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 22, 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 factor selection problem by appropriately penalizing more complex models relative to simpler ones.

In this paper we take both of these observations to heart and use monthly observations for twelve leading asset pricing factors and ten industry portfolios to carry out an exhaustive search over 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 automatically pe-

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

nalize models based on complexity, assuring that models 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. Given the large number of parameters in our SUR model

the subset models are essentially special cases of a seemingly unrelated regression (SUR) model with the same subset of factors on the right-hand side but with different asset-specific factor coefficients, and a jointly distributed vector error with an unknown precision matrix. Restricting ourselves to 12 factors (10 from Harvey and Liu (2015) and 2 additional ones), and an intercept which can be present or absent in each possible case, leads to 2^{13} possible SUR models, which along with 6 different assumptions about the error distribution, amounts to the comparison of 49152 SUR models.

Our next innovation is to utilize a carefully crafted Bayesian approach to implement the comparison of these disparate models. Careful in this context means that the prior distribution is not placed independently on each factor, since we recognize that the meaning and importance of the factor coefficients and the error precision matrix vary by model. Each of the required 49152 different prior distributions are specified in an automatic and importantly, objective fashion, by employing a small training sample that precedes our estimation sample. Careful also refers to the use of a SUR type model in which error distribution and error precision are also viewed as unknowns that have to be inferred.

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).

Talk about results here!

The rest of the paper unfolds as follows. Section 1 describes the model with Student-t errors and the estimation framework. Section 2 provides a simulation example to motivate our research. Section 4 describes the main results. Finally, section 5 concludes.

2 Model and Framework

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

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

where $d = 1, \ldots, D$ and $t = 1, \ldots, T$ and $\mathbf{f}_t' = (f_{t1}, \ldots, f_{tK})$ is a $K \times 1$ vector.

We consider a cross-section of D returns. y_{it} denotes a return for an asset i at period t. The cross-section of returns on right hand side is the same for all models. The returns are assumed to follow a common factor structure with K factors denoted as $\mathbf{f}_t$. 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 the model index, however, it is important to keep in mind that the techniques discussed below are applied to each possible combination of factors. This is a special case of the seemingly unrelated regression (SUR) model in which the regressors are *identical* across equations.

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. 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 should make an assumption about the joint distribution of the errors ε_t. Different specifications of the errors distribution can lead to the selection of different models. The normality assumption, which is often exploited due to the conjugacy properties, should be treated with caution because the tail behavior of the financial returns is different from the one imposed by Gaussian distribution. For this reason we fit a multivariate Student-t distribution that accounts for fat tails specific to the returns. As we want to remain agnostic about the shape of the distribution, we explore multiple distributions from the same family with different degrees of freedom and let the data determined the best one.

Suppose now that the errors follow a multivariate Student-t distribution:

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

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 $Var\,(\boldsymbol{\varepsilon}) = \nu\Omega/(\nu - 2)$.

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 have little understanding about their nature and thus should avoid imposing rigid assumptions. Careful treatment of the covariance matrix Ω and its inverse Ω^{-1} allows us to infer the covariance structure of errors from the data.

2.1 A Hierarchical Representation

Replacing the normal likelihood with the Student-t likelihood, however, breaks the conditional conjugacy that is usually exploited to construct an MCMC algorithm based on the Gibbs sampler. The solution to this problem is to work with a hierarchical representation in which the Student-t likelihood is introduced as a scale mixture of normal distributions (Chib and Greenberg (1995)), in particular

$$\varepsilon_t|\lambda_t \sim N\left(0, \lambda_t^{-1}\Omega\right)$$
$$\lambda_t \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

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 $p = D(K + 1)$. Using this representation, after conditioning on $(\nu, \boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots \lambda_T)'$, we are essentially back in the familiar normal case. 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}.

2.2 The Gibbs Sampler

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

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, which makes the initial draws for the regression coefficients and the inverse scale matrix the same as if were work withing with the normal model.

2.3 Marginal Likelihood for Student-t Model

We calculate the marginal likelihood using the method of Chib (1995).

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

Re-arranging Bayes' Rule we have the identity

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

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

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

Specializing this to the SUR model considered above,

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

since our priors over $\boldsymbol{\gamma}$ and Ω^{-1} are independent. 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.

2.4 Priors

One of the most challenging parts of the estimation procedure is to set up reasonable priors. In order for the marginal likelihood comparison to be valid, the marginal likelihood for each model should be compute with proper prior. Using generic prior may lead to undesirable consequences. In other words, it's important to set the priors carefully and make them specific to each model under consideration. In order to set individual level priors, we suggest to use a training sample to form qualified prior views for every model at hand. The priors based on he training sample are model-specific and thus can be used for model comparison. We split the sample into two parts: training sub-sample and fit sub-sample. The training sample is used to carefully construct priors later used to fit the model on the rest of the data. Training sample is not a part of the final estimation.

Both for Student-t and normal errors we specify two prior distributions: for the coefficients $\gamma \sim \mathcal{N}_p(\boldsymbol{\gamma}|\boldsymbol{\gamma_0}, G_0)$ and for the precision matrix $\Omega^{-1} \sim \mathcal{W}_D(\Omega^{-1}|\rho_0, R_0)$. Parameters for the two distributions are specified differently for each stage.

2.4.1 Stage 1: Training Sample

When applying the method to the training sample, we have very little information about the possible parameters values. We use diffused priors to reflect the uncertainty. For each model we start with the assumption that factor loadings can be both positive and negative. We have very little information about how these loadings interact, so our starting point would be to use diagonal covariance matrix of factor loadings. Parameters of the prior for the regression coefficients are set as follows:

$$\begin{aligned} \gamma_0 &= 0 \\ G_0 &= C_1^2 I_p \end{aligned}$$

The prior distribution of the coefficient vector is centered around zero. The covariance matrix is assumed to be diagonal. The constant C_1 controls the tightness: the larger C_1, the wider is the prior. As we have very little information about the covariance structure of the errors, a prior centered around the diagonal matrix seems to be a reasonable starting point. Covariance matrix is assumed to follow the inverse Wishart distribution with the following parameters:

$$\begin{aligned} \rho_0 &= d + C_2 \\ R_0 &= \frac{1}{C_3^2(\rho_0 - d - 1)} I_d \end{aligned}$$

Using the properties of the inverse Wishart distribution, one can easily see that the mean value of Ω implied by the prior is a diagonal matrix $R_0 \times (\rho_0 - d - 1) = C_3^2 I_d$. One way of widening the prior would be to decrease number of degrees of freedom ρ_0 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 $C_3 = 0.05$ (we are working with return in decimals). These priors are used to estimate the model using only the training sample data. The posterior draws obtained as a result of the procedure, are later used to form more informed view about the model parameters.

2.4.2 Stage 2

The first stage draws contain some information about the parameters and serve as a basis to construct proper priors for the estimation. Denote first stage posterior means of γ as $\overline{\gamma}$ and the sample covariance matrix of these draws as $\widehat{G}$. We also calculate a posterior mean of the precision matrix Ω^{-1}: $\overline{\Omega}^{-1}$. The regression coefficients prior is:

$$\begin{aligned} \gamma_0 &= \overline{\gamma} \\ G_0 &= C_4^2 \widehat{G} \end{aligned}$$

The prior of factor loadings is thus centered around the posterior mean of the first step 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. 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}. We use the duality of Wishart and inverse Wishart distribution: if $\Omega \sim \mathcal{W}_D(\rho_0, R_0)$, $\Omega^{-1} \sim \mathcal{IW}_D(\rho_0, R_0)$. The prior of the precision matrix is set:

$$\begin{aligned} \rho_0 &= d + C_5 \\ R_0 &= \frac{1}{\rho} \overline{\Omega}^{-1} \end{aligned}$$

The prior is constructed to set the mean of the precision matrix Ω^{-1} equal to the posterior mean based on the training sample: $\rho_0 \times R_0 = \overline{\Omega}^{-1}$. 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 $C_5 = 6$.

2.5 Prior Updates

Fill in something here?

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 the procedure described above. 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 (Mkt.RF, HML and SMB). The errors follow Student-t distribution with 2.5 degrees of freedom. The simulation is based on posterior means obtained when fitting the model to the real data. Other parameters are the same as in the original sample. We assume that the researcher does not know the true distribution. 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 add a Microsoft stock because of the wide-spread concern that non-factors may be selected if they are correlated with the true factors. Our results indicate this concern is not supported in the Bayesian framework. We fit in total $6 \times 2^7 = 768$ model. 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).

The simulation setup is described below:

1. Fit the Fama-French 3 factor model without an intercept to 10 value-weighted industry portfolios using the full sample. The errors are assumed to follow Student-t distribution with 4 degrees of freedom.

2. Simulate a dataset assuming the true parameters γ and Ω^{-1} to be equal to the posterior means:

   - Simulate errors:
   $$\boldsymbol{\varepsilon}_t^s \sim t_{10,2.5}\left(0, \Omega\right)$$
   - Simulate returns using values of Fama-French 3 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 likelihood. Run the usual two step estimation procedure using the training sample to construct priors for each model.

The best model selects the true factors. 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. As can be seen from the results, the procedure correctly identified the factors. Models including non-relevant factors or the non-factor (Microsoft stock) are significantly worse on the log scale.

4 Application

4.1 Data

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).

The 12 candidate factors include (aside from the constant)[2]:

- 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)
- momentum factor(MOM) as in Carhart (1997)
- liquidity(LIQv) introduced in 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).

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

We report the summary statistics in the tables below. As can be seen from the correlation matrix, many factors co-move together. This is the reason why the question of joint rather than individual significance is especially relevant for this set up. Our method is capable of addressing this issue because our selection of models is based on comparing all possible combinations of 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.

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.

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.

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.

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.

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

		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: Average Portfolio Returns for Simulated and Real Data

	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: 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: Motivational 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	1.00	-0.02	-0.06	-0.03	0.03	0.03	0.02	0.04	-0.04	0.02	-0.06	0.07
MOM	-0.02	1.00	-0.13	-0.05	-0.15	0.10	0.02	0.25	-0.02	0.04	**0.50**	**-0.64**
Mkt.RF	-0.06	-0.13	1.00	0.28	-0.32	-0.21	-0.40	**-0.53**	0.26	-0.39	-0.20	-0.13
SMB	-0.03	-0.05	0.28	1.00	-0.13	-0.38	-0.09	**-0.52**	**0.97**	-0.19	-0.39	-0.02
HML	0.03	-0.15	-0.32	-0.13	1.00	0.11	**0.71**	0.02	-0.08	**0.69**	-0.10	**0.77**
RMW	0.03	0.10	-0.21	-0.38	0.11	1.00	-0.06	**0.76**	-0.37	0.06	**0.68**	-0.07
CMA	0.02	0.02	-0.40	-0.09	**0.71**	-0.06	1.00	0.08	-0.05	**0.91**	-0.08	**0.51**
QMJ	0.04	0.25	**-0.53**	**-0.52**	0.02	**0.76**	0.08	1.00	**-0.50**	0.15	**0.69**	-0.21
ME	-0.04	-0.02	0.26	**0.97**	-0.08	-0.37	-0.05	**-0.50**	1.00	-0.15	-0.32	-0.01
IA	0.02	0.04	-0.39	-0.19	**0.69**	0.06	**0.91**	0.15	-0.15	1.00	0.05	0.49
ROE	-0.06	**0.50**	-0.20	-0.39	-0.10	**0.68**	-0.08	**0.69**	-0.32	0.05	1.00	-0.45
HMLDev	0.07	**-0.64**	-0.13	-0.02	**0.77**	-0.07	**0.51**	-0.21	-0.01	0.49	-0.45	1.00

Table 6: Factors Correlation 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 + HML + RMW + CMA + QMJ + ME + ROE + HMLDev	8	9713.33
MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE	8	9712.67
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + HMLDev	8	9712.61
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
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ROE	8	9711.74
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ROE + HMLDev	8	9711.72
constant + MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE + HMLDev	6	9711.61
constant + MOM + Mkt.RF + HML + RMW + CMA + QMJ + ME + ROE + HMLDev	8	9711.59
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME	8	9711.46
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 + HMLDev	8	9710.2
MOM + Mkt.RF + SMB + HML + RMW + CMA + QMJ + ME + IA	6	9710.11

Table 7: 20 Best Models