

1 Abstract

In this paper we revisit one of the classic questions in empirical finance: which factors in combination are useful for explaining the time-series and cross-section behavior of equity and portfolio returns? The contribution of this paper is to consider this question from a Bayesian perspective recognizing that proper evaluation of the worth of a factor has to be in the context of models with and without other factors. Answering such a question therefore requires the consideration of all possible subset models, where 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 (inverse covariance) matrix. In the case of a leading set of 12 factors, along with the intercept which can be present or absent in each possible case, this 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. We carefully compare these models with the help of objectively constructed priors (one for each of our models) using a training sample, and with the calculation of marginal likelihoods, computed by the method of (Chib, 1995). Marginal likelihoods are proportional to the posterior probability of each model and have recognized finite sample and asymptotic properties. In particular, marginal likelihoods include a penalty for complexity (in other words models with more factors do not necessarily gather greater support) and asymptotically pick either the true model (if it is in the class being considered) or find the model that is closest to the true model (if it is not in the class being considered). Our comparison focuses on test assets from the current literature on this topic: a collection of 10 asset portfolios and a collection of 10 equities. Our results show ...

2 Introduction

An abiding, key 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 have emerged as documented in (Harvey, Liu, & Zhu, 2015). Determining the empirical relevance of these factors is an ongoing statistical challenge and different avenues continue to be energetically explored. For instance, (Hou, Xue, & Zhang, 2014a) compare the

(Hou, Xue, & Zhang, 2014b) and (Fama & French, 2015) five-factor models based on the conceptual understanding of factors and the ability of the factors to explain asset-pricing anomalies. A more statistical evaluation is provided by (Harvey & Liu, 2015) who start with a collection of 12 leading factors and use a bootstrap procedure to forward select relevant factors. Stepwise selection, however, does not necessarily asymptotically produce the best model. This is particularly true in the presence of redundant factors (reference needed).

Our goal in this paper is to offer a quite different method for finding the best collection of factors. Our starting point is the recognition that the proper evaluation of the worth of a factor has to be in the context of models with and without other factors. Thus, basing a decision of the importance of a factor based solely on the t-statistic, or related statistics, ignores the question of joint significance and correlation between factors. Instead, answering the relevance question requires the consideration of all possible subset models, where 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 the same set of 12 factors of (Harvey & Liu, 2015) (along with the intercept which can be present or absent in each possible case)(note: we don't use the same factors as (Harvey & Liu, 2015)), 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. Decisions about the best model are then taken by the comparison of Bayesian marginal likelihoods, computed readily by the method of (Chib, 1995), for each of these models. This comparison of our models by Bayesian marginal likelihoods has well-recognized theoretical properties (in a small simulation exercise we also document the excellent performance of the marginal likelihood criterion in picking the true model). In particular, marginal likelihoods include a penalty for complexity. In other words, models with more factors do not necessarily gather greater support. In addition, models picked according to the marginal likelihoods, asymptotically pick either

the true model, if it is in the class being considered, or find the model that is closet to the true model, if it is not in the class being considered. Our marginal likelihood based model exploration, therefore, provides an effective, previously unexplored avenue for evaluating the joint rather than individual importance of factors. 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 & Zhou, 1990) and (Avramov & 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, Ross, & Shanken, 1989).

(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. Results for the model with normal errors and general description of the three block sample can be found in the appendix.

3 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 $d = 1, \dots, D$ and $t = 1, \dots, T$ and $\mathbf{f}_t' = (f_{t1}, \dots, f_{tK})$ is a $K \times 1$ vector. 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 $\mathbf{y}_t' = (y_{1t}, \dots, y_{Dt})$ and analogously $\boldsymbol{\varepsilon}_t' = (\varepsilon_{t1}, \dots, \varepsilon_{tD})$. Now let $\mathbf{x}_t' = (1, \mathbf{f}_t')$ and $\boldsymbol{\gamma}_d' = (\alpha_d, \boldsymbol{\beta}_d')$ so we have

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

where $X_t = I_D \otimes \mathbf{x}_t'$ and $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}_1', \dots, \boldsymbol{\gamma}_D')$. Let Y_T denote the full data sample, i.e. $\{\mathbf{y}_t, \mathbf{x}_t\}_{t=1}^T$. In order to proceed, we should also make an assumption about the distribution of errors $\boldsymbol{\varepsilon}_t$. In this paper we consider two distributions: Student-t and normal. The normal distribution is more common and easier to work with because of the conjugacy properties. However, returns are widely known to be non-Gaussian and fat-tailed. For this reason we believe that Student-t errors provide a more accurate description of the

series of interest. We further provide some details about how one can tackle the Student-t errors by representing it as a mixture of normal distributions. The case of normally distributed errors is discussed in the appendix.

3.1 Gibbs Sampler with Student-t Errors

3.2 A Hierarchical Representation

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 $Var(\boldsymbol{\varepsilon}) = \nu\Omega/(\nu-2)$. 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 & Greenberg, 1995)), in particular

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

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 $p = D(K + 1)$. Using this representation, after conditioning on $(\nu, \boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_T)'$, 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} .

3.2.1 The 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: $\gamma|\Omega^{-1}, Y_T \sim \mathcal{N}_p(\bar{\gamma}_\lambda, G_{T,\lambda})$

$$G_{T,\lambda} = \left[G_0^{-1} + \sum_{t=1}^T \lambda_t X_t' \Omega^{-1} X_t \right]^{-1}$$

$$\bar{\gamma}_\lambda = G_{T,\lambda} \left[G_0^{-1} \gamma_0 + \sum_{t=1}^T \lambda_t X_t' \Omega^{-1} \mathbf{y}_t \right]$$

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

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

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

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 we were working with the normal model.

3.2.2 Marginal Likelihood for Student-t Model

We calculate the marginal likelihood using the method of (Chib, 1995). First we describe the method for an arbitrary three-block Gibbs sampler. We will use the idea of a “reduced run” in our calculation of the marginal likelihood of the Student-t model.

General Three-Block Algorithm . Re-arranging Bayes’ Rule we have the identity

$$f(y) = \frac{f(y|\theta_1^*, \theta_2^*, \theta_3^*) \pi(\theta_1^*, \theta_2^*, \theta_3^*)}{\pi(\theta_1^*, \theta_2^*, \theta_3^*|y)}$$

for any specified values $(\theta_1^*, \theta_2^*, \theta_3^*)$ 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

$$\log f(y) = \log \pi(\theta_1^*, \theta_2^*, \theta_3^*) + \log f(y|\theta_1^*, \theta_2^*, \theta_3^*) - \pi(\theta_1^*, \theta_2^*, \theta_3^*|y)$$

The first two terms are easy: we simply evaluate the log of the prior and likelihood at the posterior mean for the three parameters. The third one, however, is more complicated.

To calculate it we use the factorization:

$$\pi(\theta_1^*, \theta_2^*, \theta_3^* | y) = \pi(\theta_1^* | y) \pi(\theta_2^* | \theta_1^*, y) \pi(\theta_3^* | \theta_2^*, \theta_1^*, y)$$

This leaves us with three a product of new terms that we need to calculate. The last of the terms in the product, $\pi(\theta_3^* | \theta_2^*, \theta_1^*, y)$ is immediately available: this conditional density is known since we used it as a step in the Gibbs sampler. All we need to do is substitute in the appropriate values for θ_2^*, θ_1^* and evaluate the density at θ_3^* . To calculate the first term in the product we need to marginalize over θ_2, θ_3 . A Monte-Carlo approximation to the appropriate integral can be computed directly from the Gibbs sampler output:

$$\hat{\pi}(\theta_1^* | y) = \frac{1}{G} \sum_{g=1}^G \pi(\theta_1^* | \theta_2^{(g)}, \theta_3^{(g)}, y)$$

To do this we rely on the fact that $\pi(\theta_1 | \theta_2, \theta_3, y)$ is a known density – we use it in the Gibbs sampler. The middle term in the product is the most difficult one to calculate. To begin, notice that

$$\pi(\theta_2^* | \theta_1^*, y) = \int \pi(\theta_2^*, \theta_3 | \theta_1^*, y) d\theta_3 = \int \pi(\theta_2^* | \theta_1^*, \theta_3^*, y) \pi(\theta_3 | \theta_1^*, y) d\theta_3$$

The idea is to construct a Monte-Carlo approximation of the integral on the right-hand-side of the preceding expression. The approximation we use is

$$\hat{\pi}(\theta_2^* | \theta_1^*, y) = \frac{1}{G} \sum_{g=1}^G \pi(\theta_2^* | \theta_1^*, \theta_3^{(g)}, y)$$

but the draws $\{\theta_3^{(g)}\}$ come *not* from the original run of the Gibbs sampler but from a so-called “reduced run” (introduced by (Chib, 1995)) in which we sample $\theta_2^{(g)}$ and $\theta_3^{(g)}$ from $\pi(\theta_2 | \theta_1^*, \theta_3, y)$ and $\pi(\theta_3 | \theta_1^*, \theta_2, y)$. In other words, the reduced run holds θ_1 *fixed* at θ_1^* , the posterior mean calculated from the draws of the *usual* Gibbs sampler. We can carry out the reduced run using the exact same algorithm as we use for the full Gibbs sampler: we just need to keep θ_1^* fixed and make sure that we store the draws $\theta_3^{(g)}$ that we will need to calculate $\hat{\pi}(\theta_2^* | \theta_1^*, y)$. Let θ denote the full collection of parameters. By Bayes’ Rule

$$\pi(\theta | Y_T) = \frac{\pi(\theta) f(Y_T | \theta)}{f(Y_T)}$$

where $f(Y_T)$ is the marginal likelihood, aka the marginal data density, aka the evidence. This identity holds true for *any* value of θ . In particular it holds at the posterior mean θ^* . Solving for $f(Y_T)$ and evaluating the result at θ^* , we have

$$f(Y_T) = \frac{\pi(\theta^*) f(Y_T | \theta^*)}{\pi(\theta^* | Y_T)}$$

Thus, we can express the *log* marginal likelihood as

$$\log f(Y_T) = \log \pi(\theta^*) + \log f(Y_T|\theta^*) - \log \pi(\theta^*|Y_T)$$

Specializing this to the SUR model considered above,

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

since our priors over γ and Ω^{-1} are independent. The (Chib, 1995) method 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.

The Contribution of the Prior Evaluating the first two terms, $\log \pi(\gamma^*)$ and $\log \pi(\Omega^{-1*})$, is easy: these are simply the priors for γ and Ω^{-1} evaluated at the posterior means. We take the sample average of the Gibbs draws to approximate γ^* and Ω^{-1*} and evaluate the Normal and Wishart distributions at these points, with parameters given by the prior:

$$\begin{aligned} \pi(\gamma^*) &= \mathcal{N}_p(\gamma^*|\gamma_0, G_0) \\ \pi(\Omega^{-1*}) &= \mathcal{W}_D(\Omega^{-1*}|\rho_0, R_0) \end{aligned}$$

The Contribution of the Likelihood In the Student-t model we have $\varepsilon_t|\mathbf{x}_t \sim \text{iid } t_{D,\nu}(0, \Omega)$ and hence, from the regression specification, it follows that $\mathbf{y}_t \sim \text{iid } t_{D,\nu}(X_t\gamma, \Omega)$. Hence, evaluating the log-likelihood at the posterior mean gives

$$\log f(Y_T|\gamma^*, \Omega^{-1*}) = \sum_{t=1}^T \log t_{D,\nu}(\mathbf{y}_t|X_t\gamma^*, \Omega^{-1*})$$

where we parameterize in terms of Ω^{-1} .

The Contribution of the Posterior It is the *final* term in the log marginal likelihood expression, the contribution of the posterior, whose computation is substantially different for the case of the Student-t model. We factorize the contribution of the posterior according to:

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

To approximate the $\log \pi(\Omega^{-1*}|Y_T)$ term we evaluate the density $\pi(\Omega^{-1}|\gamma, \mathbf{\lambda}, Y_T)$ at Ω^{-1*} and marginalize over the original Gibbs sampler draws $\{\gamma^{(g)}, \mathbf{\lambda}^{(g)}\}$, that is

$$\begin{aligned} \hat{\pi}(\Omega^{-1*}|Y_T) &= \frac{1}{G} \sum_{g=1}^G \pi(\Omega^{-1*}|\gamma^{(g)}, \mathbf{\lambda}^{(g)}, Y_T) \\ &= \frac{1}{G} \sum_{g=1}^G \mathcal{W}_D(\Omega^{-1*}|\rho_0 + T, R_{T, \mathbf{\lambda}^{(g)}}^{(g)}) \end{aligned}$$

where

$$\begin{aligned} R_{T,\lambda^{(g)}}^{(g)} &= \left[R_0^{-1} + \sum_{t=1}^T \lambda_t (\mathbf{y}_t - X_t \boldsymbol{\gamma}^{(g)}) (\mathbf{y}_t - X_t \boldsymbol{\gamma}^{(g)})' \right]^{-1} \\ &= \left[R_0^{-1} + \left(\tilde{Y} - \tilde{X} \Gamma^{(g)} \right)' \Lambda^{(g)} \left(\tilde{Y} - \tilde{X} \Gamma^{(g)} \right) \right]^{-1} \end{aligned}$$

All that remains is to approximate $\pi(\boldsymbol{\gamma}^* | \Omega^{-1*}, Y_T)$. To do this, we will need to use a reduced run similar to the one used in the calculations for the general three-block Gibbs-sampler described above. We know $\pi(\boldsymbol{\gamma} | \Omega^{-1}, \boldsymbol{\lambda}, Y_T)$ so we will evaluate this expression at $\boldsymbol{\gamma}^*$ and Ω^{-1*} and integrate out $\boldsymbol{\lambda}$ using a set of draws $\{\boldsymbol{\lambda}^{(j)}\}$ that were generated holding Ω^{-1} *fixed* at Ω^{-1*} . Accordingly, we use

$$\begin{aligned} \hat{\pi}(\boldsymbol{\gamma}^* | \Omega^{-1*}, Y_T) &= \frac{1}{J} \sum_{j=1}^J \pi(\boldsymbol{\gamma}^* | \Omega^{-1*}, \boldsymbol{\lambda}^{(j)}, Y_T) \\ &= \frac{1}{J} \sum_{j=1}^J \mathcal{N}_p(\boldsymbol{\gamma}^* | \bar{\boldsymbol{\gamma}}_{\lambda^{(j)}}^*, G_{n,\lambda^{(j)}}^*) \end{aligned}$$

where

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

3.3 Priors

One of the most challenging parts of the estimation procedure is to set up reasonable priors. We suggest to split the sample into two parts: training sub-sample and fit sub-sample. The training sample is used to form qualified prior views that will then be used as priors when fitting the model on the rest of the data. Training sample is only used to form priors and is not a part of the final estimation.

Both for Student-t and normal errors we specify two prior distributions: for the coefficients $\boldsymbol{\gamma} \sim \mathcal{N}_p(\boldsymbol{\gamma} | \boldsymbol{\gamma}_0, G_0)$ 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.

3.3.1 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.

3.3.2 Stage 2

The first stage draws contain some information about the parameters and serve as an base of constructing better priors for the rest of the sample. Denote posterior means of γ as $\bar{\gamma}$ and the sample covariance matrix of these draws as \hat{G} . We also calculate a posterior mean of the precision matrix Ω^{-1} : $\bar{\Omega}^{-1}$.

The regression coefficients prior is:

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

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:

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

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

4 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 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:

$$\epsilon_t^s \sim t_{10,2.5}(0, \Omega)$$

- Simulate returns using values of Fama-French 3 factors observed in the data:

$$\mathbf{y}_t^s = X_t \gamma + \epsilon_t^s$$

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.

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

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

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.

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

5 Application

5.1 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)¹:

- five factors by (Fama & French, 1993) and (Fama & 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)

¹We 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 (Stambaugh, 2003)
- quality factor(QMJ) as offered by (C. S. Asness, Frazzini, & Pedersen, 2014)
- alternative value factor(HMLDev) constructed by (C. Asness & 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 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.

	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)

5.2 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.

6 Conclusion

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.
- 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).

7 Appendix

7.1 Gibbs Sampler with Normal Errors

Now, suppose that

$$\boldsymbol{\varepsilon}_t | \mathbf{x}_t \sim \text{iid } \mathcal{N}_D(0, \Omega)$$

Then the likelihood is

$$\pi(Y_T | \boldsymbol{\gamma}, \Omega^{-1}) \propto |\Omega^{-1}|^{T/2} \exp \left[-\frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - X_t \boldsymbol{\gamma})' \Omega^{-1} (\mathbf{y}_t - X_t \boldsymbol{\gamma}) \right]$$

where 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 $p = D(K + 1)$.

7.1.1 Prior and Posterior Distribution

To complete the model we specify the following prior distribution

$$\pi(\boldsymbol{\gamma}, \Omega^{-1}) = \mathcal{N}_p(\boldsymbol{\gamma} | \boldsymbol{\gamma}_0, G_0) \mathcal{W}_D(\Omega^{-1} | \rho_0, R_0)$$

This prior is conditionally conjugate with the normal likelihood. In particular, we have $\gamma|\Omega^{-1}, Y_T \sim \mathcal{N}_p(\bar{\gamma}, G_T)$ where

$$\begin{aligned} G_T &= \left[G_0^{-1} + \sum_{t=1}^T X_t' \Omega^{-1} X_t \right]^{-1} \\ \bar{\gamma} &= G_T \left[G_0^{-1} \gamma_0 + \sum_{t=1}^T X_t' \Omega^{-1} \mathbf{y}_t \right] \end{aligned}$$

and $\Omega^{-1}|Y_T \sim \mathcal{W}_D(\rho_0 + T, R_T)$ where

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

7.1.2 MCMC

Using the full set of conditional posteriors, given in the preceding section, we can simulate from the joint posterior for this model using a Gibbs sampler:

1. Select a starting value $\Omega^{-1(0)}$ for the precision matrix.
2. Draw $\gamma^{(1)} \sim \mathcal{N}(\bar{\gamma}^{(1)}, G_T^{(1)})$ where

$$\begin{aligned} G_T^{(1)} &= \left[G_0^{-1} + \sum_{t=1}^T X_t' \Omega^{-1(0)} X_t \right]^{-1} \\ \bar{\gamma}^{(1)} &= G_T^{(1)} \left[G_0^{-1} \gamma_0 + \sum_{t=1}^T X_t' \Omega^{-1(0)} \mathbf{y}_t \right] \end{aligned}$$

3. Draw $\Omega^{-1(1)} \sim \mathcal{W}_D(\rho_T, R_T^{(1)})$ where

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

4. Repeat the preceding two steps a total of G times. In the g th iteration:

- (i) Draw $\gamma^{(g)} \sim \mathcal{N}(\bar{\gamma}^{(g)}, G_T^{(g)})$ where

$$\begin{aligned} G_T^{(g)} &= \left[G_0^{-1} + \sum_{t=1}^T X_t' \Omega^{-1(g-1)} X_t \right]^{-1} \\ \bar{\gamma}^{(g)} &= G_T^{(g)} \left[G_0^{-1} \gamma_0 + \sum_{t=1}^T X_t' \Omega^{-1(g-1)} \mathbf{y}_t \right] \end{aligned}$$

(ii) Draw $\Omega^{-1(g)} \sim \mathcal{W}_D \left(\rho_T, R_T^{(g)} \right)$ where

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

5. Discard the first B draws.

Note that in iteration g , $G_T^{(g)}$ and $\bar{\boldsymbol{\gamma}}^{(g)}$ are calculated using $\Omega^{-1(g-1)}$ while $R_T^{(g)}$ is calculated using $\boldsymbol{\gamma}^{(0)}$. This is because we choose to initialize the sample with a starting value $\Omega^{-1(0)}$ for the precision matrix rather than for the vector of regression coefficients.

7.1.3 Calculating the Marginal likelihood

We calculate the marginal likelihood using the method of (Chib, 1995). Let θ denote the full collection of parameters. By Bayes' Rule

$$\pi(\theta|Y_T) = \frac{\pi(\theta)f(Y_T|\theta)}{f(Y_T)}$$

where $f(Y_T)$ is the marginal likelihood, aka the marginal data density, aka the evidence. This identity holds true for *any* value of θ . In particular it holds at the posterior mean θ^* . Solving for $f(Y_T)$ and evaluating the result at θ^* , we have

$$f(Y_T) = \frac{\pi(\theta^*)f(Y_T|\theta^*)}{\pi(\theta^*|Y_T)}$$

Thus, we can express the *log* marginal likelihood as

$$\log f(Y_T) = \log \pi(\theta^*) + \log f(Y_T|\theta^*) - \log \pi(\theta^*|Y_T)$$

Specializing this to the SUR model considered above,

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

since our priors over $\boldsymbol{\gamma}$ and Ω^{-1} are independent. The (Chib, 1995) method 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.

The Contribution of the Prior Evaluating the first two terms, $\log \pi(\boldsymbol{\gamma}^*)$ and $\log \pi(\Omega^{-1*})$, is easy: these are simply the priors for $\boldsymbol{\gamma}$ and Ω^{-1} evaluated at the posterior means. We take the sample average of the Gibbs draws to approximate $\boldsymbol{\gamma}^*$ and Ω^{-1*} and evaluate the Normal and Wishart distributions at these points, with parameters given by the prior:

$$\begin{aligned} \pi(\boldsymbol{\gamma}^*) &= \mathcal{N}_p(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}_0, G_0) \\ \pi(\Omega^{-1*}) &= \mathcal{W}_D(\Omega^{-1*}|\rho_0, R_0) \end{aligned}$$

The Contribution of the Likelihood Above we assumed a normal distribution for the regression errors, specifically, $\boldsymbol{\varepsilon}_t | \mathbf{x}_t \sim \text{iid } \mathcal{N}_D(0, \Omega)$. From the regression specification it follows that $\mathbf{y}_t \sim \text{iid } \mathcal{N}_D(X_t \boldsymbol{\gamma}, \Omega)$ and thus the log likelihood evaluated at the posterior mean is

$$\log f(Y_T | \boldsymbol{\gamma}^*, \Omega^{-1*}) = \sum_{t=1}^T \log \mathcal{N}_D(\mathbf{y}_t | X_t \boldsymbol{\gamma}^*, \Omega^{-1*})$$

parameterized in terms of the precision matrix rather than the covariance matrix. Equivalently, but more conveniently, we may write

$$\log f(Y_T | \boldsymbol{\gamma}^*, \Omega^{-1*}) = \sum_{t=1}^T \log \mathcal{N}_D(\mathbf{y}_t - X_t \boldsymbol{\gamma}^* | \mathbf{0}, \Omega^{-1*})$$

The advantage of this version of the likelihood is that the parameters of the normal density are constant over t , allowing us to exploit the efficient algorithm for repeatedly evaluating a MV normal density with fixed parameters, described above. Note that we can simultaneously calculate all of the arguments for the normal density as follows:

$$(\tilde{Y} - \tilde{X} \Gamma^*)' = (\boldsymbol{\varepsilon}^*)' = \begin{bmatrix} \boldsymbol{\varepsilon}_1^* & \dots & \boldsymbol{\varepsilon}_T^* \end{bmatrix}$$

where $\boldsymbol{\varepsilon}_t^* = \mathbf{y}_t - X_t \boldsymbol{\gamma}^*$ and $\Gamma^* = (\boldsymbol{\gamma}_1^*, \dots, \boldsymbol{\gamma}_D^*)$.

The Contribution of the Posterior To evaluate the third term, we factorize the joint posterior as the product of a conditional and marginal, namely:

$$\pi(\boldsymbol{\gamma}^*, \Omega^{-1*} | Y_T) = \pi(\boldsymbol{\gamma}^* | \Omega^{-1*}, Y_T) \times \pi(\Omega^{-1*} | Y_T)$$

so that we have

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

Because we have *analytical expressions* for the conditional posteriors in this model we can evaluate the first term in the product immediately. We have $\boldsymbol{\gamma} | \Omega^{-1} \sim \mathcal{N}_p(\boldsymbol{\gamma} | \bar{\boldsymbol{\gamma}}, G_T)$ where G_T and $\bar{\boldsymbol{\gamma}}$ depend only on the prior, the data, and Ω^{-1} . To perform the required calculation, we simply evaluate the normal density at $\boldsymbol{\gamma}^*$ and evaluate G_T and $\bar{\boldsymbol{\gamma}}$ at Ω^{-1*} , that is:

$$\pi(\boldsymbol{\gamma}^* | \Omega^{-1*}, Y_T) = \mathcal{N}_p(\boldsymbol{\gamma}^* | \bar{\boldsymbol{\gamma}}^*, G_T^{-1*})$$

where

$$\begin{aligned} G_T^{-1*} &= \left[G_0^{-1} + \sum_{t=1}^T X_t' \Omega^{-1*} X_t \right] = \left[G_0^{-1} + \Omega^{-1*} \otimes \tilde{X}' \tilde{X} \right] \\ \bar{\boldsymbol{\gamma}}^* &= G_T^* \left[G_0^{-1} \boldsymbol{\gamma}_0 + \sum_{t=1}^T X_t' \Omega^{-1*} \mathbf{y}_t \right] = \text{solve} \left[G_T^{-1*}, G_0^{-1} \boldsymbol{\gamma}_0 + \text{vec}(\tilde{X}' \tilde{Y} \Omega^{-1*}) \right] \end{aligned}$$

The evaluation of the second term in the product that gives the contribution of the posterior to the marginal likelihood is a bit more involved. We write

$$\begin{aligned}\pi(\Omega^{-1*}|Y_T) &= \int \pi(\gamma, \Omega^{-1*}|Y_T) d\gamma \\ &= \int \pi(\Omega^{-1*}|\gamma, Y_T) \pi(\gamma|Y_T) d\gamma\end{aligned}$$

and approximate the second integral using the draws from the Gibbs sampler:

$$\begin{aligned}\pi(\Omega^{-1*}|Y_T) &\approx \frac{1}{G} \sum_{g=1}^G \pi(\Omega^{-1*}|\gamma^{(g)}, Y_T) \\ &= \frac{1}{G} \sum_{g=1}^G \mathcal{W}_D(\Omega^{-1*} | \rho_0 + T, R_T^{(g)})\end{aligned}$$

where

$$\begin{aligned}R_T^{(g)} &= \left[R_0^{-1} + \sum_{t=1}^T (\mathbf{y}_t - X_t \gamma^{(g)}) (\mathbf{y}_t - X_t \gamma^{(g)})' \right]^{-1} \\ &= \left[R_0^{-1} + (\tilde{Y} - \tilde{X} \Gamma^{(g)})' (\tilde{Y} - \tilde{X} \Gamma^{(g)}) \right]^{-1}\end{aligned}$$