Preliminary and Incomplete

Which Factors Matter Jointly? A Bayesian Approach

Siddhartha Chib[1], Francis J. DiTraglia[2] and Irina Pimenova[2]

[1]Olin Business School, Washington University in St. Louis
[2]Department of Economics, University of Pennsylvania

This Version: February 24, 2016

Abstract

In this paper we use Bayesian methods to revisit a classic question in empirical finance: which asset pricing factors explain the time-series and cross-section behavior of asset returns? To answer this question one must consider an extremely large collection of models formed from all possible subsets of the proposed factors. And with so many models under consideration, there is a serious risk of overfitting. We address this challenging high-dimensional model selection problem by calculating and comparing marginal likelihoods in a Bayesian seemingly unrelated regression (SUR) model with multivariate Student-t errors and an objective, training-sample prior. Unlike approaches based on hypothesis testing, marginal likelihood comparisons automatically penalize models based on complexity, avoiding the problem of over-fitting. Our proposed method performs well in a calibrated simulation, selecting the correct model even when we intentionally mis-specify the error distribution. We then consider an application using monthly returns for ten industry portfolios and twelve leading asset pricing factors. Our exhaustive search over 49,152 models, 2^{13} for each of six error specification suggests that, in addition to the usual suspects, profitability and quality are important for explaining the behavior of returns.

Keywords: Asset Pricing, Factor Models, Bayesian Model Selection
JEL Codes: C11, C52, G00, G12

1 Introduction

An abiding question in empirical finance is the following: which factors are useful for explaining the time series and cross-section behavior of equity and portfolio returns? There is by now a vast literature on this topic concerned with the development of possible factors and the empirical evaluation of those factors. Along with the original market factor introduced in Sharpe (1964) and Lintner (1965), an array of new factors has emerged as documented in Harvey et al. (2015). Determining the empirical relevance of these factors is an ongoing statistical challenge and different avenues continue to be energetically explored. For instance, Hou et al. (2014a) compare the Hou et al. (2014b) and Fama and French (2015) five-factor models based on the conceptual meaning of the factors and their ability to explain asset-pricing anomalies. A more statistical evaluation is provided by Harvey and Liu (2015) who start with a collection of 12 leading factors and use a bootstrap procedure to forward-select relevant factors.

We propose a different method for finding the best collection of factors based on two important observations. First, factors should be considered *jointly* in all possible combinations. If D is the number of factors under consideration, this requires the consideration of 2^{D+1} models, allowing for the possibility of a non-zero intercept. Stepwise selection, whether based on a t-statistic or related statistics, ignores the question of joint significance and the potentially high correlation between factors. As such it can perform poorly in practice and does not necessarily select the best model in the limit: it is not asymptotically consistent.[1] Second, with so many models under consideration, there is a serious danger of over-fitting. Indeed, as suggested by Harvey et al. (2015), many of the novel factors identified in recent years may well be spurious. To avoid this problem, any proposed selection procedure should account for the high-dimensional nature of the factor selection problem by appropriately penalizing more complex models relative to simpler ones.

In this paper we take both of these observations to heart, using monthly observations for twelve leading asset pricing factors and ten industry portfolios to carry out an exhaustive Bayesian comparison of 49,152 asset pricing specifications based on a seemingly unrelated regression (SUR) model with Student-t errors. Ten of the factors used in our exercise are used by Harvey and Liu (2015) while two others are drawn from different sources. Our model comparisons are based on the calculation of Bayesian marginal likelihoods, using the method of Chib (1995). Marginal likelihoods automat-

[1]See, for example Judd et al. (2011).

ically penalize models based on complexity, assuring that those with more factors will not rank higher merely because their greater flexibility allows them to fit the noise in the data. Moreover, model selection based on the comparison of marginal likelihoods has attractive asymptotic properties. If the true model is among the candidates under consideration, marginal likelihood comparisons will select it with probability approaching one in the limit; if it is not among the candidates, they will select the model that is closest to the truth. As we show in a calibrated simulation example, marginal likelihood comparisons also perform well in finite samples. The flexibility of marginal likelihoods as a tool for model comparison allows us to to simultaneously select over asset pricing factors and features of the error distribution. Thus we allow the data to dictate the heaviness of the tails of the return distribution, considering multivariate Student-t distributions with 4, 6, 8, 10, and 12 degrees of freedom in addition to Gaussian errors. Marginal likelihood comparisons also allow us to "test" the factor pricing models under consideration. Because all of our candidate factors are returns, factor pricing theory implies that the intercept should be zero for all test portfolios. By comparing identical specifications with and without an intercept, we can determine whether the data supports this prediction of theory. Unlike the p-value from a frequentist hypothesis test, which is a probability calculated *under* the assumption that the theory is correct, our Bayesian approach can be used to determine how much more probable it is *that* the theory is correct than that it is not.

Marginal likelihood comparisons require the specification of proper priors. Because the meaning and importance of factor coefficients vary by model, however, it would be inappropriate to place priors independently on each factor. The same is true for the error precision matrix, which we estimate without restriction from the data rather than, say, restricting to be diagonal: its meaning, too, depends on the model. To specify 49,152 priors in an automatic and objective way, we employ a small training sample that precedes our estimation sample.

Our paper adds to the growing literature on the use of Bayesian techniques in finance. Avramov (2002) and Cremers (2002), for example, use Bayesian model-averaging to explore the question of market-return predictability with multiple predictors, while Shanken (1987), Harvey and Zhou (1990) and Avramov and Chao (2006) take a Bayesian approach to consider the question of the significance (or lack thereof) of the intercept in the CAPM context, providing a Bayesian alternative to the frequentist test of this hypothesis developed by Gibbons et al. (1989).

Our main results are as follows. A number of similar factor models find support from the data and are difficult to distinguish empirically: in particular our top four models are practically identical in terms of their marginal likelihoods and differ only slightly in the factors that they contain. All of the top twenty models have a Student-t error distribution with 6 or 8 degrees of freedom and include four of the Fama-French (Fama and French, 1993, 2015) five factors – namely the market, value (HML), profitability (RMW) and investment (CMA) factors – along with momentum (Carhart, 1997) and quality (Asness et al., 2014). In contrast, the liquidity factor of Pastor and Stambaugh (2003) does not appear in any of the top models. Intriguingly, our top model includes *two* profitability factors – the Fama-French RMW factor in addition to ROE, the probability factor suggested by Hou et al. (2014b) – and two size factors: the Fama-French HML factor and an alternative version HMLdev proposed by Asness and Frazzini (2013). While these pairs of factors are closely related, they appear to be capturing enough different information to be useful in concert. Our top-rated model does not include a constant. Comparing its marginal likelihood to that of an otherwise identical specification *with* a constant, we find a difference of approximately 0.86 on the $\log_{10}$ scale. In other words, the model without a constant is just under ten times more plausible *a posteriori* than the model with a constant. Based on these results we would *not* have reason to reject the theoretical prediction of the factor pricing model for our selected model.

The remainder of the paper is organized as follows. Section 2 describes the model with Student-t errors and the estimation framework while Section 3 provides a simulation example to motivate our approach. Section 4 describes our main empirical results and Section 5 concludes.

2 The Model

Consider a linear K-factor model for D assets of the form

$$y_{dt} = \alpha_d + \mathbf{f}_t'\boldsymbol{\beta}_d + \varepsilon_{it}$$

where y_{dt} denotes the excess return for test portfolio d in period t for $d = 1, \ldots, D$ and $t = 1, \ldots, T$ and $\mathbf{f}_t' = (f_{t1}, \ldots, f_{tK})$ vector of factor returns. This is a special case of the seemingly unrelated regression (SUR) model in which the regressors are *identical* across equations. Note that the factors included into the regression $\mathbf{f}$, the number of factors K as well as the presence of the intercept depend on the model under consideration. In order to simplify the notation, we do not introduce a model index, however, it is important to keep in mind that the techniques discussed below are applied to each possible combination of factors.

Stacking observations for a given time period across assets, define $\mathbf{y}_t' = (y_{1t}, \ldots, y_{Dt})$ and analogously $\boldsymbol{\varepsilon}_t' = (\varepsilon_{t1}, \ldots \varepsilon_{tD})$. Both $\mathbf{y}_t$ and $\boldsymbol{\varepsilon}_t$ are $D \times 1$ vectors. Now let $\mathbf{x}_t' = (1, \mathbf{f}_t')$ be a $1 \times (K+1)$ vector and $\boldsymbol{\gamma}_d' = (\alpha_d, \boldsymbol{\beta}_d')$ to be a $1 \times (K+1)$ so we have

$$\mathbf{y}_t = X_t \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_t$$

where $X_t = I_D \otimes \mathbf{x}_t'$, a $D \times D(K+1)$ matrix, and $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_D')$, a $1 \times (K+1)D$ vector. If the model does not include a constant we simply remove the column of ones from the definition of $\mathbf{x}_t$. Let Y_T denote the full data sample, i.e. $\{\mathbf{y}_t, \mathbf{x}_t\}_{t=1}^T$.

In order to apply Bayesian inference, we need to specify the distribution of the errors. Because asset returns are typically thought to be heavy-tailed our models will be based on a multivariate Student-t distribution in which we will allow the data to choose the degrees of freedom.[2] We have:

$$\boldsymbol{\varepsilon}_t \sim t_{D,\nu} \left(0, \Omega\right)$$

where ν denotes the degrees of freedom of the distribution, the location parameter is zero and the scale matrix is Ω. If $\nu > 1$ then $E(\boldsymbol{\varepsilon}) = 0$. If $\nu > 2$ then $Var\left(\boldsymbol{\varepsilon}\right) = \nu\Omega/(\nu - 2)$. The SUR model lets us explicitly explore the cross-sectional dependence of errors by jointly fitting the model for all assets. The possible variance linkages between different assets are captured by the scale matrix Ω. While acknowledging the importance of these linkages and the possible influence they may have on the model selection, we avoid placing *a priori* restrictions on this matrix, such as assuming that it is diagonal, and instead infer the covariance structure of errors from the data.

To permit use of Gibbs sampler draws, we use representation of the Student-t distribution as a scale mixture of normal distributions, as in Chib and Greenberg

[2] Our model selection exercise below will also consider a Gaussian error distribution although this specification is soundly rejected by the data.

(1995). In particular,

$$\begin{aligned}\boldsymbol{\varepsilon}_t|\lambda_t &\sim N\left(0, \lambda_t^{-1}\Omega\right)\\ \lambda_t &\sim G\left(\tfrac{\nu}{2}, \tfrac{\nu}{2}\right)\end{aligned}$$

where $G(\alpha, \beta)$ denotes the Gamma distribution with shape parameter α and rate parameter β. We parameterize this problem in terms of the $D \times D$ *precision* matrix Ω^{-1} and the $p \times 1$ vector of regression coefficients $\boldsymbol{\gamma}$, where $p = D(K+1)$. In any inference that we carry out, as well as in the calculation of the marginal likelihood, we will marginalize over $\boldsymbol{\lambda}$ by simply ignoring these draws. We will place a normal prior on $\boldsymbol{\gamma}$ and a Wishart prior on Ω^{-1}.

2.1 The Gibbs Sampler

In order to approximate the posterior distribution we implement Gibbs sample following Chib and Greenberg (1996) and Carlin and Polson (1991). The sampler proceeds by fixing the degrees of freedom parameter ν. If ν is to be chosen from the data, this can be accomplished using the marginal likelihood, as described below. Holding ν fixed, the full set of conditional posteriors is as follows:

Regression Coefficients: $\boldsymbol{\gamma}|\Omega^{-1}, Y_T \sim \mathcal{N}_p\left(\bar{\boldsymbol{\gamma}}_\lambda, G_{T,\lambda}\right)$

$$\begin{aligned}G_{T,\lambda} &= \left[G_0^{-1} + \sum_{t=1}^{T} \lambda_t X_t' \Omega^{-1} X_t\right]^{-1}\\ \bar{\boldsymbol{\gamma}}_\lambda &= G_{T,\lambda}\left[G_0^{-1}\boldsymbol{\gamma}_0 + \sum_{t=1}^{T} \lambda_t X_t' \Omega^{-1} \mathbf{y}_t\right]\end{aligned}$$

Inverse Scale Matrix: $\Omega^{-1}|Y_T \sim \mathcal{W}_D\left(\rho_0 + T, R_{T,\lambda}\right)$

$$R_{T,\lambda} = \left[R_0^{-1} + \sum_{t=1}^{T} \lambda_t \left(\mathbf{y}_t - X_t\boldsymbol{\gamma}\right)\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}\right)'\right]^{-1}$$

Auxiliary Parameter: $\lambda_t|\boldsymbol{\gamma}, \nu, Y_T \sim G\left(\dfrac{\nu + D}{2}, \dfrac{\nu + \boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t}{2}\right)$, $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - X_t\boldsymbol{\gamma}$

To implement the Gibbs sampler, we simply need to draw sequentially from these distributions, in the order given above. We will require, however, starting values for both Ω^{-1} and each of the λ_t parameters. A reasonable starting value for λ_t is one.

2.2 Marginal Likelihood

We calculate the marginal likelihood using the method of Chib (1995). Consider the re-arranged Bayes rule:

$$f(y) = \frac{f(y|\boldsymbol{\gamma}, \Omega^{-1})\pi\left(\boldsymbol{\gamma}, \Omega^{-1}\right)}{\pi\left(\boldsymbol{\gamma}, \Omega^{-1}|y\right)}$$

which holds for any specified values $(\boldsymbol{\gamma}, \Omega^{-1})$ of the parameters. In particular this holds at the *posterior mean* $(\boldsymbol{\gamma}^*, \Omega^{-1*})$ which is where we will evaluate the expression. Hence, the *log* marginal likelihood is given by

$$\log f(y) = \log\left(\boldsymbol{\gamma}^*, \Omega^{-1*}\right) + \log f(y|\boldsymbol{\gamma}^*, \Omega^{-1*}) - \pi\left(\boldsymbol{\gamma}^*, \Omega^{-1*}|y\right)$$

Since our priors over $\boldsymbol{\gamma}$ and Ω^{-1} are independent, we can re-write the equation above as follows:

$$\log f(Y_T) = \log \pi(\boldsymbol{\gamma}^*) + \log \pi\left(\Omega^{-1*}\right) + \log f\left(Y_T|\boldsymbol{\gamma}^*, \Omega^{-1*}\right) - \log \pi\left(\boldsymbol{\gamma}^*, \Omega^{-1*}|Y_T\right)$$

We can evaluate the marginal likelihood using the Chib (1995) method that approximates $\log f(Y_T)$ by evaluating each of the terms on the right-hand-side of the preceding expression using the output of the Gibbs sampler.

2.3 Priors

In order for the marginal likelihood comparison to be valid, proper priors should be specified.

Given that the factor coefficients should be determined jointly and that factors differ for each model, an individual level prior should be set for every model. The same hold for the error precision matrix: its prior should be specified differently depending on the factors. In order to place model-specific priors, we suggest to use a training sample. We fit each model to the training sample data and use the model-specific posterior means obtained from the Gibbs sampler as a benchmark to specify the prior for the estimation sample. The training sample is employed solely to construct priors and is not a part of the final estimation procedure.

We place two prior distributions: normal on the coefficients and Wishart on the precision matrix. Parameters for the two distributions are specified differently for each stage.

2.3.1 Stage 1: Training Sample

In order to apply Bayesian inference to the training sample, we should specify priors for this estimation step as well.
At this stage diffuse priors are employed in order to reflect the parameter uncertainty.

Regression Coefficients:

$$\begin{aligned} \boldsymbol{\gamma} &\sim \mathcal{N}_p(\boldsymbol{\gamma}_0, G_0) \\ \boldsymbol{\gamma}_0 &= 0 \\ G_0 &= C_1^2 I_p \end{aligned}$$

The prior distribution of the coefficient vector $\boldsymbol{\gamma}$ is centered around zero. The covariance matrix of the coefficients is assumed to be diagonal. The tightness of the prior is controlled by the constant C_1: the larger C_1 is, the wider is the prior.

Inverse Scale Matrix:

$$\begin{aligned} \Omega^{-1} &\sim \mathcal{W}_D(\rho_0, R_0) \\ \rho_0 &= d + C_2 \\ R_0 &= \frac{1}{C_3^2(\rho_0 - d - 1)} I_d \end{aligned}$$

Using the properties of the inverse Wishart distribution, it can be seen that both the mean of the scale matrix Ω and the precision matrix Ω^{-1} implied by the prior are diagonal. The tightness of the prior is governed by the number of degrees of freedom ρ_0. One way of widening the prior would be to decrease number of degrees of freedom by adjusting the value of C_2. Constant C_3 is used to set the magnitude of the elements of the covariance matrix.

For the rest of the paper we set $C_1 = 2$, $C_2 = 6$ and $C_3 = 0.05$.
The posterior draws obtained from the Gibbs sampler when fitting a model to the training sample are later used to form model-specific priors used for the estimation

sample.

2.3.2 Stage 2: Estimation Sample

The first stage draws serve as a basis to construct proper priors for the estimation.

Regression Coefficients:

$$\begin{aligned}
\boldsymbol{\gamma} &\sim \mathcal{N}_p\left(\boldsymbol{\gamma}_0, G_0\right) \\
\boldsymbol{\gamma}_0 &= \overline{\gamma} \\
G_0 &= C_4^2 \widehat{G}
\end{aligned}$$

Denote first stage posterior means of the coefficient vector $\boldsymbol{\gamma}$ as $\overline{\boldsymbol{\gamma}}$ and the sample covariance matrix of these draws as $\widehat{G}$. The prior of factor loadings is thus centered around the posterior mean of the first stage draws. The standard deviation is based on the sample standard deviation of the first stage draws adjusted by the factor of C_4 to reflect uncertainty.

Inverse Scale Matrix:

$$\begin{aligned}
\Omega^{-1} &\sim \mathcal{W}_D\left(\rho_0, R_0\right) \\
\rho_0 &= d + C_5 \\
R_0 &= \frac{1}{\rho}\overline{\Omega}^{-1},
\end{aligned}$$

where $\overline{\Omega}^{-1}$ is the posterior mean of the first stage draws of Ω^{-1}. Note that our algorithm provides us draws of the precision matrix Ω^{-1}, not of the covariance matrix Ω. For this reason we form the prior in terms of the posterior mean of Ω^{-1}. The prior is constructed to set the mean of the precision matrix Ω^{-1} equal to the posterior mean based on the training sample. Again, the tightness of the prior is controlled by degrees of freedom parameter ρ_0.

For all the applications we use $C_4 = 3$ and $C_5 = 6$.

2.4 Training Sample Prior: An Example

Here we illustrate the process of setting model specific priors and show how priors vary by the model depending on the included factors.

For simplicity consider a hypothetical example with the cross-section of returns consisting only of two assets: industry portfolios for durable(Durbl) and non-durable consumer goods (NoDur). We fit two models. First model is a usual CAPM that includes an intercept and a market portfolio(Mkt.RF). An second model is a extended version of the first one that includes a value factor(HML) in addition to the constant and a the market factor.

First, we consider how the factor betas get updated over time:

$$\underbrace{\begin{array}{cc} & \begin{array}{cc} constant & Mkt.RF \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{pmatrix} \end{array}}_{\text{Stage 1 Prior}} \rightarrow \underbrace{\begin{array}{cc} & \begin{array}{cc} constant & Mkt.RF \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.000 & 0.992 \\ -0.001 & 1.046 \end{pmatrix} \end{array}}_{\text{Stage 2 Prior}} \rightarrow$$

$$\rightarrow \underbrace{\begin{array}{cc} & \begin{array}{cc} constant & Mkt.RF \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.004 & 0.793 \\ -0.002 & 1.164 \end{pmatrix} \end{array}}_{\text{Posterior}}$$

This equation demonstrates the update of beliefs about the CAPM coefficients. The first stage prior is centered around zero for all coefficients. The second stage prior based on the training sample reflects the peculiarities of the data. Finally, the posterior is an updated mean based on the estimation sample.

$$\underbrace{\begin{array}{cc} & \begin{array}{ccc} constant & Mkt.RF & HML \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{pmatrix} \end{array}}_{\text{Stage 1 Prior}} \rightarrow \underbrace{\begin{array}{cc} & \begin{array}{ccc} constant & Mkt.RF & HML \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.001 & 0.981 & -0.078 \\ -0.003 & 1.080 & 0.296 \end{pmatrix} \end{array}}_{\text{Stage 2 Prior}} \rightarrow$$

$$\rightarrow \underbrace{\begin{array}{cc} & \begin{array}{ccc} constant & Mkt.RF & HML \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.003 & 0.807 & 0.136 \\ -0.004 & 1.240 & 0.479 \end{pmatrix} \end{array}}_{\text{Posterior}}$$

This equation show the evolution of prior for an extended model. We can see that the second stage prior is different from the one we places on a CAPM model. Even though the market factor is also included into this model as well, the prior mean of the market beta is different for both portfolios because of an added factor.

Similarly, we can take look at how the precision matrix Ω^{-1} is updated. First, consider CAPM:

$$\underset{\text{Stage 1 Prior}}{\begin{array}{cc} & \begin{array}{cc} NoDur & Durbl \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 640 & 0 \\ 0 & 640 \end{pmatrix} \end{array}} \rightarrow \underset{\text{Stage 2 Prior}}{\begin{array}{cc} & \begin{array}{cc} NoDur & Durbl \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 3795 & -277 \\ -277 & 1907 \end{pmatrix} \end{array}} \rightarrow$$

$$\rightarrow \underset{\text{Posterior}}{\begin{array}{cc} & \begin{array}{cc} NoDur & Durbl \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 2283 & 155 \\ 155 & 1013 \end{pmatrix} \end{array}}$$

The following equation shows the changes in the inverse scale matrix for an extended model that includes the value portfolio:

$$\underset{\text{Stage 1 Prior}}{\begin{array}{cc} & \begin{array}{cc} NoDur & Durbl \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{pmatrix} \end{array}} \rightarrow \underset{\text{Stage 2 Prior}}{\begin{array}{cc} & \begin{array}{cc} NoDur & Durbl \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 3864 & -397 \\ -397 & 2045 \end{pmatrix} \end{array}} \rightarrow$$

$$\rightarrow \underset{\text{Posterior}}{\begin{array}{cc} & \begin{array}{cc} NoDur & Durbl \end{array} \\ \begin{array}{c} NoDur \\ Durbl \end{array} & \begin{pmatrix} 2270 & 224 \\ 224 & 1117 \end{pmatrix} \end{array}}$$

As can be seen the prior for the precision matrix varies with the model.

This example show that the proposed procedure allows automatically specify proper priors for each model.

3 Motivational Example

In order to motivate our research we simulate asset returns and demonstrate that the true factors are selected as the result of our procedure. We intentionally mis-specify the distribution of errors to show that even if the true model is not among the candidates under consideration but the imposed distribution is close enough to the true one, the correct factor structure is selected. In order to address a wide-spread concern that some non-factor assets may be selected if they are correlated with the market, we include such an asset as one of the candidate factors to demonstrate the robustness of the proposed selection procedure.

As asset returns we take 10 value-weighted Fama-French industry portfolios. The returns are assumed to be follow the famous Fama-French 3 factor structure without intercept. The factor include market portfolio (Mkt.RF), size (SMB) and value factors (HML). The true errors follow Student-t distribution with 2.5 degrees of freedom. The model parameters(coefficient vector $\boldsymbol{\gamma}$ and the error precision matrix Ω^{-1}) are equal posterior means obtained when fitting the model to the real data. The simulation is based on the sample range Apr 1986 - Dec 2014 (345 observations). The training sample includes observations Apr 1986 - Dec 1990 (57 observations). We assume that the true distribution (Student-t with 2.5 degrees of freedom) is not considered. Instead, considered distributions include normal and Student-t with 4, 6, 8, 10 and 12 degrees of freedom. The pool of candidate models includes all combinations of Fama-French 5 factors(Mkt.RF, HML, SMB, RMW and CMA), a constant and a non-factor asset - Microsoft stock (MSFT). We fit in total $6 \times 2^7 = 768$ model.

The simulation setup is described below:

1. Fit the Fama-French three factor model without an intercept to 10 value-weighted industry portfolios using the full sample under the assumption that errors follow Student-t distribution with 2.5 degrees of freedom. The posterior means $\boldsymbol{\gamma}^*$ and Ω^{-1*} obtained from the Gibbs sampler as a product of fitting this model are assumed to be the true parameter values for the simulation purposes.

2. Simulate a dataset for the full sample range assuming the true parameters to be equal to the posterior means and that errors follow Student-t distribution with 2.5 degrees of freedom.

$$\boldsymbol{\varepsilon}_t^s \sim t_{10,2.5}\left(0, \Omega^*\right)$$

   Simulate returns using the true values of the three Fama-French factors observed in the data :

$$\mathbf{y}_t^s = X_t\boldsymbol{\gamma}^* + \boldsymbol{\varepsilon}_t^s$$

3. Estimate all candidate models and evaluate the marginal likelihood. Use the training sample to specify priors for each candidate model.

As can be seen from the simulation results 3 the model with the highest marginal likelihood correctly identified three factors(Mkt.RF, HML and SMB). Moreover, even though the true error distribution (Student-t with 2.5 degrees of freedom) was not considered, the distribution of the best model (Student-t with 4 degrees of freedom) is the closest to the truth. Models including non-relevant factors or the non-factor (Microsoft stock) are significantly worse on the log scale. This simulation shows that marginal likelihood performs well in the finite sample.

4 Application

4.1 Data

Depending on the assets on the left hand side, different models may be selected. We apply our method to 10 value-weighted industry portfolios available at the Kenneth French's website: Consumer NonDurables(NoDur), Consumer Durables(Dur), Manufacturing(Manuf), Oil, Gas, and Coal Extraction and Products(Enrgy), Business Equipment(HiTec), Telephone and Television Transmission(Telcm), Wholesale, Retail, and Some Services (Shops), Healthcare, Medical Equipment, and Drugs(Hlth), Utilities(Utils), Other(Other). We choose industry-based portfolios instead of characteristics-based portfolios in order to avoid the potential bias that favors models which include factors similar to the sorting used to construct such portfolios.

The 12 candidate factors include (aside from the constant)[3]:

- five factors by Fama and French (1993) and Fama and French (2015): market(Mkt.RF), size(SMB), value(HML), profitability(RMW) and investment(CMA)
- three factors proposed by Hou et al. (2014b): size(ME), profitability(ROE) and investment(IA)

[3]We thank Lu Zhang for providing us the factors constructed in Hou et al. (2014b). Other factors are obtained from authors' web-pages.

- momentum factor(MOM) as in Carhart (1997)
- liquidity(LIQv) introduced in Pastor and Stambaugh (2003)
- quality factor(QMJ) as offered by Asness et al. (2014)
- alternative value factor(HMLDev) constructed by Asness and Frazzini (2013)

The full sample ranges from Jan 1968 - Dec 2014 (564 observations). The training sample is Jan 1968 - Dec 1979(144 observation).

We report the mean returns for the assets and factors in tables **??** and **??**. As can be seen from the correlation matrix 6, many factors co-move together. Sometimes this correlation arises because of the conceptual similarity of the factors. In our sample we have two highly correlated size factors: SMB by Fama and French (2015) and ME by Hou et al. (2014b). Another cluster of factors is related to value and investment activity. We include two value factors (HML by Fama and French (2015) and HMLDev by Asness and Frazzini (2013)) and two investment factors(CMA by Fama and French (2015) and IA by Hou et al. (2014b)). Finally, we have three profitability-based factors: RMW by Fama and French (2015), ROE by Hou et al. (2014b) and QMJ by Asness et al. (2014). The question of joint rather than individual significance is especially relevant for this set up given so many correlated factors.

4.2 Results

In total we explore 2^{13} models and consider 6 possible distribution of errors: Student-t with 4, 6, 8, 10 and 12 degrees of freedom and normal. As can be seen from the table 7, the suggested number of factors is quite large. E.g. the top model features 9 factors. Best four models are difficult to distinguish as they are similarly supported by the data. However, they have many features in common, in particular, the best models always include four out of the five Fama and French (2015) factors (everything except SMB), along with momentum, and quality minus junk. Liquidity factor Pastor and Stambaugh (2003) is never selected. Usually only one size factor (either SMB or ME) is included. However, multiple factors from other groups may be simultaneously selected. This is particularly true for profitability-based factors (at least two RMW by Fama and French (2015) and QMJ by Asness et al. (2014) are always included). The same holds for two value factors: HML by Fama and French (2015) and HMLDev by Asness and Frazzini (2013). Out of two investment factors the preference is given to the

Fama and French (2015) CMA. Multiple conceptually similar factors may be selected if they reflect different information which makes them jointly useful in explaining the cross-section of returns.

All four best models don't include an intercept. Consider the best model and compare it with an alternative specification that includes the same factors and a constant. The difference between these two models is approximately 0.86 on the $\log_{10}$ scale which can be interpreted as substantial support for the model without a constant by the Jeffrey's scale. This finding suggests that we can not reject the hypothesis that the intercept should not be included into the model.

5 Conclusion

Conclusion goes here...

References

Asness, C., Frazzini, A., 2013. The Devil in HML's details. Journal of Portfolio Management 39 (4), 49.

Asness, C. S., Frazzini, A., Pedersen, L. H., 2014. Quality minus junk. Available at SSRN 2312432.

Avramov, D., 2002. Stock return predictability and model uncertainty. Journal of Financial Economics 64 (3), 423–458.

Avramov, D., Chao, J. C., 2006. An exact bayes test of asset pricing models with application to international markets. The Journal of Business 79 (1), 293–324.

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

Carlin, B. P., Polson, N. G., 1991. Inference for nonconjugate bayesian models using the gibbs sampler. Canadian Journal of statistics 19 (4), 399–405.

Chib, S., 1995. Marginal likelihood from the gibbs output. Journal of the American Statistical Association 90 (432), 1313–1321.

Chib, S., Greenberg, E., 1995. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. Journal of Econometrics 68 (2), 339–360.

Chib, S., Greenberg, E., 1996. Markov chain monte carlo simulation methods in econometrics. Econometric theory 12 (03), 409–431.

Cremers, K. M., 2002. Stock return predictability: A Bayesian model selection perspective. Review of Financial Studies 15 (4), 1223–1249.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116 (1), 1–22.

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica: Journal of the Econometric Society, 1121–1152.

Harvey, C. R., Liu, Y., 2015. Lucky factors. Available at SSRN 2528780.

Harvey, C. R., Liu, Y., Zhu, H., 2015. ... and the cross-section of expected returns. Review of Financial Studies, hhv059.

Harvey, C. R., Zhou, G., 1990. Bayesian inference in asset pricing tests. Journal of Financial Economics 26 (2), 221–254.

Hou, K., Xue, C., Zhang, L., 2014a. A comparison of new factor models. Tech. rep., National Bureau of Economic Research.

Hou, K., Xue, C., Zhang, L., 2014b. Digesting anomalies: An investment approach. Review of Financial Studies, hhu068.

Judd, C. M., McClelland, G. H., Ryan, C. S., 2011. Data analysis: A model comparison approach. Routledge.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics, 13–37.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111 (3).

Shanken, J., 1987. A Bayesian approach to testing portfolio efficiency. Journal of financial economics 19 (2), 195–215.

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

		NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
	Apr 1986 - Dec 2014	0.0080	0.0056	0.0078	0.0080	0.0074	0.0062	0.0073	0.0084	0.0061	0.0058
Data	Apr 1986 - Dec 1990	0.0099	-0.0033	0.0034	0.0092	-0.0037	0.0084	0.0036	0.0095	0.0033	-0.0020
	Jan 1991 - Dec 2014	0.0076	0.0073	0.0086	0.0078	0.0096	0.0057	0.0080	0.0082	0.0066	0.0073
	Apr 1986 - Dec 2014	0.0035	0.0104	0.0063	0.0036	0.0062	0.0089	0.0045	0.0052	0.0043	0.0085
Simulated	Apr 1986 - Dec 1990	-0.0016	0.0039	0.0020	0.0005	0.0032	0.0106	-0.0016	0.0025	0.0047	0.0036
	Jan 1991 - Dec 2014	0.0045	0.0117	0.0071	0.0042	0.0067	0.0086	0.0057	0.0057	0.0043	0.0094

Table 1: Simulation: Average Portfolio Returns

	Mkt.RF	SMB	HML	RMW	CMA	MSFT
Apr 1986 - Dec 2014	0.0062	0.0011	0.0023	0.0036	0.0033	0.0102
Apr 1986 - Dec 1990	0.0028	-0.0068	0.0003	0.0047	0.0054	0.0327
Jan 1991 - Dec 2014	0.0069	0.0027	0.0027	0.0034	0.0029	0.0058

Table 2: Simulation:Average Factor Returns

Model	DF	log margLike
Mkt.RF + SMB + HML	4	7187.32
Mkt.RF + SMB + HML + CMA	4	7177.05
Mkt.RF + SMB + HML + MSFT	4	7170.99
Mkt.RF + SMB + HML + RMW	4	7170.24
Mkt.RF + SMB + HML	6	7169.59
Mkt.RF + SMB + HML + CMA + MSFT	4	7161.34
constant + Mkt.RF + SMB + HML	4	7160.41
Mkt.RF + SMB + HML + CMA	6	7159.73
constant + Mkt.RF + SMB + HML + CMA	4	7157.97
Mkt.RF + SMB + HML + RMW + MSFT	4	7152.92
Mkt.RF + SMB + HML	8	7152.24
constant + Mkt.RF + SMB + HML + MSFT	4	7151.56
Mkt.RF + SMB + HML + RMW	6	7151.19
constant + Mkt.RF + SMB + HML	6	7149.93
constant + Mkt.RF + SMB + HML + RMW	4	7147.47
Mkt.RF + SMB + HML + MSFT	6	7147.06
Mkt.RF + SMB + HML + RMW + CMA	4	7146.92
Mkt.RF + SMB + HML + CMA + MSFT	6	7146.04
constant + Mkt.RF + SMB + HML + CMA + MSFT	4	7145.13
constant + Mkt.RF + SMB + HML + MSFT	6	7139.5

Table 3: Simulation: 20 Best Models

	NoDur	Durbl	Manuf	Enrgy	HiTec	Telcm	Shops	Hlth	Utils	Other
Jan 1968 - Dec 2014	0.0068	0.0046	0.0055	0.0064	0.0048	0.0054	0.0063	0.0066	0.0049	0.0050
Jan 1968 - Dec 1979	0.0006	-0.0009	-0.0003	0.0048	-0.0012	0.0013	0.0004	0.0013	0.0003	0.0005
Jan 1980 - Dec 2014	0.0090	0.0064	0.0075	0.0070	0.0069	0.0068	0.0084	0.0084	0.0065	0.0065

Table 4: Average Portfolio Returns

	LIQv	MOM	Mkt.RF	SMB	HML	RMW	CMA	QMJ	ME	IA	ROE	HMLDev
Jan 1968 - Dec 2014	0.0043	0.0066	0.0049	0.0020	0.0038	0.0026	0.0037	0.0039	0.0028	0.0046	0.0055	0.0039
Jan 1968 - Dec 1979	0.0014	0.0080	0.0003	0.0039	0.0059	-0.0004	0.0049	0.0018	0.0049	0.0061	0.0043	0.0073
Jan 1980 - Dec 2014	0.0053	0.0061	0.0065	0.0014	0.0030	0.0037	0.0033	0.0046	0.0020	0.0040	0.0059	0.0028

Table 5: Average Factor Returns

	LIQv	MOM	Mkt.RF	SMB	HML	RMW	CMA	QMJ	ME	IA	ROE	HMLDev
LIQv												
MOM	-0.02											
Mkt.RF	-0.06	-0.13										
SMB	-0.03	-0.05	0.28									
HML	0.03	-0.15	-0.32	-0.13								
RMW	0.03	0.1	-0.21	-0.38	0.11							
CMA	0.02	0.02	-0.4	-0.09	0.71	-0.06						
QMJ	0.04	0.25	-0.53	-0.52	0.02	0.76	0.08					
ME	-0.04	-0.02	0.26	0.97	-0.08	-0.37	-0.05	-0.5				
IA	0.02	0.04	-0.39	-0.19	0.69	0.06	0.91	0.15	-0.15			
ROE	-0.06	0.5	-0.2	-0.39	-0.1	0.68	-0.08	0.69	-0.32	0.05		
HMLDev	0.07	-0.64	-0.13	-0.02	0.77	-0.07	0.51	-0.21	-0.01	0.49	-0.45	

Table 6: FactorsCorrelation Matrix (based on full sample)

Model	DF	log margLike
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE + HMLDev	6	9713.58
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + HMLDev	6	9713.53
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE	6	9713.47
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME	6	9712.45
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ROE + HMLDev	6	9712.26
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ROE	6	9712.21
constant + MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE + HMLDev	6	9711.61
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + IA + HMLDev	6	9711.1
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + IA + HMLDev	6	9710.89
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + HMLDev	6	9710.62
constant + MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ROE + HMLDev	6	9710.29
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + IA	6	9710.11
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + ROE	6	9710.07
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + IA + ROE	6	9709.84
constant + MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE	6	9709.69
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + IA	6	9709.68
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME	6	9709.58
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + ROE + HMLDev	6	9708.89
constant + MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ROE	6	9708.65
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + IA + ROE + HMLDev	6	9708.33

Table 7: 20 Best Models