On Flexible Linear Factor Stochastic Volatility Models

Malefaki, Valia

University of Athens

January 2015

On Flexible Linear Factor Stochastic Volatility Models

Valia Malefaki

January 2015

Abstract In this thesis I discuss flexible Bayesian treatment of the linear factor stochastic volatility model with latent factors, which proves to be essential in order to preserve parsimony when the number of cross section in the data grows. Based on the Bayesian model selection literature, I introduce a flexible prior specification which allows carrying out restriction search on the mean equation coefficients of the factor model – the loadings matrix. I use this restriction search as a data-based alternative to evaluate the cross sectional restrictions suggested by arbitrage pricing theory. A mixture innovation model is also proposed which generalizes the standard stochastic volatility specification and can also be interpreted as a restriction search in variance equation parameters. I comment on how to use the mixture innovation model to catch both gradual and abrupt changes in the stochastic evolution of the covariance matrix of high-dimensional financial datasets. This approach has the additional advantages of dating when large jumps in volatility have occurred in the data and determining whether these jumps are attributed to any of the factors, the innovation errors, or combinations of those.

1) Introduction

Linear factor models are of paramount importance in empirical finance with main applications in asset pricing. They are also used as a method of treating parsimoniously multivariate financial data especially in settings where the factor covariances are allowed to be heteroskedastic, usually following GARCH-type or stochastic volatility specifications. The purpose of this paper is to apply efficient Bayesian algorithms which can be used for parsimonious representation of linear factor models and for testing restrictions on their parameters. Unlike studies in finance that treat the factors as observable data, the factors here are assumed to be latent (as in Ross (1976, 1977)) and need to be estimated along with the model parameters.

The first of the proposed algorithms is inspired from the Bayesian model selection literature and in particular from the works of George and McCullogh (1993, 1997). This involves a stochastic search variable selection (SSVS) algorithm, which is used extensively in univariate regressions to search the model space for the most probable models, i.e. models that have regression coefficients different than zero with high probability. This algorithm is based on a prior which is a mixture of two Normal densities, one that shrinks the posterior values of the parameter towards zero and one that leaves the parameter unrestricted. Similar priors have been used in factor models by Cremers (2002), Avramov (2002) and Ericsson and Karlsson (2004). These studies treat the factors as observables (like Fama-French factors) and one task here is to show that Bayesian model selection priors can be easily generalized in the case of unobserved factors. Nevertheless, the most important feature of the proposed prior specification is that restriction search is implemented on each element of the mean equation coefficient matrices. Hence in a factor model with many asset returns, each factor is allowed to explain or not each individual return, unlike Ericsson and Karlsson (2004) where a factor either explains all of the returns or not. Following the suggestions of George and McCullogh (1997) and Chipman et al. (2001), it is straightforward to show that the restriction search algorithm is fairly automatic and simple to use, while at the same time is an attractive data-based alternative to the cross sectional restrictions suggested by the Arbitrage Pricing Theory.

At a second stage the log volatilities evolve as autoregressions as in Pitt and Shephard (199?), but with innovations that are mixtures of two normal components. The standard stochastic volatility specification, assumes that the log-volatilities change smoothly at each point in time following an AR(1) specification. In the extension proposed here, log-volatilities vary according to a random walk

with an error that is a mixture of two Normal components: one with unrestricted variance estimated by the data, and one with variance zero. Note that this modelling approach restricts the volatility equation parameters, since if for a time period the zero variance component is drawn from the data, the log-volatility remains equal to the value of the previous time period. The methods used to estimate the mixture innovation in the log-volatility equation is based on the recent algorithm for dynamic mixtures of Gerlach et al. (2000).

Both extensions induce flexibility as many different specifications may be nested without increasing substantially the computational cost. The output of the stochastic search algorithm gives the posterior belief (i.e. data-based probability) that a certain variable loads on each of the latent factors. The mixture innovation approach has the additional advantages of dating when large jumps in volatility have occurred in the data and determining whether these jumps are attributed to any of the factor or innovation errors. Also, different jump patterns are allowed for the volatility of each of the innovation and factor errors. The two extensions combined result in a parsimonious representation of the factor stochastic volatility model, especially when the number of assets is large or, equivalently, when the number of time series observations is restricted. The SSVS prior restricts the mean equation parameters (factor loadings) by shrinking some of their elements towards zero, while the mixture innovation specification restricts the variance equation by restricting the volatility to drift at each point in time.

It has to be clarified at this point that both algorithms will implement restriction in a data-based fashion. That is, the data are restricted according to some probability rules that are updated by the likelihood. I show that it is easy to adopt priors that have minimal influence to the posterior densities of the parameters, but also priors that reflect strong beliefs about which parameters should be restricted (for example, according to some theory or the researchers' experience).

2) Factor Models for Asset Returns

i) The linear factor model

I begin by specifying the standard linear factor model which will be the building block for all the proposed extensions in the analysis. Let $y_t = \{y_{1t}, \ldots, y_{pt}\}'$ denote the observations on p financial

variables (assets) at time $t = 1,\ldots,n$, which conditioned on k unobserved factors $f_t = (f_{1t},\ldots,f_{kt})'$ follow a linear model of the form:

$$y_t = \alpha + \beta f_t + \varepsilon_t \tag{0.1}$$

where β is the $[p \times k]$ matrix of factor loadings, $f_t \sim N(0, I_k)$, $\varepsilon_t \sim N(0, \Sigma)$ and $\Sigma = diag(\sigma_1^2,\ldots,\sigma_n^2)$. Along with the condition of independence between the idiosyncratic errors and the factors, $E(f_t\varepsilon_t') = 0$, these are the main assumptions of the factor model retained in modern treatments like Geweke and Zhou (1996) and Lopes and West (2004). Unlike principal components analysis which recovers the true factors up to a rotational indeterminacy (Chamberlain and Rothchild (1983) and Connor and Korajczyk (1988)), likelihood-based estimation of the factors requires to set identification restrictions, usually on the loadings matrix β. Following the standard practice in the literature (for example Geweke and Zhou (1996), Aguilar and West (2001)) I set $\beta_{ij} = 0$ for $j > i$ and $\beta_{ii} > 0$ for $i \le k$; for more discussion of these issues see Lopes and West (2004).The factor model indeed offers a parsimonious representation for multivariate time series as its implication is that the covariance of the series of interest, y_t, is decomposed as:

$$\Sigma_y = \beta\Sigma_f\beta' + \Sigma_\varepsilon \tag{0.2}$$

where $\Sigma_y, \Sigma_f, \Sigma_\varepsilon$ is the covariance matrix of y_t, f_t and ε_t respectively, so that in the case of the homoscedastic factor model (1.1), $\Omega = \text{var}(y_t \mid \beta, \Sigma) = \beta\beta' + \Sigma$. According to this formula in the case of, say, $p = 50$ and $k = 5$ there are only $(p \times k) + p = 305$ elements to estimate in the conditional covariance matrix Ω (corresponding to the number of elements in β plus the number of elements in Σ), while the full unconditional covariance matrix of y has $p(p+1)/2 = 1275$ elements.

Dependent on the application of the factor model different assumptions can be made about the constant. Estimates of α resulting either from maximum likelihood or the Bayesian posterior mean under a non-informative prior, are equal to the sample mean $\bar{y} = \sum_{t=1}^{n} y_t$. Hence, in many occasions, for the sake of brevity the data are demeaned and no intercept is estimated. In other instances a restriction may be imposed on α, like in the competitive equilibrium version of the arbitrage pricing

theory (APT) where the exact linear factor pricing holds in the economy (see Connor (1984)). The factor model for p asset returns, $r_t = \{r_{1t}, \ldots, r_{pt}\}'$, has the following form, adopted from equation (1.1)

$$r_t = E(r_t) + \beta f_t + \varepsilon_t \tag{0.3}$$

where $E(\square)$ denotes mathematical expectation and hence in terms of equation (1.1), $a = E(r_t)$. Now the competitive equilibrium version of the APT (Connor and Korajczyk (1988)) implies the restriction on $a = E(r_t)$

$$E(r_t) = r_t^F + \beta\gamma \tag{0.4}$$

where r_t^F is the risk-free rate of return and γ are the risk premia which for now are assumed to be time-invariant. The equilibrium version of the APT implies a factor model for excess asset returns (i.e. in excess of the risk-free rate r_t^F)

$$\begin{aligned} \tilde{r}_t &= r_t - r_t^F \\ &= \beta\gamma + \beta f_t + \varepsilon_t \\ &= \beta(\gamma + f_t) + \varepsilon_t \end{aligned} \tag{0.5}$$

The factors f_t are mean zero, but if we set $f_t^* = \gamma + f_t$ it turns out that $f_t^* \square N(\gamma, I_k)$ so that γ can be interpreted as the factor means.

ii) The linear factor stochastic volatility model

Recently, Pitt and Shephard (1999), Aguilar and West (2001) and Chib, Nardari and Shephard (2005) relax the hypothesis of homoscedasticity in the factor model. This is accomplished by allowing the diagonal covariance matrices on the factor and idiosyncratic errors to be time-varying. Their formulation is

$$y_t = \alpha + \beta f_t + \varepsilon_t \tag{0.6}$$

$$\begin{pmatrix} \varepsilon_t \\ f_t \end{pmatrix} \sim N\left\{0, \begin{pmatrix} V_t & 0 \\ 0 & D_t \end{pmatrix}\right\} \tag{0.7}$$

where

$$V_t = V_t(h_t) = diag\{\exp(h_{1,t}),\ldots,\exp(h_{p,t})\}$$
$$D_t = D_t(h_t) = diag\{\exp(h_{p+1,t}),\ldots,\exp(h_{p+k,t})\} \quad (0.8)$$

and h_t is the $[(p+k)\times 1]$ vector of latent log-variances, each following independent stochastic volatility processes for $j = 1,\ldots,p+k$:

$$h_{j,t} = \kappa_j + \varphi_j(h_{j,t-1} - \kappa_j) + \sigma_j \eta_{j,t}^h \quad (0.9)$$

with $\eta_{j,t} \sim N(0,1)$ and $E(\eta_{j,t}\eta_{l,t}) = 0$ for $l \neq j$. This is the simple factor stochastic volatility (henceforth FSV) model, sometimes also called simply the multivariate stochastic volatility model as it is a parsimonious extension of its univariate counterpart. I follow the standard convention in the literature (see for example Chib, Nardari and Shephard (2005)), and impose the identification and normalization restriction that $\beta_{ij} = 0$ for $j > i$ and $\beta_{ii} = 1$ for $i \leq k$. Observe that in the case of the factor stochastic volatility model, we have the following decomposition of the covariance matrix of our data $\Omega_t = \text{var}(y_t \mid \beta, V_t, D_t) = \beta D_t \beta' + V_t$.

3) Parsimony in factor models

It is reasonable to assume that there is no immediate need to preserve degrees of freedom in the already parsimonious linear factor and factor stochastic volatility representations of the covariance matrix of the observed data. However, when the data grow large the number of free parameters grow large, an issue more evident in the factor stochastic volatility model. Additionally, unless an estimate of these elements is exactly zero, every variable is allowed to load on every factor. Hence I employ a data-based stochastic search for parameter restrictions inspired by the Bayesian variable selection literature for univariate regressions. As it was mentioned in the introduction, each element of the $[p\times k]$ loadings matrix is subject to restriction search, hence the number of all possible models is $2^{p\times k}$ (an element is zero or not). West (2003), in the context of sparse factor modeling, was the first to discuss the possibility of using priors on the elements β_{ij}, $i = 1,\ldots,p$, $j = 1,\ldots,k$, of β that induce zeros with high probability. Nowadays it is a standard practice to use these priors in the demanding task of modeling the factor structure in high dimensional DNA microarray gene expression data where

scientists have to analyze only a handful of observations on thousands of genes ($n = 49$ and $p = 6128$ in the application of West, 2003). In particular the prior proposed here for each β_{ij} is not based on the "slab and spike" mixture prior used in West (2003), but on the "smooth" version proposed by George and McCullogh (1993), which takes the form

$$(\beta_{ij} \mid \pi_{ij}) \sim (1-\pi_{ij})N(0, c_{ij1}^2) + \pi_{ij} N(0, c_{ij2}^2), \quad i = 1, \ldots, p, \; j = 1, \ldots, k \tag{0.10}$$

with $c_{ij1} \ll c_{ij2}$ and π_{ij} a random variable taking $0-1$ values, having the interpretation of the component weights on this mixture of Normals prior density.

The idea of selecting (c_{ij1}, c_{ij2}) is to set c_{ij1} small (close or equal to zero) and c_{ij2} large so that β_{ij} will be restricted when $\pi_{ij} = 0$ and unrestricted when $\pi_{ij} = 1$. The iterative nature of Markov Chain Monte Carlo (MCMC) simulation, and in particular the Gibbs sampler, makes the implementation of the restriction search. At each cycle of the sampler we obtain consecutively draws from the posterior of each parameter, conditional on the values of the rest parameters. Hence, when $\pi_{ij} = 0$ the prior for β_{ij} becomes $(\beta_{ij} \mid \pi_{ij}) \sim N(0, c_{ij1}^2)$ which restricts its conditional posterior (conditional on π_{ij}) to be close to the prior mean (which is zero) when $c_{ij1} \to 0$. Similarly, when $\pi_{ij} = 1$ the prior becomes $(\beta_{ij} \mid \pi_{ij}) \sim N(0, c_{ij2}^2)$ which is uninformative when $c_{ij2} \to \infty$. A semiautomatic approach would be to extract estimates of the factors $\hat{f}_t$ using some other method, like standard principal components, run a regression of the observed data on $\hat{f}_t$ and save the standard error associated with the least squares estimate of each β_{ij}, say $\hat{\sigma}_{\beta_{ij}}$. Then following the recommendations of George and McCullogh (1997), we can set $c_{ij1} = \kappa_1 \hat{\sigma}_{\beta_{ij}}$ and $c_{ij2} = \kappa_2 \hat{\sigma}_{\beta_{ij}}$ with typical values $\kappa_1 = 1/10$ and $\kappa_2 = 10$. Accordingly, a fully automatic approach is to set uninformative inverse gamma priors on (c_{ij1}, c_{ij2}). Additionally, George and Foster (2000) discuss several choices of c_{ij1} and c_{ij2} based on

empirical Bayes methods, which they prove that result in maximum posterior models which correspond to several information criteria[1].

The weights π_{ij} determine the probability of setting a restriction on a parameter and subsequently must be updated from the information in the data. For that reason a prior is set on each element of $\pi = \{\pi_{11}, \ldots, \pi_{ij}\}$, and a sensible choice for this $0-1$ random variable is the Bernoulli density:

$$p(\pi_{ij} = 1) = \rho_{ij}, \quad p(\pi_{ij} = 0) = 1 - \rho_{ij}, \tag{0.11}$$

which implies simply that $\pi \mid \rho \sim \prod_i \prod_j \rho_{ij}^{\pi_{ij}} (1-\rho_{ij})^{(1-\pi_{ij})}$ for $i = 1, \ldots, p$, $j = 1, \ldots, k$, since each π_{ij} is independent of one another. Because there is no actual uninformative choice for ρ_{ij}[2], a second hierarchical layer is placed (hyperprior) on these hyperparameters

$$\rho_{ij} \mid \underline{\lambda}_1, \underline{\lambda}_2 \sim Beta(\underline{\lambda}_1, \underline{\lambda}_2) \tag{0.12}$$

That way, the prior expectation about the value of π_{ij} is not fixed over each MCMC draw, but they are generated randomly from the Beta distribution which has mean $\underline{\lambda}_1 / (\underline{\lambda}_1 + \underline{\lambda}_2)$ and variance $\dfrac{\underline{\lambda}_1 \cdot \underline{\lambda}_2}{(\underline{\lambda}_1 + \underline{\lambda}_2)^2 \cdot (\underline{\lambda}_2 + \underline{\lambda}_2 + 1)}$.

To implement restriction search in the volatility equation, a mixture innovation process is assumed which efficiently utilizes the algorithm of Gerlach, Carter and Kohn (2001); see Koop et al. (2009) and Korobilis (2010) for applications of this method. In this case (1.9) admits the following representation:

[1] Several modifications for this prior have been proposed, see for instance George and Mc Culloch (1993), and Korobilis (2013a). Alternative priors for implementing parameter shrinkage and selection have been proposed, among others, by Kyung et al (2010), Liang et al. (2008), Maruyama and George (2010), Park and Casella (2008), and Zhou (2006).

[2] For $\rho_{ij} > 0.5$ larger models (i.e. models with less parameters restricted) are favored a priori and for $\rho_{ij} < 0.5$ smaller models are favored. The choice $\rho_{ij} = 0.5$ results in a uniform prior, but not in a prior which is non-informative about model size. The reader is referred to Chipman *et al* (2001, p. 77) where all aspects of Bayesian model selection priors are discussed in detail.

$$h_{j,t} = h_{j,t-1} + K^h_{j,t}\sigma_j \eta^h_{j,t} \tag{0.13}$$

for each $j = 1,\ldots, p+k$, and where the introduced index vectors K^h_t are assumed to be Bernoulli (0-1 values) random variables. The interpretation here is that conditional on the values of K_t, structural breaks occur in the log-variance process. A break is indicated when $K_t = 1$, so that the errors $e_{jt} = K_{j,t}\sigma_j \eta_{j,t}$ have a "large" (equal to σ_j^2) variance to allow large abrupt changes in the level of h_t, while $K_t = 0$ implies that h_t equals h_{t-1} - the value of the previous time period (and hence remains constant for that period). The mixture innovation specification nests both a model with constant volatilities when $K^h_t = 0$ for every $t = 1,\ldots,n$, or the conventional stochastic volatility model when $K^h_t = 1$ for every $t = 1,\ldots,n$. In between of those two, lies the interpretation of jumps (structural breaks) occurring in the log-volatility process when $K^h_t = 1$ only for some of the time periods t. The index variable is allowed to be different for each of the p innovation errors and for each of the k factors, thus allowing different assets and factors to have different dates for their breaks in the variance. Note that in (1.13) it is implied that $\varphi_j = 1$ and subsequently the mean κ_j is not identified and thus $\kappa_j = 0$. The random walk specification is not restrictive compared to an AR(1) specification like the one in (1.9), as it allows for complex paths of the log-variances as well. The fact that a random walk specification is explosive should not be of concern when low frequency (monthly) asset returns are used, since the process (1.13) will be in place only for a finite period of time and not forever. Otherwise, when data of higher frequencies are used, φ_j should be strictly less than one in order to prevent h_t hitting any upper or lower bound.

A complete model specification requires to place a hierarchical prior on the random variables K^h_t. As it was the case with the 0-1 variables $\pi = \{\pi_{11},\ldots,\pi_{ij}\}$, a Bernoulli prior is the natural choice for K^h_t

$$K^h \mid p \sim \prod_j \prod_t p_j^{K^h_{jt}} (1-p_j)^{(1-K^h_{jt})} \tag{0.14}$$

for $t=1,\ldots,n$ and $j=1,\ldots,p+k$. Similarly to (1.12), a second hierarchical layer in this prior is assumed. I use a Beta prior, which is the conjugate choice, for each p_j^h of the form

$$p_j^h \mid \underline{\kappa}_1, \underline{\kappa}_2 \sim Beta(\underline{\kappa}_1, \underline{\kappa}_2) \tag{0.15}$$

Each p_j^h is treated as independent, but notice that it is straightforward to account for correlation between breaks in different blocks of parameters by assuming dependence between the K's, contemporaneously and/or at leads and lags.

To sum up, two different algorithms are considered here for parsimonious representation of factor stochastic volatility models. One pertains to the loadings matrix and one to the factor's and innovation's covariance matrices. These algorithms can also be used in many other instances of the factor model. In many cases it is rational to assume stochastic mean returns and factor loadings, i.e. $a = a_{t|t-1}$ and $\beta = \beta_{t|t-1}$. For example in the context of the equilibrium version of the APT in (1.5), there is a large literature that suggests that the 'betas' (β) and the risk premia (γ) should be time-varying. Examples of work in this area are Jagannathan and Wang (1996), Hodrick and Zhang (2001), Lettau and Ludvigson (2001), Cochrane (2001) and Jostova and Philipov (2004). In that respect, these studies suggest that (following the formulation of Cochrane (2001)) for $i=1,...,p$ and $t=1,\ldots,n$:

$$\begin{aligned} \gamma_{it}(z_{t-1}) &= \Gamma z_{t-1} \\ \beta_{it}(z_{t-1}) &= b_{i0} + b_{i1} z_{t-1} \end{aligned} \tag{0.16}$$

where z_{t-1} is a $[q \times 1]$ vector of macroeconomic and financial (explanatory) variables. Equation (1.5) now takes the form

$$\begin{aligned} \tilde{r}_t &= \beta_t(\gamma_t + f_t) + \varepsilon_t \\ &= (b_0 + b_1 z_{t-1})(\Gamma z_{t-1} + f_t) + \varepsilon_t \end{aligned} \tag{0.17}$$

In other factor model specifications for asset prices (Adrian and Franzoni (2005)), exchange rates (Lopes and Carvalho (2006)) or large macro datasets (Stock and Watson (2007), DelNegro and Otrok (2008) and Korobilis (2008)) the assumption that each row of the loadings matrix evolves as random walk is adopted. This implies the following form

$$\beta_{i,t} = \beta_{i,t-1} + Q_i \eta_{i,t}^{\beta} \tag{0.18}$$

with Q_i a $k \times k$ covariance matrix and $\eta_t^{\beta} \sim N(0, I_k)$. Notice that with the additional assumption of time varying loadings, $\Omega_t = \text{var}(y_t \mid \beta_t, V_t, D_t) = \beta_t D_t \beta_t' + V_t$. Whether we assume formulation (1.16) or (1.18), the two algorithms can be used in a straightforward way. This flexibility in the specification of β may prove to be of paramount importance as, according to Ghysels (1998), misspecification of the time-varying nature may result in larger pricing errors than assuming simply a constant loadings matrix. This is a task which is left for future examination.

4) Model selection and Bayesian model averaging

Model comparison in general factor models, is feasible using marginal likelihoods. In the simple factor model (eq. (1.1)), Lopes and West (2004), review six popular methods to estimate the marginal likelihood. For the factor model with stochastic volatility (eq. (1.6) – (1.8)), Chib, Nardari and Shephard (2006) rely on the method of Chib (1995) and Chib and Jeliakov (2001). Using Bayes theorem the logarithm of the marginal likelihood is

$$\ln p(y \mid M) = \ln p(y \mid \theta, M) + \ln p(\theta \mid M) - \ln p(\theta \mid y, M)$$

where $\ln p(y \mid M)$ is the log-marginal likelihood, $\ln p(y \mid \theta, M)$ is the log likelihood, $\ln p(\theta \mid M)$ is the prior density of the parameters θ and $\ln p(\theta \mid y, M)$ is the posterior density of the parameters. All the densities above are conditional on a certain model specification, model M. The expression above if estimated for specific value of θ of high posterior density (e.g. posterior mean or mode), gives an accurate approximation to the marginal likelihood of a model M. Model comparison then relies

Using the full prior specification for the loadings, it is easy to show that model selection can be implemented using the average of the posterior probabilities $\bar{\pi}_{ij} = E(\pi_{ij} \mid data)$, where $p(\pi_{ij} \mid data)$ is the posterior distribution of the probabilities π_{ij} conditional on the 'data'. The posterior of $p(\pi_{ij} \mid data)$ is a sequence of 'zeros' and 'ones' obtained from the MCMC sampler, so that the average of the samples from the posterior is exactly the quantity $\bar{\pi}_{ij}$, which is interpreted as 'the probability that the

respective parameter β_{ij} is zero or not". Hence, using the finding of Barbieri and Berger (2004) that the median probability model is the optimal among all possible model specifications, the best fitting model will be the one with parameters β_{ij} that have respective probability $\bar{\pi}_{ij}$ greater than 0.5. This means that we have to re-estimate the factor model, this time imposing restrictions on the parameters which had low probability on the first MCMC run. As an alternative to model selection, in many occasions it is preferable to average across possible model specifications, so as to reduce model (specification) uncertainty (Cremers, (2002), Hoeting et al. (1999)). This is called Bayesian model averaging and has many advantages, especially when we want to use the model to extrapolate the data in the future (forecasting, value-at-risk). The average model is obtained directly from the output of the restriction search algorithm, and the mean probabilities $\bar{\pi}_{ij}$. A probability of, say, 0.6 initially says that this parameter was used in 60% of the draws from the posterior, while the rest 40% its value was zero. This is automatically done for each of the $p \times k$ elements of the loadings matrix, meaning that averaging is done across all 2^{pk} possible model specifications. If we were to use the model in, say, forecasting, the forecast would be the BMA forecast, as each parameter β_{ij} is contributes $\bar{\pi}_{ij}\%$ in constructing the final forecast.

5) Monte Carlo simulation

In this Section we perform a small Monte Carlo experiment in order to assess the suitability of the different proposed algorithms. We follow Lopes and West (2004) and simulate observations from a static factor model with parameters:

$$\beta' = \begin{pmatrix} 0.99 & 0 & 0 & 0.99 & 0.99 & 0 & 0 & 0 & 0 \\ 0 & 0.95 & 0 & 0 & 0 & 0.95 & 0.95 & 0 & 0 \\ 0 & 0 & 0.90 & 0 & 0 & 0 & 0 & 0.90 & 0.90 \end{pmatrix},$$

and

$$\Sigma = diag(0.02, 0.19, 0.36, 0.02, 0.02, 0.19, 0.19, 0.36, 0.36).$$

In particular, we simulate 1000 samples with $n = 50, p = 9, k = 3$. We use MCMC to estimate all 1,000 datasets and we evaluate the average posterior mean of the posterior probabilities of inclusion. These probabilities are

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \\ \pi_{41} & \pi_{42} & \pi_{43} \\ \pi_{51} & \pi_{52} & \pi_{53} \\ \pi_{61} & \pi_{62} & \pi_{63} \\ \pi_{71} & \pi_{72} & \pi_{73} \\ \pi_{81} & \pi_{82} & \pi_{83} \\ \pi_{91} & \pi_{83} & \pi_{93} \end{pmatrix} = \begin{pmatrix} 1 & - & - \\ .019 & 1 & - \\ .0063 & .048 & 1 \\ 1 & .0031 & .003 \\ 1 & .003 & .003 \\ .0363 & .9921 & .0037 \\ .0077 & .9999 & .0039 \\ .0059 & .0436 & .7949 \\ .0046 & .0121 & .9486 \end{pmatrix}$$

which indicate that the SSVS algorithm is working very well. In red color we can see the where the algorithm hasn't performed optimally, due to possible correlation in the data, and the fact that the 1,000 samples we have generated are very short ($n = 50$).

As a second experiment we generate from a larger factor stochastic volatility model with *n=350, p=24 & k=6*, which is of the form

$$y_t = \alpha + \beta f_t + \varepsilon_t$$
$$h_{j,t} = \kappa_j + \varphi_j (h_{j,t-1} - \kappa_j) + \sigma_j \eta_{j,t}^h$$

$\kappa_j = 0.9$, $\phi_j = 0.9$ and $\sigma_j = \sqrt{0.2}$.

In the prior $p(\pi_{ij} = 1) = \rho_{ij}$, $p(\pi_{ij} = 0) = 1 - \rho_{ij}$, I try values $\rho = 0.5$ ('informative') and $\rho = 0.85$ (impose a priori many restrictions). The loadings used in this experiment are shown in the table below:

Table 1: Coefficients used to generate from the factor stochastic volatility model					
		β			
1	*0*	*0*	*0*	*0*	*0*
0	*1*	*0*	*0*	*0*	*0*
0	*0*	*1*	*0*	*0*	*0*
0	*0*	*0*	*1*	*0*	*0*
0	*0*	*0*	*0*	*1*	*0*
0	*0*	*0*	*0*	*0*	*1*
0.99	*0*	*0*	*0*	*0*	*0*
0.99	*0*	*0*	*0*	*0*	*0*
0.99	*0*	*0*	*0*	*0*	*0*
0	*0.99*	*0*	*0*	*0*	*0*
0	*0.99*	*0*	*0*	*0*	*0*
0	*0.99*	*0*	*0*	*0*	*0*
0	*0*	*0.99*	*0*	*0*	*0*
0	*0*	*0.99*	*0*	*0*	*0*
0	*0*	*0.99*	*0*	*0*	*0*
0	*0*	*0*	*0.99*	*0*	*0*
0	*0*	*0*	*0.99*	*0*	*0*
0	*0*	*0*	*0.99*	*0*	*0*
0	*0*	*0*	*0*	*0.99*	*0*
0	*0*	*0*	*0*	*0.99*	*0*
0	*0*	*0*	*0*	*0.99*	*0*
0	*0*	*0*	*0*	*0*	*0.99*
0	*0*	*0*	*0*	*0*	*0.99*
0	*0*	*0*	*0*	*0*	*0.99*

The estimated probabilities under the different values of ρ are shown in Tables 2 and 3 below.

Table 2: Posterior probabilities, $\rho = 0.85$					
		π			
–	–	–	–	–	–
0.008	–	–	–	–	–
0.0053	*0.006*	–	–	–	–
0.0033	*0.0043*	*0.0043*	–	–	–
0.0053	*0.008*	*0.004*	*0.0217*	–	–
0.0033	*0.007*	*0.005*	*0.0053*	*0.0037*	–
1	*0.0047*	*0.0043*	*0.0077*	*0.0043*	*0.0123*
1	*0.0063*	*0.004*	*0.0037*	*0.0037*	*0.006*
1	*0.0043*	*0.0103*	*0.0077*	*0.0043*	*0.0033*
0.0443	*1*	*0.004*	*0.0113*	*0.005*	*0.0017*
0.006	*1*	*0.004*	*0.002*	*0.0077*	*0.0033*
0.004	*1*	*0.0053*	*0.0053*	*0.0077*	*0.0017*
0.0143	*0.0083*	*1*	*0.0057*	*0.0017*	*0.005*
0.0183	*0.0043*	*1*	*0.0047*	*0.0033*	*0.0043*
0.0167	*0.0063*	*1*	*0.0037*	*0.0043*	*0.004*
0.027	*0.007*	*0.0057*	*1*	*0.0063*	*0.0027*
0.003	*0.006*	*0.0063*	*1*	*0.004*	*0.0053*
0.0037	*0.0047*	*0.0047*	*1*	*0.005*	*0.0023*
0.004	*0.0017*	*0.0033*	*0.006*	*1*	*0.002*
0.0047	*0.0023*	*0.0077*	*0.004*	*1*	*0.0057*
0.006	*0.0033*	*0.0037*	*0.0057*	*1*	*0.004*
0.0043	*0.004*	*0.0057*	*0.01*	*0.1077*	*1*
0.0037	*0.0037*	*0.0227*	*0.0057*	*0.0027*	*1*
0.0042	*0.0047*	*0.0063*	*0.01*	*0.005*	*0.99*

Table 3: Posterior probabilities, $\rho = 0.50$					
		π			
–	–	–	–	–	–
0.024	–	–	–	–	–
0.0316	*0.0322*	–	–	–	–
0.0295	*0.0218*	*0.0245*	–	–	–
0.0281	*0.0431*	*0.0238*	*0.084*	–	–
0.0168	*0.035*	*0.0265*	*0.0217*	*0.0235*	–
1	*0.0278*	*0.0251*	*0.033*	*0.0215*	*0.0606*
1	*0.0271*	*0.019*	*0.027*	*0.0196*	*0.029*
1	*0.0206*	*0.0381*	*0.0266*	*0.0243*	*0.0175*
0.154	*1*	*0.0215*	*0.0492*	*0.0298*	*0.0165*
0.024	*1*	*0.0238*	*0.0238*	*0.041*	*0.0181*
0.02	*1*	*0.0265*	*0.038*	*0.0352*	*0.0195*
0.0668	*0.0443*	*1*	*0.0178*	*0.0221*	*0.0206*
0.098	*0.0253*	*1*	*0.021*	*0.0196*	*0.0253*
0.0921	*0.0236*	*1*	*0.0218*	*0.0208*	*0.0195*
0.1265	*0.0238*	*0.034*	*1*	*0.0208*	*0.0195*
0.025	*0.0355*	*0.022*	*1*	*0.024*	*0.0206*
0.021	*0.026*	*0.0257*	*1*	*0.0233*	*0.0206*
0.0183	*0.02*	*0.018*	*0.0236*	*1*	*0.0253*
0.0254	*0.0201*	*0.0251*	*0.0221*	*1*	*0.0467*
0.0217	*0.026*	*0.0226*	*0.021*	*1*	*0.0318*
0.0213	*0.0448*	*0.0215*	*0.0428*	*0.2686*	*1*
0.0318	*0.0235*	*0.0978*	*0.0245*	*0.0225*	*1*
0.0241	*0.0221*	*0.0319*	*0.041*	*0.0308*	*1*

In both cases, we can see that the performance of the proposed algorithm is excellent, accurately picking the correct restrictions every time.

6) Conclusions and further thoughts

I discuss how flexible Bayesian analysis can help enhance linear factor models for testing asset pricing theory. The proposed extensions are computationally feasible using MCMC methods. The benefit of some of the proposed algorithms has been assessed using a small Monte Carlo exercise. Further extensions can be the application of the various flexible factor models in Fama-French type portfolio data. That way, one can obtain a reduced pricing error of APT by using Bayesian model selection methods as in this study.

References

Aguilar, O. and West, M., (2000), 'Bayesian Dynamic Factor Models and Portfolio Allocation', *Journal of Business and Economic Statistics* **18**, 338–357.

Avramov, D., (2002), 'Stock Return Predictability and Model Uncertainty', *Journal of Financial Economics* **64**, 423–458.

Avramov, D., (2004), 'Stock Return Predictability and Asset Pricing Models', *Review of Financial Studies* **17**, 699–738.

Avramov, D. and Chao, J., (2006), 'An Exact Bayes Test of Asset Pricing Models with Application to International Markets', *Journal of Business* **79**, 293 – 323.

Bai, J. and Ng, S. (2007), 'Boosting Diffusion Indexes', Unpublished manuscript, Columbia University.

Belmonte, M., Koop, G. and Korobilis, D., (2014), 'Hierarchical Shrinkage in Time-Varying Parameter Models', *Journal of Forecasting* **33**, 80-94.

Chib, S., (1998), 'Estimation and Comparison of Multiple Change-Point Models', *Journal of Econometrics* **86**, 221–241.

Chib, S., Nardari, F. and Shephard, N., (2006), 'Analysis of High Dimensional Multivariate Stochastic Volatility Models', *Journal of Econometrics* **134**, 341–371.

Cremers, M., (2002), Stock Return Predictability: A Bayesian Model Selection Perspective, *Review of Financial Studies* **15**, 1223-1249.

Cremers, M., (2006), 'Multifactor Efficiency and Bayesian Inference', *Journal of Business* **79**, 2951–2998.

George, E. and McCulloch, R., (1993), 'Variable Selection via Gibbs Sampling', *Journal of the American Statistical Association* **88**, 881–889.

George, E. and McCulloch, R., (1997), 'Approaches to Bayesian Variable Selection', *Statistica Sinica* **7**, 339–379.

Gerlach, R., Carter, C. and Kohn, E., (2000), 'Efficient Bayesian Inference for Dynamic Mixture Models', *Journal of the American Statistical Association* **95**, 819–828.

Geweke, J. and Zhou., G., (1996), 'Measuring the Price of Arbitrage Pricing Theory', *Review of Financial Studies* **9**, 557–587.

Ghysels, E., (1998), 'On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?', *Journal of Finance* **53**, 549–573.

Harvey, C. and Zhou, G., (1990), 'Bayesian Inference in Asset Pricing Tests', *Journal of Financial Economics* **26**, 221–254.

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T., (1999), 'Bayesian Model Averaging: A Tutorial', *Statistical Science* **14**, 382–401.

Koop, G., Leon-Gonzalez, R., amd Strachan, R.W., (2009), 'On The Evolution Of The Monetary Policy Transmission Mechanism', *Journal of Economic Dynamics and Control* **33**, 997-1017.

Korobilis, D., (2008), 'Forecasting In Vector Autoregressions With Many Predictors', MPRA Paper 21122, University Library of Munich, Germany.

Korobilis, D., (2009), 'Assessing The Transmission Of Monetary Policy Using Dynamic Factor Models', MPRA Paper 27593, University Library of Munich, Germany, revised Nov 2010.

Korobilis, D. (2013a), 'Bayesian Forecasting with Highly Correlated Predictors', *Economics Letters* **118**, 148-150.

Korobilis, D. (2013b), 'Hierarchical Shrinkage Priors for Dynamic Regressions with Many Predictors', *International Journal of Forecasting* **29**, 43-59.

Kyung, M., Gill, J., Ghoshz, M. and Casella, G. (2010), 'Penalized Regression, Standard Errors, and Bayesian Lassos', *Bayesian Analysis* **5**, 369-412.

Liang, F., Paulo, R., Molina, G., Clyde, M. A. and Berger, J. O. (2008), 'Mixtures of g-priors for Bayesian Variable Selection', *Journal of the American Statistical Association* **103**, 410-423.

Lopes, H.F. and Carvalho, C.M., (2007), 'Factor Stochastic Volatility with Time-Varying Loadings and Markov Switching Regimes', *Journal of Statistical Planning and Inference* **137**, 3082–3091.

McCulloch, R. and Tsay, E., (1993), 'Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series', *Journal of the American Statistical Association* **88**, 968–978.

Maruyama, Y. and George, E. I. (2010), 'gBF: A Fully Bayes Factor with a Generalized g-prior', Technical Report, University of Pennsylvania, available at http://arxiv.org/abs/0801.4410

Nardari, F. and Scruggs, J., (2007), 'Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions', *Journal of Financial and Quantitative Analysis* **42**, 857 – 892.

Park, T. and Casella, G. (2008), 'The Bayesian Lasso', *Journal of the American Statistical Association* **103**, 681-686.

Pitt, M.K. and Shephard, N., (1999), 'Time Varying Covariances: A Factor Stochastic Volatility Approach' (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (Eds.), *Bayesian Statistics* **6**, Oxford University Press, Oxford, pp. 547–570.

Sentana, E. and Fiorentini, G., (2001), 'Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models', *Journal of Econometrics* **102**, 143–164.

Shanken, J., (1987), 'A Bayesian Approach to Testing Portfolio Efficiency', *Journal of Financial Economics* **19**, 195–215.

Zellner, A. (1986), 'On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions'. in: P. Goel and A. Zellner (Eds.) Bayesian Inference and Decision Techniques (North-Holland, Amsterdam).

Zou, H. (2006), 'The Adaptive Lasso and Its Oracle Properties', *Journal of the American Statistical Association* **101**, 1418-1429.

Zou, H. and Hastie, T. (2005), 'Regularization and Variable Selection via the Elastic Net', *Journal of the Royal Statistical Society, Series B* **67**, 301-320.

Appendix:

A. Posterior inference in the factor model with restriction search

Consider the simple case of the linear factor model (is straightforward to generalize the restriction search algorithm to the factor stochastic volatility model with constant loadings) in matrix form:

$$Y = F\beta' + \varepsilon$$

The Gibbs sampler iterates for $s = 1,\ldots,S$ in order to get in total S draws from the conditional densities:

Initialize β^0, π^0, Σ^0. *For* $s = 1,\ldots,S$:

i) Sample the factors F for $t = 1,\ldots,n$ from:

$$f_t^s \mid \beta^{s-1}, \Sigma^{s-1} \sim N\left(\left(I_k + \beta'\Sigma^{-1}\beta\right)^{-1}\beta'\Sigma^{-1}y_t,\ \left(I_k + \beta'\Sigma^{-1}\beta\right)^{-1}\right)$$

ii) Sample the loadings β from:

- For $i = 1,\ldots,k$

$$\beta_i^s \mid F^s, \pi^{s-1}, \Sigma^{s-1} \sim N\left(\left(\left(D_{i,1:i}D_{i,1:i}\right)^{-1} + \sigma_i^{-2}F_i'F_i\right)\sigma_i^{-2}F_i'y_i,\ \left(\left(D_{i,1:i}D_{i,1:i}\right)^{-1} + \sigma_i^{-2}F_i'F_i\right)^{-1}\right)\mathbf{1}\left(\beta_{ii} > 0\right)$$

- For $i = k+1,\ldots,p$

$$\beta_i^s \mid F^s, \pi^{s-1}, \Sigma^{s-1} \sim N\left(\left(\left(D_iD_i\right)^{-1} + \sigma_i^{-2}F'F\right)\sigma_i^{-2}F'y_i,\ \left(\left(D_iD_i\right)^{-1} + \sigma_i^{-2}F'F\right)^{-1}\right)$$

where $D_i = diag\left(d_{1,i},...,d_{k,i}\right)$ with $d_{ij} = \begin{cases} c_{1ij} & if \quad \pi_{ij} = 0 \\ c_{2ij} & if \quad \pi_{ij} = 1 \end{cases}$, for $i = 1,\ldots,p$, $j = 1,\ldots,k$, and $D_{i,1:i}$ indexes the upper $i \times i$ block of D_i, i.e. $D_{i,1:i} = diag\left(d_{1,i},...,d_{i,i}\right)$, when $i < k$ and is equal to D_i when $i = k$.

iii) Sample the mixing proportions π for $i = 1,\ldots,p$, $j = 1,\ldots,k$ from:

$$\pi_{ij}^s \mid \pi_{/-ij}^{s-1}, F^s, \beta^s, \Sigma^{s-1} \sim Bernoulli\left(\delta_{1ij}/\left(\delta_{1ij} + \delta_{2ij}\right)\right)\mathbf{1}\left(i \geq j\right)$$

with $\delta_{1ij} = \frac{1}{c_{1ij}}\exp\left\{-\left(\frac{\beta_{ij}}{2c_{1ij}}\right)^2\right\}\rho_{ij}$, $\quad \delta_{2ij} = \frac{1}{c_{2ij}}\exp\left\{-\left(\frac{\beta_{ij}}{2c_{2ij}}\right)^2\right\}\left(1-\rho_{ij}\right)$

iv) Sample the diagonal error covariance Σ for $i = 1,\ldots,p$ from:

$$(\sigma_i^s)^2 \mid F^s, \beta^s \sim iG\Big(\big(v+n\big)/2, \quad \big(va^2 + m_i\big)/2\Big)$$

where $m_i = \left(y_i - F\beta_i'\right)'\left(y_i - F\beta_i'\right)$

if $s \leq S$ *go to* i)

else if $s > S$ *stop and save draws.*

Posterior quantities of interest are now easy to calculate. For example the average posterior probabilities for model selection are:

$$E(\pi_{ij} \mid y) = \frac{1}{S} \cdot \sum_{s=1}^{S} \pi_{ij}^s$$

In the case of the model with stochastic volatility step iv) will change accordingly (see Chib et al. (2006)) and in step i) Σ is replaced with Σ_t for each $t = 1,\ldots,n$.