operating profitability, investment, earnings-to-price, cashflow-to-price, dividend yield, momentum, short-term reversal, long-term reversal, beta, variance and residual variance. The bivariate sort portfolios include size and book-to-market, size and operating profitability, size and investing, book-to-market and operating profitability, book-to-market and investment, and operating profitability and investment. The industry portfolios comprise 49 industries.

the number of possible models is enormous, varying from 2^{75} in the first sub-sample period to 2^{81} in the last sub-sample period. However, we do not find that a single model (a set of factors) or factors (other than the market factor) are consistently selected across sub-sample periods. Additionally, among the large number of candidate factors, only 13 factors are ever selected over the three sub-sample periods. These factors are mkt (the market return), aeavol (abnormal earnings announcement volume), chmom (change in 6-month momentum), chanalyst (change in number of analysts covering stock), ear (earnings announcement return), ep (earnings-to-price), herf (industry sales concentration), mom1m (1-month momentum), ms (Mohanram (2005a)'s financial statement score), pctacc (percent accruals), saleinv (sales to inventory), tb (tax income to book income), and intercept. When factors whose marginal posterior probabilities are above 0.5 are counted in any of the sub-sample period, we have only four factors, i.e, mkt, chmom, herf, and mom1m.

During the first sub-sample period, from January 1980 to December 1991, there is a substantial amount of model uncertainty, as the posterior probability of the best model is quite low. The best model includes the market factor and chmom (change in 6-month momentum), with a posterior probability of 0.24. Other models include either mom1m (1-month momentum) and/or ms (Mohanram (2005a)'s financial statement score). Lower probability models include other factors formed on ep (earnings-to-price) or tb (tax income to book income). The only factors that have marginal posterior probabilities higher than 0.5 (our prior) during this period are the market factor and chmom (change in 6-month momentum).

In the second sub-sample period, which comprises the period from January 1992 to December 2003, model uncertainty is much lower, with the best model including only the market factor with a high posterior probability of 0.64. The second best model includes aeavol (abnormal earnings announcement volume) or pctacc (percent accruals), the latter with very low posterior probability.

Finally, in the last sub-sample period, from January 2004 to December 2016, two models appear with similar posterior probabilities. The best model, with posterior probability 0.24, includes the intercept and 3 factors. In addition to the market return, this model includes herf (industry sales concentration) and mom1m (1-month momentum). The second best model, with posterior probability 0.20, drops the intercept and includes chanalyst (change in number of analysts covering stock). We note that many models include the intercept, which we interpret as evidence that these models do not explain the individual stock returns well during this period.

These results suggest some interesting points. First, factor selection varies quite a lot through time, with no specific model dominating the others. The only factor for which appears consistently across all sub-sample periods is the excess market return. The other factor that appears in more than one sub-sample period is mom1m (1-month momentum), which is included in the best models in the first and last sub-sample periods. All other factors matter only during one sub-sample period. Second, the high posterior probability models do not include the popular factors (other than the market return) that have been proposed in the literature, e.g., FF3, FF5, CZ or HXZ. The additional factors that are selected by our Bayesian variable selection method are related to other anomalies or characteristics such as short-term reversal, momentum, earnings announcement volume, change in the number of analysts covering stocks, and industry concentration. Third, the total number of factors selected in these models over all sub-sample periods is small relative to the total number of candidate factors. Only 13 factors (out of more than 80) are ever selected by the variable selection methodology, and from these, only 4 factors have marginal posterior probabilities higher than 0.5 in any of the sub-sample periods.

4.1.2. Results using five sub-sample periods and the reduced set of factors

The selection of the intercept during the last sub-sample period suggests that the 13 models (8 factors) do not fully explain individual stock returns. Moreover, the large numbers of models in the first and third sub-sample periods indicate model uncertainty during these periods. In this subsection we explore the possibility that these results arise because of the relatively long sub-sample periods, *i.e.* 12 years. When the true model is time-varying, a linear regression model over a long period requires a larger number of factors or even does not explain asset returns (Jagannathan and Wang, 1996).

Five shorter sub-sample periods are used to test individual stocks: January 1980 to June 1987, July 1987 to December 1994, January 1995 to June 2002, July 2002 to December 2009, and January 2010 to December 2016. The number of stocks varies from 1,014 in the first sub-sample period to 1,225 in the last sub-sample period. By selecting factors that are significant in any of the regressions in Green *et al.* (2017), the numbers of factors we test for these five sub-sample periods range from 44 to 49. The results are reported in Table 4.

[Table 4 about here.]

The results show similarities with those of 3 longer sub-sample periods and the full set of candidate factors in Table 3. No single model or factor (other than the market factor) is consistently selected across sub-sample periods except for the excess market return. However, a few important differences from the shorter sub-sample periods can be summarized as follows. First, model uncertainty decreases with test period. At most 4 factors are selected rather than 7 factors in 3, and only 10 factors (out of almost 50) are selected by the variable selection methodology. These are: mkt (market excess return), aeavol (abnormal earnings announcement volume), bm (book-to-market), chmom (change in 6-month momentum), ear (earnings announcement return), mom1m (1-month momentum), mve_ia (industry adjusted size), pchsale_pchrect (change in sales - change in A/R), pctacc (percent accruals), and sue (unexpected quarterly earnings). From these, only 6 have marginal posterior probabilities higher than 0.5 (mkt, chmom, ear, mom1m, mve_ia, and sue). Second, despite the smaller numbers of factors, the intercept is not selected in any of the sub-sample periods, which can be interpreted that these 10 factors are enough to explain individual stocks. These results suggest that the possibility of model mis-specification decreases when sample period becomes shorter.

Starting in Panel A, a single model is selected during the first sub-sample period; the model includes the mkt (market excess return) and chmom (change in 6-month momentum) factors. Panel B shows that, during the period from July 1987 to December 1994, ear (earnings announcement return) and mom1m (1-month momentum) factors need to be added on top of these two factors. Model uncertainty is still very low, with only two models having relevant posterior probabilities. The next three sub-sample periods show higher model uncertainty, and different factors are included in the best models. In the period from January 1995 to June 2002 (Panel C), the best model includes only the market return, with a posterior probability of 0.44. The next best models require either pctacc (percent accruals), mom1m (1-month momentum), or aevol (abnormal earnings announcement volume), although with much lower posterior probabilities. During the period from July 2002 to December 2009 (Panel D), the best model includes the market return and sue (unexpected quarterly earnings), with a posterior probability of 0.44. The next best models include bm (book-to-market) or pchsale_pchrect (change in sales - change in A/R) with lower posterior probabilities. Finally, in the last period, from January 2010 to December 2016, we see more model uncertainty regarding the best model, as the posterior probabilities of the three best models are quite similar (0.36, 0.28, and 0.24). The best models include, in addition to the market return, mom1m (1-month momentum), mve_ia (industry-adjusted market value of equity), sue (unexpected quarterly earnings) or a combination thereof.

The only factor for which we find consistent evidence for all sub-sample periods is the excess market return. Other factors which are selected in some sub-sample periods, such as chmom (change in 6-month momentum), sue (unexpected quarterly earnings) and mom1m (1-month momentum), are not those in models such as FF5 and HXZ. Exceptions are the inclusion of bm (book-to-market) in Panel D (although with low posterior probability) and a size factor (mve_ie) in Panel E. The additional factors that are selected to explain individual stocks are related to other anomalies such as short-term reversal, earnings announcement returns, surprise earning etc.

4.1.3. Which factors explain the returns on microcap stocks?

The main results with non-microcap stocks may be different from those with microcap stocks. Many previous studies report that microcap stocks perform differently mainly due to their illiquidity (Baker & Wurgler, 2006; Stambaugh *et al.* , 2012; Antoniou *et al.* , 2015). Therefore, in this subsection, we apply our methodology to the set of microcap stocks only in each sub-sample period, to investigate which factors matter to explain their returns[22].

[Table 5 about here.]

The results are reported in Table 5. There are similarities in the selected factors over different sub-sample periods with respect to those of non-microcap stocks in Table 4. For example, chmom (change in 6-month momentum) is selected for both groups of stocks during the first sub-sample period and ear (earnings announcement return) is selected in the second sub-sample period. One interesting result is that, during the period January 1995 to June 2002 (Panel C), the two highest posterior probability models do not include the market return factor. The two best models, with together represent a posterior probability of 0.88, include aeavol (abnormal earnings announcement volume) and chanalyst (change in number of analysts covering stock). A possible explanation is that the microcap universe during the period of the high-tech bubble of the 1990s includes a high number of small technology stocks whose prices were extremely sensitive to these variables during this unusual period, and not very sensitive to overall market movements, as many investors were captivated by the possibility of finding the next "hot" technology stock (during the build-up of the bubble) or concerned about any news regarding their technology stocks during the bursting of the bubble.

[22] We report results for five sub-sample periods and the reduced set of factors. The results with three sub-sample periods and the full set of factors do not differ significantly and are available upon request.

Interestingly, smaller numbers of factors are selected compared to non-microcap stocks, and these factors are not the popular ones. Overall, only 6 factors are ever selected in any of the periods: mkt (market excess return), aeavol (abnormal earnings announcement volume), chmom (change in 6-month momentum), ear (earnings announcement return), pchsale_pchrect (change in sales - change in A/R) and sue (unexpected quarterly earnings). From these, only 3 have a marginal posterior probability higher than 0.5: mkt, aeavol, and ear.

4.2. Portfolios as Test Assets

For portfolios of stocks, we test the whole sample period from 1980 to 2016 because there is no survivorship bias in this case. The best model for each set of portfolios and the associated posterior probabilities are reported in Table 6. Several interesting results are reported as follows.

First, when portfolios are formed on firm characteristics, factors related to these characteristics are typically included in the best models. For example, the best model to explain portfolios formed on size includes mve_ia (industry-adjusted size); for portfolios formed on book-to-market, the bm factor is included; for portfolios formed on operating profitability, roic (return on invested capital), which is highly correlated with the factor formed on roe (return on equity), is included, and so on. This pattern also holds for portfolios formed on bivariate sorts. For example, the 25 Fama-French portfolios formed on size and book-to-market require a size factor (mve_ia) and lev (leverage, which has almost 0.70 correlation with the bm factor).

The pattern of dependence between the variable used for portfolio formation and the selected factors reflects the concerns expressed by Lo & Mackinlay (1990), Ferson *et al.* (1999), Berk (2000), Roll (1977) and Lewellen *et al.* (2010). The selected models appear to be incorrectly promising because none of the high posterior probability models include the intercept even for the long testing period, i.e, 444 monthly data.

Second, model uncertainty increases for the portfolios and varies significantly across the different sets of factors. The posterior probabilities of the best models varies from 0.10 for the portfolios formed on long-term reversal to 0.57 for the portfolios formed on operating profitability. For the double sorted portfolios, the posterior probabilities of the best models are around 0.2. Even for the portfolios formed on characteristics related to widely used factor models, the best models also include factors other then the firm charactersitics that are used to form portfolios, for example, Fama & French (2015), Hou *et al.* (2015b), and others.

The model uncertainty indicates that sorting stocks to one or two firm characteristics does not completely remove the effects of other firm characteristics. If firm characteristics are not correlated, then portfolios formed on one firm characteristic should not be explained by factors formed on other firm characteristics. The model uncertainty confirms the problems raised by Fama and French (2008) that forming portfolios based on one or more firm characteristics does not guarantee that these portfolios are not affected by other firm characteristics.

Third, once again, the one most important factor is invariably the market factor. In untabulated results, we calculate the average posterior probability of all factors across all sets of portfolios, and find that the only factor with an average posterior probability higher than 0.5 is the excess market return, which is consistent with the results we obtained using individual stocks.

Finally, we find that the results for the 49 industry portfolios support a five-factor model with the market factor, beta, illiquidity, leverage and organizational capital. Organizational capital represents Selling, general and administrative expenses which is a distinct factor that can only be found in the sector portfolios[23].

[Table 6 about here.]

4.3. Robustness of Results

We investigate how sensitive our results are with respect to the informativeness of priors. We perform robustness test by changing the value of c, a scaling parameter related to the prior variance of the regression coefficient vector $\boldsymbol{\beta}$. Instead of $c = 1$ that we used for our main results, we use $c = 5$, a much less informative prior. We perform the calculations for non-microcap and microcap stocks using the five sub-sample periods, and for each set of portfolios using the whole sample.

The results with less informative priors ($c = 5$) are almost identical to those with $c = 1$ in terms of factor selection and model probabilities. For non-microcap stocks with $c = 5$, the results are not different from those we report in Table 4, and thus we do not report them. For microcap stocks, the results of which are reported in Table 7, there is slight difference in the posterior probability of the market factor, but the difference is only marginal.

[Table 7 about here.]

[23] See Eisfeldt & Papanikolaou (2013).

Table 8 reports the best models and associated posterior probabilities for the different sets of portfolios when the procedure is run with $c = 5$. In most cases, the best model includes fewer factors compared to the results obtained with $c = 1$, which is expected as the prior is less informative, and it becomes *a priori* less likely that factor sensitivities would be generated in regions where the associated γ_j is high, as discussed for example in O'Hara & Sillanpää (2009). Model uncertainty is smaller, in the sense that the best models have higher posterior probabilities, which reflects the fact that, as the priors of the regression coefficients become more diffuse, fewer factors are selected, increasing model probabilities. In a few cases, some factors are dropped and others are included, but similar patterns we reported previously still hold, *i.e.* factors related to the characteristics used for portfolio formation remain in the model.

Overall, we conclude that our results are not sensitive to the prior specification, particularly for individual stocks, where we find virtually identical results.

[Table 8 about here.]

4.4. Comparison with Other Studies

Although there have been several studies that apply a Bayesian approach to asset pricing, comparison is challenging due to the differences in data, both in terms of factors as well as test assets. Specifically, compared to previous studies that use a Bayesian variable selection procedure to identify asset pricing factors (Ericsson & Karlsson (2003), Ouysse & Kohn (2010), Puelz *et al.* (2017)), the most important difference is that we also apply our method to thousands of individual stocks, while these studies only use portfolios. Another relevant difference is that Ericsson & Karlsson (2003) and Ouysse & Kohn (2010) include macroeconomic factors, while we chose to focus on tradable factors based on cross-sectional patterns reported in the literature. Our set of candidate factors is also much broader.

Our tests using a large collection of sets of portfolios revealed a strong pattern of dependence between the portfolio formation criteria and the selected factors, suggesting skepticism in interpreting results of studies that apply these techniques using portfolios which are related to the candidate factors. Our results using a set of 49 industry portfolios (which are not directly formed based on sorting accounting or return characteristics) suggest a model which includes, in addition to the market factor, factors related to beta, illiquidity, leverage, and organizational capital. In comparison, *e.g.* Ericsson & Karlsson (2003)'s results using 10 industry portfolios support the Carhart model with the addition of macroeconomic factors (credit risk spread and industrial production). For portfolios formed on size and book-to-market ratio, our results are comparable to Puelz *et al.* (2017) and Ericsson & Karlsson (2003), but not surprisingly they favor model which include factors (co)related to size and book-to-market, with the addition of illiquidity in our case.

Recently, Barillas & Shanken (2017) developed a Bayesian asset pricing test which can be calculated in closed form and, in principle, be used to test all possible models using a set of candidate factors. However, in their empirical tests they only considered the factors in FF5, HXZ, as well as a different version of HML proposed by Asness & Frazzini (2013) and momentum (a total of 10 factors). Their tests, conducted on the factors themselves and on sets of portfolios formed on either size and momentum or book-to-market and investment, support a six-factor model with the market return, the HXZ versions of investment (IA) and profitability (ROE), the FF5 version of size (SMB), the modified HML factor from Asness & Frazzini (2013), and the momentum factor. These results are not unexpected, as the portfolios are related to the factors.

Since we build tradable factors based on the characteristics studied by Green *et al.* (2017), it is interesting to compare our results with theirs. They identify 9 characteristics which are significant determinants of non-microcap stocks[24]. In comparison with our results, the only commonalities are earnings announcement return and 1-month momentum, while we also find that change in 6-month momentum, market value of equity, and unexpected quarterly earnings are important factors, for some periods. When they include microcaps, 3 additional characteristics (book-to-market, change in 6-month momentum, and zero trading days) are also significant. There are similarities with our results, as we find that change in the number of analysts, earnings announcement, and change in 6-month momentum are important factors to explain microcap stocks. However, our results suggest that these factors are not consistently selected in different sub-sample periods. Also, similarly to Green *et al.* (2017), we find that the factors from prominent models such as FF and HXZ are not relevant to explain individual stocks.

Our work is also related to recent studies that test factors using procedures to directly account for data mining issues. For example, using a multiple testing framework based on a bootstrapping procedure with individual stocks, Harvey & Liu (2016) test a set of 14 factors that includes many of the ones in our study,

[24]These are cash, change in the number of analysts, earnings announcement return, one-month momentum, the number of consecutive quarters with an increase in earnings over the same quarter a year ago, annual R&D to market cap, return volatility, share turnover, and volatility of share turnover.

and find evidence that the most important factor is, by far, the market return, with only a small role for the profitability factor. We note that, while Harvey & Liu (2016)'s approach and set of factors is quite different from ours, their conclusion regarding the importance of the market return for individual stocks is mirrored in our results.

5. Conclusion

The asset pricing literature has proposed hundreds of factors to explain asset returns, most within the last two decades. It is unlikely that so many factors matter to determine security prices; rather, some are likely to be redundant, while others (or even most) may be product of data mining. In this paper we propose a Bayesian variable selection methodology to explore the most promising linear factor models, given a set of candidate factors and a set of assets. The proposed methodology builds on the literature on Bayesian variable selection in multivariate regression models and provides a computationally feasible means of exploring model selection in large panels of data.

We apply the methodology to identify the most relevant factors to explain returns on individual stocks, as well as an extensive set of portfolios. We consider a large set of 83 candidate factors, including 82 tradable factors based on various firm characteristics identified in the literature, as well as the market return suggested by Sharpe (1964).

Using individual stocks, we find that (i) the only factor that matters across all sub-sample periods is the market excess return; (ii) factor selection varies substantially over time, with no specific model dominating the others in the various sub-sample periods we investigate; (iii) other factors (in addition to the market return) which are selected for specific sub-sample periods are not the factors in widely used models such as the ones proposed by Fama & French (1992, 1996), Chen & Zhang (2010), Hou *et al.* (2015b) and Fama & French (2015). The additional factors that are selected in certain periods to explain individual stocks are related to other anomalies or characteristics such as short-term reversal, change in 6-month momentum, earnings announcement return, change in the number of analysts covering stocks, industry concentration, unexpected quarterly earnings, and industry-adjusted size; (iv) the total number of factors selected in these models over all sub-sample periods is small relative to the total number of candidate factors, *i.e.* only 10 factors (out of more than 80) are ever selected by the variable selection methodology, and from these, only 5 to 6 have a marginal posterior probability higher than 0.5; and (v) the factors that matter to explain microcap stocks include factors formed on change in six-month momentum, abnormal earnings announcement volume and change in number of analysts covering stock.

Our work builds on the literature on asset pricing factor selection, by showing that, despite the large number of factors that have been proposed, only a handful appear to explain the returns on individual stocks, with the market return remaining the most important factor. We leave for future research refinements of the model to allow even more efficient exploration of the model space when the numbers of factors and assets are large.

References

Abarbanell, Jeffery S, & Bushee, Brian J. 1998. Abnormal returns to a fundamental analysis strategy. *Accounting Review*, 19–45.

Ali, Ashiq, Hwang, Lee-Seok, & Trombley, Mark A. 2003. Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, **69**(2), 355–373.

Almeida, Heitor, & Campello, Murillo. 2007. Financial constraints, asset tangibility, and corporate investment. *The Review of Financial Studies*, **20**(5), 1429–1460.

Amihud, Yakov. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, **5**(1), 31–56.

Amihud, Yakov, & Mendelson, Haim. 1989. The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns. *The Journal of Finance*, **44**(2), 479–486.

Anderson, Christopher W, & Garcia-Feijóo, LUIS. 2006. Empirical evidence on capital investment, growth options, and security returns. *The Journal of Finance*, **61**(1), 171–194.

Ando, Tomohiro. 2011. Bayesian Variable Selection for the Seemingly Unrelated Regression Model with a Large Number of Predictors. *J. Japan Statist. Soc.*, **41**(2), 187–203.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. 2006. The cross-section of volatility and expected returns. *Journal of Finance*, **61**, 259–299.

Antoniou, Constantinos, Doukas, John A, & Subrahmanyam, Avanidhar. 2015. Investor sentiment, beta, and the cost of equity capital. *Management Science*, **62**(2), 347–367.

Asness, Clifford, & Frazzini, Andrea. 2013. The devil in HML's details. *The Journal of Portfolio Management*, **39**(4), 49–68.

Asness, Clifford S, Porter, R Burt, & Stevens, Ross L. 2000. Predicting stock returns using industry-relative firm characteristics.

Baker, Malcolm, & Wurgler, Jeffrey. 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, **61**(4), 1645–1680.

Balakrishnan, Karthik, Bartov, Eli, & Faurel, Lucile. 2010. Post loss/profit announcement drift. *Journal of Accounting and Economics*, **50**(1), 20–41.

Bandyopadhyay, Sati P, Huang, Alan G, & Wirjanto, Tony S. 2010. *The accrual volatility anomaly*. Tech. rept. working paper, University of Waterloo.

Banz, Rolf W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics*, **9**, 3–18.

Barbee Jr, William C, Mukherji, Sandip, & Raines, Gary A. 1996. Do Sales–Price and Debt–Equity Explain Stock Returns Better than Book–Market and Firm Size? *Financial Analysts Journal*, **52**(2), 56–60.

Barillas, Francisco, & Shanken, Jay. 2017. Comparing Asset Pricing Models. *The Journal of Finance*, forthcoming.

Basu, S. 1977. Investment performance of common stocks in relation to their price-earnings ratios: a test of the efficient market hypothesis. *Journal of Finance*, **32**, 663–682.

Bauman, W Scott, & Dowen, Richard. 1988. Growth Projections And commin Stock Returns. *Financial Analysts Journal*, **44**(4), 79.

Belo, Frederico, Lin, Xiaoji, & Bazdresch, Santiago. 2014. Labor hiring, investment, and stock return predictability in the cross section. *Journal of Political Economy*, **122**(1), 129–177.

Benjamini, Yoav, & Yekutieli, Daniel. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165–1188.

Berk, Jonathan B. 2000. Sorting out sorts. *The Journal of Finance*, **55**(1), 407–427.

Bhandari, Laxmi Chand. 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. *The journal of finance*, **43**(2), 507–528.

Brandt, Michael W, Kishore, Runeet, Santa-Clara, Pedro, & Venkatachalam, Mohan. 2008. Earnings announcements are full of surprises. *SSRN eLibrary*.

Brown, David P, & Rowe, Bradford. 2007. *The productivity premium in equity returns*. Tech. rept. working paper, University of Wisconsin.

Brown, P J, Vannucci, M, & Fearn, T. 1998. Multivariate Bayesian variable selection and prediction. *J. R. Statist. Soc. B*, **60**(3), 627–641.

Brown, P J, Vannucci, M, & Fearn, T. 2002. Bayes Model Averaging With Selection of Regressors. *J. R. Statist. Soc. B*, **64**(3), 519–536.

Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of finance*, **52**(1), 57–82.

Chandrashekar, Satyajit, & Rao, Ramesh KS. 2009. *The productivity of corporate cash holdings and the cross-section of expected stock returns*. Tech. rept. McCombs Research Paper Series No. FIN-03-09.

Chen, Long, & Zhang, Lu. 2010. A better three-factor model that explains more anomalies. *Journal of Finance*, **65**(2), 563–595.

Chen, Nai-Fu, Roll, Richard, & Ross, Stephen A. 1986. Economic Forces and the Stock Market. *The Journal of Business*, **59**(3), 383–403.

Chordia, Tarun, Subrahmanyam, Avanidhar, & Anshuman, V Ravi. 2001. Trading activity and expected stock returns. *Journal of Financial Economics*, **59**(1), 3–32.

Chordia, Tarun, Goyal, Amit, & Saretto, Alessio. 2017. *p-hacking: Evidence from two million trading strategies*.

Clyde, Merlise, & George, Edward I. 2004. Model Uncertainty. *Statistical Science*, **19**(1), 81–94.

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

Connor, Gregory, & Korajczyk, Robert A. 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, **21**(2), 255–289.

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

Datar, Vinay T, Naik, Narayan Y, & Radcliffe, Robert. 1998. Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, **1**(2), 203–219.

Dellaportas, R, Forster, J, & Ntzoufras, J. 1999. Bayesian Variable Selection Using the Gibbs Sampler. *Pages 1–19 of: Generalized Linear Models: A Bayesian Perspective*.

Desai, Hemang, Rajgopal, Shivaram, & Venkatachalam, Mohan. 2004. Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review*, **79**(2), 355–385.

Diether, Karl B, Malloy, Christopher J, & Scherbina, Anna. 2002. Differences of opinion and the cross section of stock returns. *The Journal of Finance*, **57**(5), 2113–2141.

Eisfeldt, Andrea L, & Papanikolaou, Dimitris. 2013. Organization capital and the cross-section of expected returns. *The Journal of Finance*, **68**(4), 1365–1406.

Elgers, Pieter T, Lo, May H, & Pfeiffer Jr, Ray J. 2001. Delayed security price adjustments to financial analysts' forecasts of annual earnings. *The Accounting Review*, **76**(4), 613–632.

Ericsson, Johan, & Karlsson, Sune. 2003. *Choosing factors in a multifactor asset pricing model: A Bayesian approach*. Tech. rept. SSE/EFI Working Paper Series in Economics and Finance.

Fairfield, Patricia M, Whisenant, J Scott, & Yohn, Teri Lombardi. 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. *The accounting review*, **78**(1), 353–371.

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

Fama, E. F., & MacBeth, J. D. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy*, **71**, 43–66.

Fama, Eugene F., & French, Kenneth R. 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance*, **51**(1), 55–84.

Fama, Eugene F, & French, Kenneth R. 2015. A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, **116**, 1–22.

Feng, Guanhao, Giglio, Stefano, & Xiu, Dacheng. 2017. *Taming the Factor Zoo*.

Ferson, Wayne E, Sarkissian, Sergei, & Simin, Timothy. 1999. The alpha factor asset pricing model: A parable. *Journal of Financial Markets*, **2**(1), 49–68.

Francis, Jennifer, LaFond, Ryan, Olsson, Per M, & Schipper, Katherine. 2004. Costs of equity and earnings attributes. *The accounting review*, **79**(4), 967–1010.

George, E I, & McCulloch, Robert E. 1993. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, **88**(423), 881–889.

George, E I, & McCulloch, Robert E. 1997. Approaches for Bayesian Variable Selection. *Statistica Sinica*, **7**, 339–373.

Gettleman, Eric, & Marks, Joseph M. 2006. *Acceleration strategies*. Tech. rept. Working Paper, Seton Hall Univeristy.

Gibbons, Michael R., Ross, Stephen A., & Shanken, Jay. 1989. A Test of the Efficiency of a Given Portfolio. *Econometrica*, **57**(5), 1121.

Giles, David E A. 2003. *Computer-Aided Econometrics*. New York, N.Y.: Marcel Dekker, Inc.

Green, Jeremiah, Hand, John RM, & Zhang, X Frank. 2017. The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, hhx019.

Greene, W H. 2003. *Econometric Analysis*. Upper Saddle River, N.J.: Prentice Hall.

Hafzalla, Nader, Lundholm, Russell, & Matthew Van Winkle, E. 2011. Percent accruals. *The Accounting Review*, **86**(1), 209–236.

Hall, Anthony, Hwang, Soosung, & Satchell, Stephen E. 2002. Using Bayesian variable selection methods to choose style factors in global stock return models. *Journal of Banking and Finance*, **26**(12), 2301–2325.

Hans, C, Dobra, A, & West, M. 2007. Shotgun stochastic search for ?large p? regression. *Journal of the American Statistical Association*, **102**, 507?516.

Harvey, Campbell R, & Liu, Yan. 2016. Lucky factors. *Department of Finance*.

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

Hawkins, Eugene H, Chamberlin, Stanley C, & Daniel, Wayne E. 1984. Earnings expectations and security prices. *Financial Analysts Journal*, **40**(5), 24–38.

Holthausen, Robert W, & Larcker, David F. 1992. The prediction of stock returns using financial statement information. *Journal of Accounting and Economics*, **15**(2-3), 373–411.

Hou, Kewei, & Moskowitz, Tobias J. 2005. Market frictions, price delay, and the cross-section of expected returns. *The Review of Financial Studies*, **18**(3), 981–1020.

Hou, Kewei, & Robinson, David T. 2006. Industry concentration and average stock returns. *The Journal of Finance*, **61**(4), 1927–1956.

Hou, Kewei, Xue, Chen, & Zhang, Lu. 2015a. Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, **28**(3), 650–705.

Hou, Kewei, Xue, Chen, & Zhang, Lu. 2015b. Digesting anomalies: An investment approach. *The Review of Financial Studies*, **28**(3), 650–705.

Huang, Alan Guoming. 2009. The cross section of cashflow volatility and expected stock returns. *Journal of Empirical Finance*, **16**(3), 409–429.

Jegadeesh, Narasimhan. 1990. Evidence of predictable behavior of security returns. *The Journal of finance*, **45**(3), 881–898.

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

Jiang, Guohua, Lee, Charles MC, & Zhang, Yi. 2005. Information uncertainty and expected returns. *Review of Accounting Studies*, **10**(2), 185–221.

Kama, Itay. 2009. On the market reaction to revenue and earnings surprises. *Journal of Business Finance & Accounting*, **36**(1-2), 31–50.

Kewei, Hou, Chen, Xue, & Zhang, Lu. 2017. *Replicating Anomalies*.

Kim, Chang-Jin, & Nelson, Charles R. 1999. State-space models with regime switching: classical and Gibbs-sampling approaches with applications. *MIT Press Books*, **1**.

Kuo, Lynn, & Mallick, Bani. 1998. Variable selection for regression models. *Sankhya*, **60**(1).

Lakonishok, Josef, Shleifer, Andrei, & Vishny, Robert W. 1994. Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance*, **49**(5), 1541–1578.

Lehmann, Bruce N, & Modest, David M. 1988. The empirical foundations of the arbitrage pricing theory. *Journal of financial economics*, **21**(2), 213–254.

Lerman, Alina, Livnat, Joshua, & Mendenhall, Richard R. 2008. *The high-volume return premium and post-earnings announcement drift*. Tech. rept. Working paper, New York University.

Lev, Baruch, & Nissim, Doron. 2004. Taxable income, future earnings, and equity values. *The Accounting Review*, **79**(4), 1039–1074.

Lewellen, Jonathan, Nagel, Stefan, & Shanken, Jay. 2010. A Skeptical Appraisal of Asset-Pricing Tests A Skeptical Appraisal of Asset-Pricing Tests. *Journal of Financial Economics*, **96**(May 2005), 175–194.

Liu, Weimin. 2006. A liquidity-augmented capital asset pricing model. *Journal of financial Economics*, **82**(3), 631–671.

Lo, Andrew W., & Mackinlay, A. Craig. 1990. When are Contrarian Profits Due to Stock Market Overreaction? *The Review of Financial Studies*, **3**(2), 175–205.

McLean, R David, & Pontiff, Jeffrey. 2016. Does academic research destroy stock return predictability? *The Journal of Finance*, **71**(1), 5–32.

Mohanram, Partha S. 2005a. Separating winners from losers among lowbook-to-market stocks using financial statement analysis. *Review of accounting studies*, **10**(2), 133–170.

Mohanram, Partha S. 2005b. Separating Winners from Losers among LowBook-to-Market Stocks using Financial Statement Analysis. *Review of Accounting Studies*, **10**, 133–170.

Moskowitz, Tobias J., & Grinblatt, Mark. 1999. Do industries explain momentum? *Journal of Finance*.

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

O'Hara, R B, & Sillanpää, M J. 2009. A Review of Bayesian Variable Selection Methods: What, How and Which. *Bayesian Analysis*, **4**(1), 85?118.

Ou, Jane A, & Penman, Stephen H. 1989. Financial statement analysis and the prediction of stock returns. *Journal of accounting and economics*, **11**(4), 295–329.

Ouysse, Rachid, & Kohn, Robert. 2010. Bayesian variable selection and model averaging in the arbitrage pricing theory model. *Computational Statistics and Data Analysis*, **54**(12), 3249?3268.

Palazzo, Berardino. 2012. Cash holdings, risk, and expected returns. *Journal of Financial Economics*, **104**(1), 162–185.

Pastor, Lubos, & Stambaugh, Robert F Rf F. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*.

Piotroski, Joseph D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, 1–41.

Pontiff, Jeffrey, & Woodgate, Artemiza. 2008. Share issuance and cross-sectional returns. *The Journal of Finance*, **63**(2), 921–945.

Puelz, David, Hahn, P Richard, Carvalho, Carlos M, *et al.* . 2017. Variable selection in seemingly unrelated regressions with random predictors. *Bayesian Analysis*.

Rendleman, Richard J, Jones, Charles P, & Latane, Henry A. 1982. Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics*, **10**(3), 269–287.

Richardson, Scott A, Sloan, Richard G, Soliman, Mark T, & Tuna, Irem. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of accounting and economics*, **39**(3), 437–485.

Roll, Richard. 1977. A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of financial economics*, **4**(2), 129–176.

Roll, Richard, & Ross, Stephen A. 1980. An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of Finance*.

Rosenberg, B., Reid, K., & Lanstein, R. 1985. Persuasive Evidence of Market Inefficiency. *Journal of Portfolio Management*, **11**, 9–11.

Scherbina, Anna. 2008. Suppressed negative information and future underperformance. *Review of Finance*, 533–565.

Sharpe, W.F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of finance*, **19**(3), 425–442.

Sloan, RG. 1996. Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *Accounting Review*, **71**(3), 289–315.

Smith, Michael, & Kohn, Robert. 2000. Nonparametric seemingly unrelated regression. *Journal of Econometrics*, **98**(2), 257–281.

Soliman, Mark T. 2008. The use of DuPont analysis by market participants. *The Accounting Review*, **83**(3), 823–853.

Stambaugh, Robert F, Yu, Jianfeng, & Yuan, Yu. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, **104**(2), 288–302.

Sullivan, Ryan, Timmermann, Allan, & White, Halbert. 1999. Data-snooping, technical trading rule performance, and the bootstrap. *The journal of Finance*, **54**(5), 1647–1691.

Sullivan, Ryan, Timmermann, Allan, & White, Halbert. 2001. Dangers of data mining: The case of calendar effects in stock returns. *Journal of Econometrics*, **105**(1), 249–286.

Thomas, Jacob K, & Zhang, Huai. 2002. Inventory changes and future returns. *Review of Accounting Studies*, **7**(2), 163–187.

Titman, Sheridan, Wei, KC John, & Xie, Feixue. 2004. Capital investments and stock returns. *Journal of financial and Quantitative Analysis*, **39**(4), 677–700.

Tuzel, Selale. 2010. Corporate real estate holdings and the cross-section of stock returns. *The Review of Financial Studies*, **23**(6), 2268–2302.

Wang, Hao. 2010. Sparse seemingly unrelated regression modelling: Applications in finance and econometrics. *Computational Statistics and Data Analysis*, 2866–2877.

Yan, Xuemin, & Zheng, Lingling. 2017. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *The Review of Financial Studies*, **30**(4), 1382–1423.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of American Statistical Association*, 500–509.

Zellner, A. 1971. *An introduction to bayesian inference in econometrics*. John Wiley & Sons.

Figure 1: **Histogram of absolute pairwise factor correlations**

The figure plots the distribution of the pairwise absolute correlations of 83 factors, including 82 factors formed on firm characteristics obtained following Green *et al.* (2017), and the market excess return.

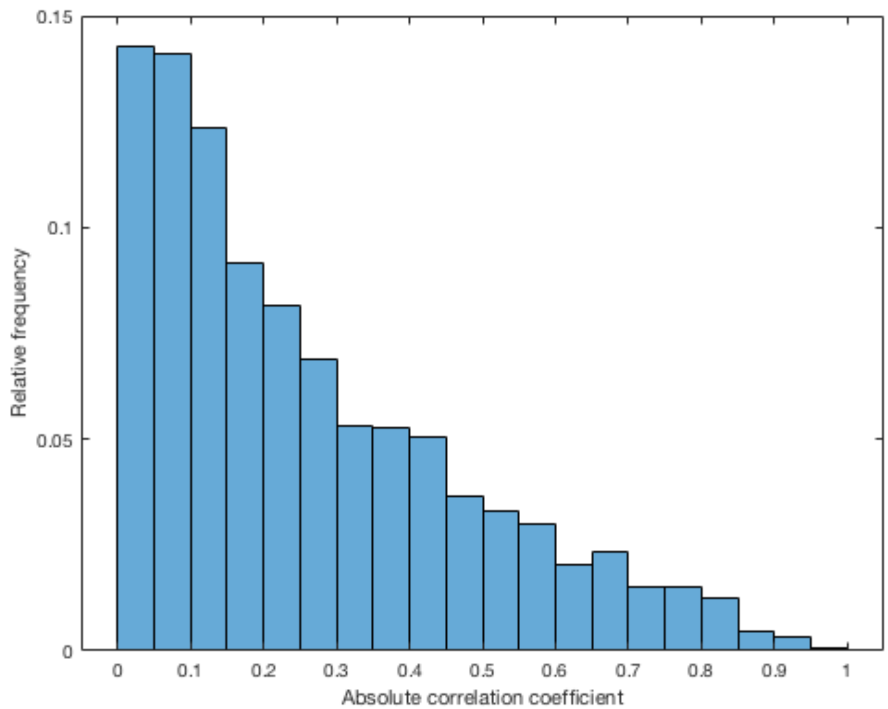

Table 1: Statistics of Candidate Factors

The 82 factors are constructed from value-weighted portfolios sorting all non-microcap stocks into deciles based on the characteristics of Green *et al.* (2017). The factor returns are calculated as the difference between the top and bottom deciles. We also report statistics for the market excess return, calculated as the excess return on the market, value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t, minus the one-month Treasury bill rate. The table reports the first date at which the factor has been calculated, the total number of months, the average monthly return, the standard deviation, and the t-statistic. An */bold line denotes a Dependent False Discovery Rate (DFDR) p-value lower than 0.05, calculated using the method of Benjamini & Yekutieli (2001).

Factor	First Date	#Months	Average Return	Standard Deviation	Tstat	Factor	First Date	#Months	Average Return	Standard Deviation	Tstat
mkt	198001	444	0.65%	4.46%	3.05	lev	198001	444	0.27%	4.47%	1.29
absacc	198001	444	0.06%	3.58%	0.33	mom12m	198001	444	0.68%	7.40%	1.93
acc	**198001**	**444**	**-0.45%**	**2.61%**	**-3.61***	mom1m	198001	444	-0.27%	5.39%	-1.06
aeavol	198001	444	0.00%	2.69%	0.00	mom36m	198001	444	-0.58%	5.22%	-2.35
age	198001	444	0.04%	4.46%	0.20	ms	198001	444	0.16%	3.25%	1.02
agr	198001	444	-0.45%	3.30%	-2.90	mve	198001	444	-0.52%	4.68%	-2.36
baspread	198001	444	-0.10%	8.25%	-0.26	mve_ia	198001	444	-0.16%	3.22%	-1.06
beta	198001	444	-0.06%	8.79%	-0.14	**nanalyst**	**200707**	**114**	**-0.91%**	**2.27%**	**-4.26***
bm	198001	444	0.44%	4.46%	2.09	operprof	198001	444	0.36%	2.98%	2.52
bm_ia	198001	444	0.29%	4.44%	1.38	orgcap	198001	444	0.58%	5.29%	2.32
cash	198001	444	0.27%	4.68%	1.23	pchcapx_ia	198001	444	-0.13%	3.80%	-0.74
cashdebt	198001	444	0.11%	3.42%	0.71	pchcurrat	198001	444	-0.22%	1.74%	-2.67
cashpr	198001	444	-0.40%	3.38%	-2.47	pchdepr	198001	444	0.16%	2.34%	1.42
cfp	198001	444	0.47%	4.90%	2.01	pchgm_pchsale	198001	444	0.20%	2.36%	1.78
cfp_ia	198001	444	-0.05%	4.27%	-0.24	pchsaleinv	198001	444	0.20%	2.34%	1.83
chatoia	198001	444	0.34%	2.52%	2.85	pchsale_pchinvt	198001	444	0.17%	2.29%	1.58
chcsho	**198001**	**444**	**-0.51%**	**3.01%**	**-3.54***	pchsale_pchrect	198001	444	0.08%	2.10%	0.77
chempia	198001	444	0.00%	2.97%	0.02	pchsale_pchxsga	198001	444	-0.14%	2.83%	-1.06
chfeps	198901	336	0.25%	3.73%	1.23	pctacc	198001	444	-0.17%	2.74%	-1.28
chinv	**198001**	**444**	**-0.57%**	**2.98%**	**-4.05***	pricedelay	198001	444	0.04%	2.59%	0.31
chmom	198001	444	-0.49%	4.64%	-2.23	ps	198001	444	0.49%	4.24%	2.42
chnanalyst	198904	333	-0.02%	2.20%	-0.20	realestate	198501	384	0.26%	4.57%	1.12
chpmia	198001	444	-0.17%	3.55%	-1.02	retvol	198001	444	-0.31%	7.78%	-0.83
chtx	198001	444	0.18%	3.15%	1.18	roaq	198001	444	0.37%	4.17%	1.87
cinvest	198001	444	0.07%	2.09%	0.66	roavol	198001	444	-0.18%	4.49%	-0.86
currat	198001	444	-0.14%	4.59%	-0.64	roeq	198001	444	0.34%	4.34%	1.63
depr	198001	444	0.06%	5.20%	0.23	roic	198001	444	0.35%	3.98%	1.84
disp	198901	336	-0.35%	5.00%	-1.30	rsup	198001	444	-0.23%	3.53%	-1.40
ear	198001	444	0.32%	2.33%	2.93	salecash	198001	444	-0.04%	4.26%	-0.20
egr	198001	444	-0.43%	3.18%	-2.84	saleinv	198001	444	0.25%	2.95%	1.80
ep	198001	444	0.29%	5.41%	1.14	salerec	198001	444	0.44%	3.51%	2.63
fgr5yr	198901	336	0.15%	6.59%	0.43	**sfe**	**198901**	**336**	**-1.06%**	**4.88%**	**-3.99***
gma	198001	444	0.17%	3.21%	1.11	sgr	198001	444	-0.15%	3.66%	-0.86
grcapx	198001	444	-0.37%	2.88%	-2.70	sp	198001	444	0.44%	4.25%	2.17
grltnoa	198001	444	-0.42%	2.76%	-3.20	stdcf	198001	444	-0.28%	4.12%	-1.42
herf	200001	204	-0.11%	4.29%	-0.38	std_dolvol	198001	444	0.24%	3.19%	1.57
hire	198001	444	-0.34%	3.33%	-2.12	std_turn	198001	444	0.00%	5.50%	0.00
idiovol	198001	444	-0.21%	7.82%	-0.56	sue	198001	444	0.41%	3.34%	2.57
ill	198001	444	0.31%	3.78%	1.70	tang	198001	444	0.17%	3.89%	0.94
indmom	199408	269	0.26%	6.80%	0.63	tb	198001	444	0.10%	2.69%	0.81
invest	**198001**	**444**	**-0.54%**	**3.08%**	**-3.68***	turn	198001	444	-0.10%	5.78%	-0.38
						zerotrade	198001	444	0.07%	5.53%	0.26

Table 2: Statistics of Factor Correlations

The table reports summary statistics of the pairwise absolute correlations of a set of 83 factors, including 82 factors on firm characteristics and the market excess return.

Statistics of (absolute) correlations among candidate factors								
# Factors	**# correlations**	**Min**	**10th percentile**	**25th percentile**	**Median**	**75th percentile**	**90th percentile**	**Maximum**
83	3403	0.000	0.033	0.081	0.179	0.333	0.498	0.971
10 highest absolute correlations								
ill_mve	0.971							
turn_zerotrade	0.963							
baspread_retvol	0.960							
pchsaleinv_pchsale_pchinvt	0.934							
baspread_beta	0.895							
beta_retvol	0.880							
roaq_roeq	0.877							
cashdebt_roic	0.877							
idiovol_std_turn	0.874							
baspread_idiovol	0.869							

Table 3: Posterior model probabilities for individual stocks, 3 sub-sample periods

We apply the Bayesian variable selection method to non-microcap and non-penny stocks in each sub-sample period and report the selected models with their posterior probabilities. The set of candidate factors includes all available factors in each sub-sample period.

Panel A. January 1980 - December 1991, # stocks = 807, # factors =75

Model	**# Factors**	**Posterior probability**
mkt, chmom	2	0.24
mkt, mom1m	2	0.16
mkt, mom1m, ms	3	0.16
mkt	1	0.08
mkt, chmom, ms	3	0.08
mkt, chmom, mom1m	3	0.08
mkt, chmom, mom1m, ms	4	0.08
mkt, ep	2	0.04
mkt, ep, ms	3	0.04
mkt, chmom, tb	3	0.04

Factors with marginal posterior probability > 0: mkt, chmom, ep, mom1m, ms, tb
Factors with marginal posterior probability > 0.5: mkt, chmom

Panel B. January 1992 - December 2003, # stocks = 893, # factors = 81

Model	**# Factors**	**Posterior probability**
mkt	1	0.64
mkt,aeavol	2	0.32
mkt,aeavol,pctacc	3	0.04

Factors with marginal posterior probability > 0: mkt, aeavol, pctacc
Factors with marginal posterior probability > 0.5: mkt

Panel C. January 2004 - December 2016, # stocks = 967, # factors = 83

Model	**# Factors**	**Posterior probability**
intercept,mkt,herf,mom1m	4	0.24
mkt,chnanalyst,herf,mom1m	4	0.2
mkt,mom1m	2	0.08
intercept,mkt,herf,mom1m,saleinv	5	0.08
intercept,mkt,chnanalyst,herf,mom1m	5	0.08
mkt,herf,mom1m	3	0.04
mkt,ear,herf,mom1m	4	0.04
mkt,chnanalyst,herf,mom1m,ms	5	0.04
mkt,chnanalyst,ear,herf,mom1m	5	0.04
intercept,mkt,herf,mom1m,ms	5	0.04
intercept,mkt,chnanalyst,herf,mom1m,saleinv	6	0.04
intercept,mkt,chnanalyst,herf,mom1m,ms	6	0.04
intercept,mkt,chnanalyst,ear,herf,mom1m,saleinv	7	0.04

Factors with marginal posterior probability > 0: intercept, mkt, chnanalyst, ear, herf, mom1m, ms, saleinv
Factors with marginal posterior probability > 0.5: intercept, mkt, herf

Table 4: Posterior model probabilities obtained with non-microcap stocks, 5 sub-sample periods

We apply the Bayesian variable selection method to all non-microcap stocks in each sub-sample period and report the models with the highest posterior probability. The set of candidate factors includes all significant factors in Green *et al.* (2017) in addition to the intercept and the excess market return.

Panel A. January 1980 - June 1987, # stocks = 1,014, # factors = 44

Model	# Factors	Posterior probability
mkt,chmom	2	1

Factors with marginal posterior probability > 0: mkt, chmom
Factors with marginal posterior probability > 0.5: mkt, chmom

Panel B. July 1987 - December 1994, # stocks = 1,114, # factors = 44

Model	# Factors	Posterior probability
mkt,chmom,ear,mom1m	4	0.8
mkt,chmom,ear	3	0.2

Factors with marginal posterior probability > 0: mkt, chmom, ear, mom1m
Factors with marginal posterior probability > 0.5: mkt, chmom, ear, mom1m

Panel C. January 1995 - June 2002, # stocks = 1,112, # factors = 48

Model	# Factors	Posterior probability
mkt	1	0.44
mkt,pctacc	2	0.16
mkt,mom1m	2	0.12
mkt,mom1m,pctacc	3	0.08
mkt,aeavol	2	0.08

Factors with marginal posterior probability > 0: mkt, aeavol, mom1m, pctacc
Factors with marginal posterior probability > 0.5: mkt

Panel D. July 2002 - December 2009, # stocks = 1,296, # factors = 48

Model	# Factors	Posterior probability
mkt,sue	2	0.44
mkt,bm	2	0.16
mkt,bm,sue	3	0.16
mkt	1	0.08
mkt,pchsale_pchrect	2	0.08

Factors with marginal posterior probability > 0: mkt, bm, pchsale_pchrect, sue
Factors with marginal posterior probability > 0.5: mkt, sue

Panel E. January 2010 - December 2016, # stocks = 1,225, # factors = 49

Model	# Factors	Posterior probability
mkt,mom1m,mve_ia	3	0.36
mkt,mve_ia	2	0.28
mkt,mom1m,mve_ia,sue	4	0.24
mkt,mve_ia,sue	3	0.08
mkt,mom1m,sue	3	0.04

Factors with marginal posterior probability > 0: mkt, mom1m, mve_ia, sue
Factors with marginal posterior probability > 0.5: mkt, mom1m, mve_ia

Table 5: Posterior model probabilities obtained with microcap stocks

We apply the Bayesian variable selection method to microcap stocks only in each sub-sample period and report the models with the highest posterior probability. The set of candidate factors includes factors that are significant in any regression of Green *et al.* (2017) in addition to the excess market return and intercept.

Panel A. January 1980 - June 1987, # stocks = 868, # factors = 44

Model	**# Factors**	**Posterior probability**
mkt	1	0.44
mkt, aeavol	2	0.40
mkt, chmom	2	0.12
mkt, aeavol, chmom	3	0.04

Factors with marginal posterior probability > 0: mkt, aeavol, chmom
Factors with marginal posterior probability > 0.5: mkt

Panel B. July 1987 - December 1994, # stocks = 1,138, # factors = 44

Model	**# Factors**	**Posterior probability**
mkt, ear	2	1.00

Factors with marginal posterior probability > 0: mkt, ear
Factors with marginal posterior probability > 0.5: mkt, ear

Panel C. January 1995 - June 2002, # stocks = 1,289, # factors = 48

Model	**# Factors**	**Posterior probability**
aeavol	1	0.48
aeavol, chnanalyst	2	0.40
mkt, aeavol	2	0.04
mkt, aeavol, chnanalyst	3	0.04

Factors with marginal posterior probability > 0: mkt, aeavol, chnanalyst
Factors with marginal posterior probability > 0.5: aeavol

Panel D. July 2002 - December 2009, # stocks = 1,119, # factors = 48

Model	**# Factors**	**Posterior probability**
mkt, pchsale_pchrect	2	0.44
mkt	1	0.40
mkt, sue	2	0.12
mkt, pchsale_pchrect, sue	3	0.04

Factors with marginal posterior probability > 0: mkt, pchsale_pchrect, sue
Factors with marginal posterior probability > 0.5: mkt

Panel E. January 2010 - December 2016, # stocks = 1,058, # factors = 49

Model	**# Factors**	**Posterior probability**
mkt	1	1.00

Factors with marginal posterior probability > 0: mkt
Factors with marginal posterior probability > 0.5: mkt

Table 6: Summary of best models identified using different sets of portfolios, 1980-2016

We apply the Bayesian variable selection methodology to sets of portfolios formed according to various criteria, for the period 1980 to 2016. The table reports the best model, *i.e.* the model with highest posterior probability, the number of factors in the model, and the posterior probability.

Portfolio formation	**# Portfolios**	**Best model**	**# Factors**	**Probability**
Univariate Sorts				
Size	10	mkt, ill, mve_ia, std_dolvol	4	0.21
Book-to-market	10	mkt, bm, idiovol, lev	4	0.31
Operating profitability	10	mkt, roavol, roic	3	0.57
Investment	10	mkt, agr, roavol, sgr	4	0.47
Earnings-to-price	10	mkt, age, ep, lev	4	0.32
Cashflow-to-price	10	mkt, age, ep, lev	4	0.33
Dividend Yield	10	mkt, age, beta, cashpr, salecash	5	0.13
Momentum	10	mkt, age, mom12m, roavol	4	0.20
Short-term reversal	10	mkt, mom1m, std_turn	3	0.54
Long-term reversal	10	mkt, lev, mom36m, std_turn	4	0.10
Beta	10	mkt, age, beta, idiovol	4	0.29
Variance	10	mkt, retvol, salecash, stdcf	4	0.18
Residual variance	10	mkt, idiovol, retvol, stdcf	4	0.30
Bivariate Sorts				
Size and book-to-market	25	mkt, ill, lev, mve_ia, roavol, std_dolvol	6	0.20
Size and operating profitability	25	mkt, age, idiovol, ill, mve_ia, roavol, std_dolvol, std_turn, zerotrade	9	0.20
Size and investing	25	mkt, beta, ill, mve_ia, sgr, stdcf	6	0.20
Book-to-market and operating profitability	25	mkt, baspread, beta, bm, cash, lev, retvol, roavol, salecash, stdcf	10	0.20
Book-to-market and investment	25	mkt, bm, lev	3	0.20
Operating profitability and investment	25	mkt, roic, std_turn, turn	4	0.20
Industry Portfolios				
Industries	49	mkt, beta, ill, lev, orgcap	5	0.40

Table 7: Robustness test - posterior model probabilities obtained with microcap stocks, $c = 5$

We apply the Bayesian variable selection method to all microcap stocks in each sub-sample period and report the models with the highest posterior probability. The set of candidate factors includes all significant factors in Green *et al.* (2017) as well as the excess market return. We report results for $c = 5$, where c is a scaling parameter related to the prior variance of the regression coefficient vector $\boldsymbol{\beta}$.

Panel A. January 1980 - June 1987, # stocks = 868, # factors = 44

Model	**# Factors**	**Posterior probability**
mkt	1	0.64
mkt, aeavol	2	0.20
mkt, chmom	2	0.16

Factors with marginal posterior probability > 0: mkt, aeavol, chmom
Factors with marginal posterior probability > 0: mkt

Panel B. July 1987 - December 1994, # stocks = 1,138, # factors = 44

Model	**# Factors**	**Posterior probability**
mkt, ear	2	1.00

Factors with marginal posterior probability > 0: mkt, ear
Factors with marginal posterior probability > 0: mkt, ear

Panel C. January 1995 - June 2002, # stocks = 1,289, # factors = 48

Model	**# Factors**	**Posterior probability**
aeavol	1	0.48
aeavol, chnanalyst	2	0.40
mkt, aeavol	2	0.04
mkt, aeavol, chnanalyst	3	0.04

Factors with marginal posterior probability > 0: mkt, aeavol, chnanalyst,
Factors with marginal posterior probability > 0: aeavol

Panel D. July 2002 - December 2009, # stocks = 1,119, # factors = 48

Model	**# Factors**	**Posterior probability**
mkt, pchsale_pchrect	2	0.44
mkt	1	0.40
mkt, sue	2	0.12
mkt, pchsale_pchrect, sue	3	0.04

Factors with marginal posterior probability > 0: mkt, pchsale_pchrect, sue
Factors with marginal posterior probability > 0: mkt

Panel E. January 2010 - December 2016, # stocks = 1,058, # factors = 49

Model	**# Factors**	**Posterior probability**
mkt	1	1.00

Factors with marginal posterior probability > 0: mkt
Factors with marginal posterior probability > 0: mkt

Table 8: Summary of best models identified using different sets of portfolios, 1980-2016, $c = 5$

We apply the Bayesian variable selection methodology to sets of portfolios formed according to various criteria, using factors for which data is available for the period 1980 to 2016. The table reports the best model, *i.e.* the model with highest posterior probability, the number of factors in the model, and the posterior probability. We report results for $c = 5$, where c is a scaling parameter related to the prior variance of the regression coefficient vector $\boldsymbol{\beta}$.

Portfolio formation	**# Portfolios**	**Best model**	**# Factors**	**Probability**
Univariate Sorts				
Size	10	mkt,ill,mve,mve_ia	4	0.30
Book-to-market	10	mkt,lev	2	0.80
Operating profitability	10	mkt,roic	2	0.60
Investment	10	mkt,agr,roavol	3	0.22
Earnings-to-price	10	mkt,chcsho,ep	3	0.20
Cashflow-to-price	10	mkt,bm,roavol	3	0.32
Dividend Yield	10	mkt,absacc,cashpr	3	0.22
Momentum	10	mkt,absacc,mom12m	3	0.26
Short-term reversal	10	mkt,mom1m	2	0.34
Long-term reversal	10	mkt,chcsho,mom36m	3	0.22
Beta	10	mkt,beta	2	0.75
Variance	10	mkt,idiovol	2	0.70
Residual variance	10	mkt,beta,idiovol,retvol	4	0.20
Bivariate Sorts				
Size and book-to-market	25	mkt,ill,lev,mve_ia	4	0.40
Size and operating profitability	25	mkt,beta,ill,roic,std_dolvol	5	0.20
Size and investing	25	mkt,beta,ill,mve_ia,std_dolvol	5	0.60
Book-to-market and operating profitability	25	mkt,bm,idiovol,lev,roavol,std_turn	6	0.40
Book-to-market and investment	25	mkt,bm,lev	3	0.80
Operating profitability and investment	25	mkt,beta,roic,tang	4	0.40
Industry Portfolios				
Industries	49	mkt,beta,ill,lev,orgcap	5	0.40

Appendix A. Bayesian Estimation of the SUR Model

This section details the estimation of SUR model using the Gibbs sampler. The SUR model with common regressors can be written as

$$\mathbf{r}_i = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{e}_i, \qquad i = 1, \ldots, N \tag{A.1}$$

where, for each equation $i = 1, \ldots, N$, $\mathbf{r}_i$ is the $T \times 1$ vector of observed responses, $\mathbf{X}$ is the matrix of regressors with dimension $T \times K$, $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,K})'$ is a $K \times 1$ vector of unknown regression coefficients and $\mathbf{e_i}$ is a $T \times 1$ vector of disturbances. The system can be stacked in a single equation $\tilde{\mathbf{r}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}}$ in the following way (see *e.g.* Greene (2003)):

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_N \end{bmatrix} \tag{A.2}$$

Letting $\tilde{\mathbf{e}} = (\mathbf{e}_1' \quad \mathbf{e}_2' \quad \ldots \quad \mathbf{e}_N')'$, the basic assumption of the SUR model is $\mathbb{E}(\tilde{\mathbf{e}}\tilde{\mathbf{e}}') = \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I_T}$. We assume $\tilde{\mathbf{e}} \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I_T})$ and the following prior distributions for $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}$:

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &\sim N(\mathbf{b}_0, \mathbf{B}_0) \\ \boldsymbol{\Sigma} &\sim IW(\nu_0, \boldsymbol{\Phi}_0). \end{aligned} \tag{A.3}$$

The likelihood for the full system of equations is given by

$$L(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\mathbf{X}, \mathbf{r}) = (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Omega}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})\right). \tag{A.4}$$

Let $\boldsymbol{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$ and define $\tilde{\mathbf{X}}^* = \mathbf{P}\mathbf{X}$, $\tilde{\mathbf{r}}^* = \mathbf{P}\tilde{\mathbf{r}}$. Then $\tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}} = \tilde{\mathbf{X}}^{*\prime}\tilde{\mathbf{X}}^*$ and $\tilde{\mathbf{X}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{r}} = \tilde{\mathbf{X}}^{*\prime}\tilde{\mathbf{r}}$ and we can write

$$L(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\mathbf{X}, \mathbf{r}) \propto \exp\left(-\frac{1}{2}(\tilde{\mathbf{r}}^* - \tilde{\mathbf{X}}^*\tilde{\boldsymbol{\beta}})'(\tilde{\mathbf{r}}^* - \tilde{\mathbf{X}}^*\tilde{\boldsymbol{\beta}})\right). \tag{A.5}$$

Appendix A.1. Distribution of $\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \tilde{\mathbf{r}}$

We will use the notation $f(\cdot)$ to denote a generic probability density function, and $f(\cdot|\cdot)$ to denote a conditional density. The prior for $\tilde{\boldsymbol{\beta}}$ is given by

$$f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}) \propto \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{\beta}} - \mathbf{b}_0)'\mathbf{B}_0^{-1}(\tilde{\boldsymbol{\beta}} - \mathbf{b}_0)\right), \tag{A.6}$$

where $\mathbf{b}_0, \mathbf{B}_0$ are known. Therefore the posterior conditional distribution of $\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \tilde{\mathbf{r}}$ is

$$\begin{aligned} f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \mathbf{r}) &\propto f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}) L(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}|\mathbf{X}, \tilde{\mathbf{r}}) \\ &\propto \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{\beta}} - \mathbf{b}_0)' \mathbf{B}_0^{-1} (\tilde{\boldsymbol{\beta}} - \mathbf{b}_0)\right) \exp\left(-\frac{1}{2}(\tilde{\mathbf{r}}^* - \tilde{\mathbf{X}}^* \tilde{\boldsymbol{\beta}})'(\tilde{\mathbf{r}}^* - \tilde{\mathbf{X}}^* \tilde{\boldsymbol{\beta}})\right). \end{aligned}$$

Expanding the products and collecting terms on $\boldsymbol{\beta}$, we have

$$f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \mathbf{r}) \propto \exp\left(-\frac{1}{2}\left[\tilde{\boldsymbol{\beta}}'(\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}^{*\prime}\tilde{\mathbf{X}}^*)\tilde{\boldsymbol{\beta}} - 2\tilde{\boldsymbol{\beta}}'(\mathbf{B}_0^{-1}\mathbf{b}_0 + \tilde{\mathbf{X}}^{*\prime}\tilde{\mathbf{r}}^*)\right]\right).$$

Letting $\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}^{*\prime}\tilde{\mathbf{X}}^*)^{-1}$ we obtain

$$f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \mathbf{r}) \propto \exp\left(-\frac{1}{2}\left[\tilde{\boldsymbol{\beta}}'\mathbf{B}_1^{-1}\tilde{\boldsymbol{\beta}} - 2\tilde{\boldsymbol{\beta}}'(\mathbf{B}_0^{-1}\mathbf{b}_0 + \tilde{\mathbf{X}}^{*\prime}\tilde{\mathbf{r}}^*)\right]\right).$$

Finally, completing the quadratic form and letting

$$\mathbf{b}_1 = (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}})^{-1}(\mathbf{B}_0\mathbf{b}_0 + \tilde{\mathbf{X}}'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{r}}),$$

we obtain

$$f(\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \mathbf{r}) \propto \exp\left(-\frac{1}{2}\left[\left(\tilde{\boldsymbol{\beta}} - \mathbf{b}_1\right)' \mathbf{B}_1^{-1} \left(\tilde{\boldsymbol{\beta}} - \mathbf{b}_1\right)\right]\right),$$

therefore recognizing that $\tilde{\boldsymbol{\beta}}|\boldsymbol{\Omega}, \mathbf{r} \sim N(\mathbf{b}_1, \mathbf{B}_1)$.

Appendix A.2. Sequential generation of $\boldsymbol{\beta}_i|\tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}}$

Recall that $\tilde{\mathbf{X}}$ has dimension $NT \times NK$ and $\boldsymbol{\Omega}$ has dimension $NT \times NT$. Therefore, for large panels (when N is large), the expressions above will require multiplication and inversion of large matrices. An alternative and quicker approach for large panels consists of sampling each $\boldsymbol{\beta}_i$ conditionally on the remaining $\boldsymbol{\beta}_j$, $j \neq i$ and $\boldsymbol{\Sigma}$. Let $\tilde{\boldsymbol{\beta}}_{-i}$ denote the full vector $\tilde{\boldsymbol{\beta}}$ with the entries corresponding to i removed. Assume that

$$\boldsymbol{\beta}_i|\tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\Sigma} \sim N(\mathbf{b}_{0,i}, \mathbf{B}_{0,i}), \qquad i = 1, \ldots, N.$$

For simplicity, let's assume that $i = 1$, that is, we are interested in generating $\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \boldsymbol{\Sigma}$. Partition the SUR system as follows:

$$\tilde{\mathbf{r}} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_{-1} \end{bmatrix}, \tilde{\boldsymbol{\beta}} = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \tilde{\boldsymbol{\beta}}_{-1} \end{bmatrix}, \tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{X}}_{-1} \end{bmatrix},$$

where $\tilde{\mathbf{X}}_{-1}$ collects the structure of $\tilde{\mathbf{X}}$ for the remaining $N-1$ equations. Then we can write

$$\begin{aligned}\tilde{\mathbf{r}} - \mathbf{X}\tilde{\boldsymbol{\beta}} &= \begin{bmatrix}\mathbf{r}_1 \\ \mathbf{r}_{-1}\end{bmatrix} - \begin{bmatrix}\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{X}}_{-1}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\beta}}_1 \\ \tilde{\boldsymbol{\beta}}_{-1}\end{bmatrix} \\ &= \begin{bmatrix}\tilde{\mathbf{r}}_1 - \mathbf{X}\tilde{\boldsymbol{\beta}}_1 \\ \tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1}\end{bmatrix} \qquad \text{(A.7)}\end{aligned}$$

Recall that $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T$ and let $\{\boldsymbol{\Sigma}^{-1}\}_{ij} = \sigma^{ij}$ denote element (i, j) of $\boldsymbol{\Sigma}^{-1}$. The corresponding partition of $\boldsymbol{\Omega}^{-1}$ is

$$\boldsymbol{\Omega}^{-1} = \left[\begin{array}{c|ccc}\sigma^{11}\mathbf{I} & \sigma^{12}\mathbf{I} & \cdots & \sigma^{1N}\mathbf{I} \\ \hline \sigma^{21}\mathbf{I} & \sigma^{22}\mathbf{I} & \cdots & \sigma^{2N}\mathbf{I} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma^{N1}\mathbf{I} & \sigma^{N2} & \cdots & \sigma^{NN}\mathbf{I}\end{array}\right] = \begin{bmatrix}\sigma^{11}\mathbf{I} & \mathbf{A} \\ \mathbf{A}' & \boldsymbol{\Omega}_{-1}^{-1}\end{bmatrix}. \qquad \text{(A.8)}$$

In the partition of $\boldsymbol{\Omega}^{-1}$ above, we note that $\sigma^{11}\mathbf{I}$ has dimension $T \times T$, $\mathbf{A}$ has dimension $T \times (N-1)T$, and $\boldsymbol{\Omega}_{-1}^{-1}$ has dimension $(N-1)T \times (N-1)T$. Using A.7 and A.8 we can now write the weighted sum of residuals as follows.

$$(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}) =$$

$$\begin{bmatrix}(\mathbf{r}_1 - \mathbf{X}\boldsymbol{\beta}_1)' & (\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1})'\end{bmatrix}\begin{bmatrix}\sigma^{11}\mathbf{I} & \mathbf{A} \\ \mathbf{A}' & \boldsymbol{\Omega}_{-1}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{r}_1 - \mathbf{X}\boldsymbol{\beta}_1 \\ \tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1}\end{bmatrix}$$

Expanding the right-hand side and collecting terms, we obtain

$$\begin{aligned}&(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}) = \\ &\sigma^{11}(\mathbf{r}_1 - \mathbf{X}\tilde{\boldsymbol{\beta}}_1)'(\mathbf{r}_1 - \mathbf{X}\tilde{\boldsymbol{\beta}}_1) + 2(\mathbf{r}_1 - \mathbf{X}\boldsymbol{\beta}_1)'\mathbf{A}(\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1}) \qquad \text{(A.9)} \\ &+ (\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1})'\boldsymbol{\Omega}_{-1}^{-1}(\mathbf{r}_1 - \mathbf{X}\boldsymbol{\beta}_1)\end{aligned}$$

Now the posterior of $\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}}$ can be calculated as

$$\begin{aligned} f(\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}}) &\propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_1 - \mathbf{b}_{0,1})'\mathbf{B}_{0,1}^{-1}(\boldsymbol{\beta}_1 - \mathbf{b}_{0,1})\right) \\ &\quad \times \exp\left(-\frac{1}{2}(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\tilde{\mathbf{r}} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})\right) \\ &\propto \exp\Bigg(-\frac{1}{2}\Big[(\boldsymbol{\beta}_1 - \mathbf{b}_{0,1})'\mathbf{B}_{0,1}^{-1}(\boldsymbol{\beta}_1 - \mathbf{b}_{0,1}) \\ &\quad + \sigma^{11}(\mathbf{r}_1 - \mathbf{X}\tilde{\boldsymbol{\beta}}_1)'(\mathbf{r}_1 - \mathbf{X}\tilde{\boldsymbol{\beta}}_1) + 2(\mathbf{r}_1 - \mathbf{X}\boldsymbol{\beta}_1)'\mathbf{A}(\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1}) \\ &\quad + (\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1})'\boldsymbol{\Omega}_{-1}^{-1}(\mathbf{r}_1 - \mathbf{X}\boldsymbol{\beta}_1)\Big]\Bigg), \end{aligned}$$

where we have substituted (A.9). Expanding the expression above and removing terms that are constant or do not depend on $\boldsymbol{\beta}_1$ yields:

$$\begin{aligned} f(\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}}) &\propto \exp\Bigg(-\frac{1}{2}\Big[\boldsymbol{\beta}_1'\mathbf{B}_{0,1}^{-1}\boldsymbol{\beta}_1 - 2\boldsymbol{\beta}_1'\mathbf{B}_{0,1}^{-1}\mathbf{b}_{0,1} + \sigma^{11}(\boldsymbol{\beta}_1'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_1 \\ &- 2\mathbf{r}_1'\mathbf{X}\boldsymbol{\beta}_1) - 2\boldsymbol{\beta}_1\mathbf{X}'\mathbf{A}(\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1})\Big]\Bigg) \\ &\propto \exp\Bigg(-\frac{1}{2}\Big[\boldsymbol{\beta}_1'(\mathbf{B}_{0,1}^{-1} + \sigma^{11}\mathbf{X}'\mathbf{X})\boldsymbol{\beta}_1 \\ &- 2\boldsymbol{\beta}_1'(\mathbf{B}_{0,1}^{-1}\mathbf{b}_{0,1} + \sigma^{11}\mathbf{X}'(\mathbf{r}_1 - (\sigma^{11})^{-1}\mathbf{A}(\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1})))\Big]\Bigg) \end{aligned}$$

Now letting:

$$\begin{aligned} \mathbf{r}_1^* &= \mathbf{r}_1 - (\sigma^{11})^{-1}\mathbf{A}(\tilde{\mathbf{r}}_{-1} - \tilde{\mathbf{X}}_{-1}\tilde{\boldsymbol{\beta}}_{-1}) \\ \mathbf{B}_{1,1} &= (\mathbf{B}_{0,1}^{-1} + \sigma^{11}\mathbf{X}'\mathbf{X})^{-1} \\ \mathbf{b}_{1,1} &= (\mathbf{B}_{0,1}^{-1} + \sigma^{11}\mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_{0,1}^{-1}\mathbf{b}_{0,1} + \sigma^{11}\mathbf{X}'\mathbf{r}_1^*) \end{aligned}$$

and completing the squares, we obtain

$$f(\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}}) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}_1' - \mathbf{b}_{1,1}\right)'\mathbf{B}_{1,1}^{-1}\left(\boldsymbol{\beta}_1' - \mathbf{b}_{1,1}\right)\right),$$

therefore establishing $\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}} \sim N(\mathbf{b}_{1,1}, \mathbf{B}_{1,1})$. More generally, we could have placed any of the equations in the first position in our partition, so it follows that $\boldsymbol{\beta}_i|\tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}} \sim N(\mathbf{b}_{1,i}, \mathbf{B}_{1,i})$, with

$$\begin{aligned} \mathbf{r}_i^* &= \mathbf{r}_i - (\sigma^{ii})^{-1}\mathbf{A}(\tilde{\mathbf{r}}_{-i} - \tilde{\mathbf{X}}_{-i}\tilde{\boldsymbol{\beta}}_{-i}) \\ \mathbf{B}_{1,i} &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}'\mathbf{X})^{-1} \\ \mathbf{b}_{1,i} &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_{0,i}^{-1}\mathbf{b}_{0,i} + \sigma^{ii}\mathbf{X}'\mathbf{r}_i^*), \end{aligned}$$

where $\mathbf{A}$ now is defined appropriately to contain the terms for $j \neq i$.

Note that $\mathbf{r}_i^*$ is the vector of responses for equation i, subtracted from a weighted average of the residuals from the remaining $N-1$ equations, where the weights depend on the elements of $\mathbf{\Sigma}^{-1}$. Thus, the posterior variance of $\boldsymbol{\beta}_1|\tilde{\boldsymbol{\beta}}_{-1}, \mathbf{\Sigma}, \tilde{\mathbf{r}}$ depends on the covariance of the residuals of the equations. If these are zero, that is, if the system is compose of actually unrelated regressions, then $\mathbf{r}_1^* = \mathbf{r}_1$ and the posterior covariance matrix reduces to the one that would be obtained for the single regression equation i, as one would expect.

Appendix A.3. Distribution of $\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}$

Since $\mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I}_T$, it suffices to derive the conditional distribution of $\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}$. The prior for $\mathbf{\Sigma}$ is an inverted Whishart distribution with parameters ν_0 and $\mathbf{\Phi}_0$:

$$f(\mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-\frac{\nu_0+N+1}{2}} \exp\left(-\frac{1}{2}\,\mathrm{Tr}\left(\mathbf{\Phi}_0\mathbf{\Sigma}^{-1}\right)\right).$$

To derive the posterior of $\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}$, it is convenient to write the likelihood function in a different way, by arranging the system such that, instead of stacking all T observations for each equation, we will stack the N equations for each observation. For an arbitrary observation t, let $\mathbf{r}_t = (y_{t,1}, y_{t,2}, \ldots, y_{t,N})'$ denote the $N \times 1$ vector of observed responses, $\mathbf{x}_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,K})'$ denote the $K \times 1$ vector of predictors, and $\mathbf{e}_t = (e_{t,1}, e_{t,2}, \ldots, e_{t,N})'$ denote the vector of error terms. Then we can write

$$\mathbf{r}_t' = \mathbf{x}_t'\begin{bmatrix}\boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_N\end{bmatrix} + \mathbf{e}_t', \qquad t = 1, \ldots, T. \tag{A.10}$$

The SUR correlation structure now can be represented conveniently as $\mathbb{E}(\mathbf{e}_t\mathbf{e}_t') = \mathbf{\Sigma}$. The likelihood at each observation is $L_t = (2\pi)^{-\frac{N}{2}}\,|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{e}_t'\mathbf{\Sigma}^{-1}\mathbf{e}_t\right)$ and the full likelihood can be written as

$$\begin{aligned} L &= \prod_{t=1}^{T} L_t = (2\pi)^{-\frac{NT}{2}}\,|\mathbf{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\mathbf{e}_t'\mathbf{\Sigma}^{-1}\mathbf{e}_t\right) \\ &\propto |\mathbf{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\,\mathrm{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{S}\right)\right), \end{aligned} \tag{A.11}$$

where $\mathbf{S} = \sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t'$ and we have used the fact that $\mathbf{e}_t'\mathbf{\Sigma}^{-1}\mathbf{e}_t$ is a scalar (thus equal to its trace), and the properties of the trace operator.

We can now write the conditional distribution $\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}$ as follows:

$$\begin{aligned} f(\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}) &\propto f(\mathbf{\Sigma})L(\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}) \\ &\propto |\mathbf{\Sigma}|^{-\frac{\nu_0+N+1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(\mathbf{\Phi}_0\mathbf{\Sigma}^{-1}\right)\right) \times |\mathbf{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(\mathbf{\Sigma}^{-1}\mathbf{S}\right)\right) \\ &\propto |\mathbf{\Sigma}|^{-\frac{\nu_0+N+T+1}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(\mathbf{\Sigma}^{-1}(\mathbf{\Phi}_0+\mathbf{S})\right)\right), \end{aligned}$$

which establishes $\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}} \sim IW(\nu_0 + T, \mathbf{\Phi}_0 + \mathbf{S})$.

Appendix B. Bayesian Variable Selection in SUR

This section derives the conditional distributions required for our variable selection methodology using the Gibbs sampler. Let $\mathbf{X}_\gamma$ represent the matrix $\mathbf{X}$ where each column has been multiplied by the corresponding γ_j. Then we can write the model with variable selection as $\mathbf{r}_i = \mathbf{X}_\gamma\boldsymbol{\beta}_i + \mathbf{e}_i, i = 1, \ldots, N$. Stacking the N equations, we can also represent the model as:

$$\begin{bmatrix}\mathbf{r}_1\\ \mathbf{r}_2\\ \vdots\\ \mathbf{r}_N\end{bmatrix} = \begin{bmatrix}\mathbf{X}_\gamma & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_\gamma & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_\gamma\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\\ \vdots\\ \boldsymbol{\beta}_N\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \vdots\\ \mathbf{e}_N\end{bmatrix}$$

or

$$\tilde{\mathbf{r}} = \tilde{\mathbf{X}}_\gamma\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{e}}. \tag{B.1}$$

Note that, conditional on $\boldsymbol{\gamma}$, the model reduces to a SUR with the corresponding predictors for which $\gamma_j = 1$. Therefore, we can use the results derived in the previous section for $\tilde{\boldsymbol{\beta}}|\mathbf{\Sigma}, \boldsymbol{\gamma}, \tilde{\mathbf{r}}$ and $\mathbf{\Sigma}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \tilde{\mathbf{r}}$, substituting $\tilde{\mathbf{X}}$ by $\tilde{\mathbf{X}}_\gamma$.

Appendix B.1. Distribution of $\tilde{\boldsymbol{\beta}}|\mathbf{\Sigma}, \boldsymbol{\gamma}, \tilde{\mathbf{r}}$

Using the results from the previous section, treating $\mathbf{\Sigma}$ and $\boldsymbol{\gamma}$ as known, the posterior distribution of $\tilde{\boldsymbol{\beta}}|\mathbf{\Sigma}, \boldsymbol{\gamma}, \tilde{\mathbf{r}}$ is $N(\mathbf{b}_1, \mathbf{B}_1)$, where

$$\begin{aligned} \mathbf{b}_1 &= (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}_\gamma'\mathbf{\Omega}^{-1}\tilde{\mathbf{X}}_\gamma)^{-1}(\mathbf{B}_0\mathbf{b}_0 + \tilde{\mathbf{X}}_\gamma'\mathbf{\Omega}^{-1}\tilde{\mathbf{r}}), \\ \mathbf{B}_1 &= (\mathbf{B}_0^{-1} + \tilde{\mathbf{X}}_\gamma'\mathbf{\Omega}^{-1}\tilde{\mathbf{X}}_\gamma)^{-1}. \end{aligned}$$

We can also use the sequential generation of $\boldsymbol{\beta}_i, i = 1, \ldots, N$ as in Section Appendix A.2. In this case, we rewrite the partition in equation A.7 in terms of

$\tilde{\mathbf{X}}_\gamma$ and define $\tilde{\mathbf{X}}_{\gamma,-i}$ as the matrix that collects the structure of $\tilde{\mathbf{X}}_\gamma$ for the remaining $N-1$ equations. Then, assuming $\boldsymbol{\gamma}$ known, we have $\boldsymbol{\beta}_i|\tilde{\boldsymbol{\beta}}_{-i}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}, \tilde{\mathbf{r}} \sim N(\mathbf{b}_{1,i}, \mathbf{B}_{1,i})$, with

$$\begin{aligned}
\mathbf{r}_i^* &= \mathbf{r}_i - (\sigma^{ii})^{-1}\mathbf{A}(\tilde{\mathbf{r}}_{-i} - \tilde{\mathbf{X}}_{\gamma,-i}\tilde{\boldsymbol{\beta}}_{-i}) \\
\mathbf{B}_{1,i} &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1} \\
\mathbf{b}_{1,i} &= (\mathbf{B}_{0,i}^{-1} + \sigma^{ii}\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}(\mathbf{B}_{0,i}^{-1}\mathbf{b}_{0,i} + \sigma^{ii}\mathbf{X}'\mathbf{r}_i^*).
\end{aligned}$$

Appendix B.2. Distribution of $\boldsymbol{\Sigma}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \tilde{\mathbf{r}}$

Using the results from the previous section, treating $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\gamma}$ as known, we have $\boldsymbol{\Sigma}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \tilde{\mathbf{r}} \sim IW(\nu_0 + T, \boldsymbol{\Phi}_0 + \mathbf{S}_\gamma)$, where $\mathbf{S}_\gamma$ is calculated using the residuals from equation B.1.

Appendix B.3. Distribution of $\boldsymbol{\gamma}|\boldsymbol{\Sigma}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}$

The simplest approach to generate $\boldsymbol{\gamma}|\boldsymbol{\Sigma}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}$ is to use the Gibbs sampler to generate each value of $\boldsymbol{\gamma}$ component-wise, that is, we can generate each γ_j, conditionally on the remaining $\gamma_i, i \neq j$, which we denote as $\boldsymbol{\gamma}_{-j}$, $\boldsymbol{\Sigma}$, and $\tilde{\boldsymbol{\beta}}$. For a given j, denote by $L_{j,1} = L(\gamma_j = 1|\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}})$ the likelihood function evaluated at $\gamma_j = 1$, considering $\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{\beta}}$ known, and likewise by $L_{j,0} = L(\gamma_j = 0|\boldsymbol{\gamma}_{-j}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}})$ the likelihood evaluated at $\gamma_j = 0$. Then, using the fact that the prior distribution of the γ_j is $B(1, \pi_j)$, $j = 1, \ldots, N$, we have

$$P(\gamma_j = 1|\gamma_{-j}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}) = \frac{\pi_j L_{j,1}}{\pi_j L_{j,1} + (1-\pi_j)L_{j,0}}. \tag{B.2}$$

Let $\boldsymbol{\gamma}_j^1$ and $\boldsymbol{\gamma}_j^0$ represent the vector $\boldsymbol{\gamma}$ with the $j-th$ position fixed at 1 or 0, respectively. That is,

$$\begin{aligned}
\boldsymbol{\gamma}_j^1 &= [\gamma_1, \cdots, \gamma_{j-1}, 1, \gamma_{j+1} \cdots \gamma_K]', \\
\boldsymbol{\gamma}_j^0 &= [\gamma_1, \cdots, \gamma_{j-1}, 0, \gamma_{j+1} \cdots \gamma_K]'.
\end{aligned}$$

Further, let $\mathbf{e}_t^1$ and $\mathbf{e}_t^0$ represent the residuals, at observation t, if $\gamma_j = 1$ and if $\gamma_j = 0$, respectively. Let $\mathbf{S}_\gamma^1$ and $\mathbf{S}_\gamma^0$ represent the corresponding residual matrices. Then we can write, using A.11:

$$\begin{aligned}
L_{j,1} &= (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_\gamma^1)\right) \\
L_{j,0} &= (2\pi)^{-\frac{NT}{2}} |\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_\gamma^0)\right)
\end{aligned}$$

Substituting the above into B.2, we get

$$\begin{aligned} P(\gamma_j = 1|\gamma_{-j}, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{r}}) &= \frac{\pi_j \exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_\gamma^1)\right)}{\pi_j \exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_\gamma^1)\right) + (1-\pi_j)\exp\left(-\frac{1}{2}\operatorname{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_\gamma^0)\right)} \\ &= \left(1 + \frac{1-\pi_j}{\pi_j}\exp\left[-\frac{1}{2}\operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1}(\mathbf{S}_\gamma^1 - \mathbf{S}_\gamma^0)\right)\right]\right)^{-1}, \qquad \text{(B.3)} \end{aligned}$$

where we have taken the inverse of the expression on the right-hand side twice.

Appendix C. Factor Construction

[Table 9 about here.]

Table C.9: The Factor Zoo: candidate factors/firm characteristics

The table lists the 82 firm characteristics used to construct tradable factors.

Acronym	**Firm Characteristic/Factor**	**Reference**
mkt	Market return	**?**
absacc	Absolute accruals	Bandyopadhyay *et al.* (2010)
acc	Working capital accruals	Sloan (1996)
aeavol	Abnormal earnings announcement volume	Lerman *et al.* (2008)
age	# years since first Compustat coverage	Jiang *et al.* (2005)
agr	Asset growth	Cooper *et al.* (2008)
baspread	Bid-ask spread	Amihud & Mendelson (1989)
beta	Beta	Fama & MacBeth (1973)
bm	Book-to-market	Rosenberg *et al.* (1985)
bm_ia	Industry-adjusted book to market	Asness *et al.* (2000)
cash	Cash holdings	Palazzo (2012)
cashdebt	Cash flow to debt	Ou & Penman (1989)
cashpr	Cash productivity	Chandrashekar & Rao (2009)
cfp	Cash-flow-to-price ratio	Desai *et al.* (2004)
cfp_ia	Industry-adjusted cash-flow-to-price ratio	Asness *et al.* (2000)
chatoia	Industry-adjusted change in asset turnover	Soliman (2008)
chcsho	Change in shares outstanding	Pontiff & Woodgate (2008)
chempia	Industry-adjusted change in employees	Asness *et al.* (2000)
chfeps	Change in forecasted EPS	Hawkins *et al.* (1984)
chinv	Change in inventory	Thomas & Zhang (2002)
chmom	Change in 6-month momentum	Gettleman & Marks (2006)
chnanalyst	Change in number of analysts	Scherbina (2008)
chpmia	Industry-adjusted change in profit margin	Soliman (2008)
chtx	Change in tax expense	Thomas & Zhang (2002)
cinvest	Corporate investment	Titman *et al.* (2004)
currat	Current ratio	Ou & Penman (1989)
depr	Depreciation / PP&E	Holthausen & Larcker (1992)
disp	Dispersion in forecasted EPS	Diether *et al.* (2002)
ear	Earnings announcement return	Brandt *et al.* (2008)
egr	Growth in common shareholder equity	Richardson *et al.* (2005)
ep	Earnings to price	Basu (1977)
fgr5yr	Forecasted growth in 5-year EPS	Bauman & Dowen (1988)
gma	Gross profitability	Novy-Marx (2013)

Table C.9: (continued)

Acronym	Firm Characteristic/Factor	Reference
grcapx	Growth in capital expenditures	Anderson & Garcia-Feijóo (2006)
grltnoa	Growth in long-term net operating assets	Fairfield *et al.* (2003)
herf	Industry sales concentration	Hou & Robinson (2006)
hire	Employee growth rate	Belo *et al.* (2014)
idiovol	Idiosyncratic return volatility	Ali *et al.* (2003)
ill	Illiquidity	Amihud (2002)
indmom	Industry momentum	Moskowitz & Grinblatt (1999)
invest	Capital expenditures and inventory	Chen & Zhang (2010)
lev	Leverage	Bhandari (1988)
mom12m	12-month momentum	Jegadeesh (1990)
mom1m	1-month momentum	Jegadeesh & Titman (1993)
mom36m	36-month momentum	Jegadeesh & Titman (1993)
ms	Financial statement score	Mohanram (2005b)
mve	Size	Banz (1981)
mve_ia	Industry-adjusted size	Asness *et al.* (2000)
nanalyst	Number of analysts covering stock	Elgers *et al.* (2001)
operprof	Operating profitability	Fama & French (2015)
orgcap	Organizational capital	Eisfeldt & Papanikolaou (2013)
pchcapx_ia	Industry adjusted change in capex	Abarbanell & Bushee (1998)
pchcurrat	change in current ratio	Ou & Penman (1989)
pchdepr	change in depreciation	Holthausen & Larcker (1992)
pchgm_pchsale	change in gross margin - change in sales	Abarbanell & Bushee (1998)
pchsaleinv	change sales-to-inventory	Ou & Penman (1989)
pchsale_pchinvt	change in sales - change in inventory	Abarbanell & Bushee (1998)
pchsale_pchrect	change in sales - change in A/R	Abarbanell & Bushee (1998)
pchsale_pchxsga	change in sales - change in SG&A	Abarbanell & Bushee (1998)
pctacc	Percent accruals	Hafzalla *et al.* (2011)
pricedelay	Price delay	Hou & Moskowitz (2005)
ps	Financial statements score	Piotroski (2000)
realestate	Real estate holdings	Tuzel (2010)
retvol	Return volatility	Ang *et al.* (2006)
roaq	Return on assets	Balakrishnan *et al.* (2010)
roavol	Earnings volatility	Francis *et al.* (2004)

Table C.9: (continued)

Acronym	**Firm Characteristic/Factor**	**Reference**
roeq	Return on equity	Hou *et al.* (2015a)
roic	Return on invested capital	Brown & Rowe (2007)
rsup	Revenue surprise	Kama (2009)
salecash	Sales to cash	Ou & Penman (1989)
saleinv	Sales to inventory	Ou & Penman (1989)
salerec	Sales to receivables	Ou & Penman (1989)
sfe	Scaled earnings forecast	Elgers *et al.* (2001)
sgr	Sales growth	Lakonishok *et al.* (1994)
sp	Sales to price	Barbee Jr *et al.* (1996)
stdcf	Cash flow volatility	Huang (2009)
std_dolvol	Volatility of liquidity (dollar trading volume)	Chordia *et al.* (2001)
std_turn	Volatility of liquidity (share turnover)	Chordia *et al.* (2001)
sue	Unexpected quarterly earnings	Rendleman *et al.* (1982)
tang	Debt capacity/firm tangibility	Almeida & Campello (2007)
tb	Tax income to book income	Lev & Nissim (2004)
turn	Share turnover	Datar *et al.* (1998)
zerotrade	Zero trading days	Liu (2006)