

# Bayesian Variable Selection and Estimation of Risk Premiums in the APT model

Rachida Ouysse \*      Robert Kohn<sup>†</sup>

September 26, 2009

## Abstract

Empirical tests of the arbitrage pricing theory using measured variables rely on the accuracy of standard inferential theory in approximating the distribution of the estimated risk premiums and factor betas. The techniques employed thus far perform factor selection and model inference sequentially. Recent advances in Bayesian variable selection are adapted to an approximate factor model to investigate the role of measured economic variables in the pricing of securities. In finite samples, exact statistical inference is carried out using posterior distributions of functions of risk premiums and factor betas. The role of the panel dimensions in posterior inference is investigated. New empirical evidence is found of time-varying risk premiums with higher and more volatile expected compensation for bearing systematic risk during contraction phases. In addition, investors are rewarded for exposure to “Economic” risk.

*JEL Classification:* C1, C22, C52

*Keywords:* Factor models, observed factors, arbitrage pricing theory, risk premiums, factor betas, Bayesian variable selection, posterior inference, maximum likelihood estimation, Markov-chain Monte Carlo.

---

\*Corresponding author, School of Economics, The University Of New South Wales, Sydney 2052 Australia. Email: rouysse@unsw.edu.au

<sup>†</sup>School of Economics, The University Of New South Wales, Sydney 2052 Australia. Email: r.kohn@unsw.edu.au

## 1 Introduction

After the documented empirical failure of the market beta to explain the cross-sectional variation in asset returns (Reinganum [1981], Stambaugh [1982], Fama and French [1992]), the arbitrage pricing theory model (APT) of [Ross, 1976] has generated an increased interest in the application of multifactor models in investigating and testing asset pricing theory. The APT has the attractive feature of making minimal assumptions about the nature of the economy. However, despite its popularity, the tractability of the APT comes at the cost of certain ambiguities such as an approximate pricing relation and an unknown set of factors. Most empirical tests of the APT with observed economic variables employ the two-pass approach of Fama and MacBeth [1973] (eg., Chen et al. [1986], Ferson and Harvey [1991]). This methodology involves (i) estimating the factor betas in a first-pass time series regression of asset returns on a given set of factors, and (ii) estimating the risk premiums by a second-pass cross-sectional regression of the asset returns on the betas estimated in the first-pass. In addition to the errors in variables problem, the two-pass approach suffers miss-specification bias due to *model risk*.

Our article provides methodological and empirical contributions to the existing literature. First, we adapt recent advances in Bayesian variable selection to an approximate factor model with observable factors. This framework handles both model and parameter uncertainty in a straightforward and formal way. The uncertainty about the factor structure is therefore embedded in the estimation of the risk premiums and the factor betas. Second, we add to the understanding of the role of the panel dimensions (number of cross sections and number of observations) in posterior inference. In particular, we investigate the usefulness of sample information in identifying the “best” factor structure. Third, we present new empirical evidence of time-varying factor risk premiums that suggests that pricing of industry and size portfolio returns is determined by systematic economic risks instead of firm-specific variables.

Much work has been done to formalize the search for latent factors using factor analysis and principal components. See for example Connor and Korajczyk [1988, 1993], Bai and Ng [2002], Bai [2003] and Galimbeti et al. [2009]. Optimization of information criteria is one class of tools successfully used in this literature. For each model considered, these criteria provide a score that accounts for the trade off between parsimony and precision. Our paper is related to the smaller literature on variable selection when the factors are observed measured variables.

Estimation and inference in observable factor models faces two challenges. Firstly, as the number of candidate variables  $K$  increases, the high-dimensionality of the model space ( $2^K$ ) poses a problem for exhaustive search tools such as optimizing information criteria. Ouyse [2006] proposes a two-step procedure where the candidate variables are first ordered using an  $R$ -squared principle. Then, information criteria are applied to the ordered set of variables to select the size of the set. The procedure is consistent under  $N, T \rightarrow \infty$  asymptotics and the search requires only  $2K$  regressions. In the context of univariate regression, Kapetanios [2007] proposes the use of simulated annealing and genetic algorithms to search efficiently the model space. Hofmann et al. [2007] reduce the search space by using a radius metric to preorder the variables inside the regression tree.

The second challenge is post-model selection inference. Standard statistical inference does not take into account the pretesting that precedes model estimation. Asymptotically, the inference is correct conditional on the model selected being consistent.

The Bayesian approach offers an alternative for exact finite sample inference by embedding both the uncertainty about the model and the parameters in the posterior distributions. Previous literature applying Bayesian methods to the study of the APT is mainly concerned with latent factors. McCulloch and Rossi [1990, 1991] develop a Bayesian framework for testing the

restrictions implied by the APT. Their approach is a two-pass procedure in which the factors are first extracted using the asymptotic principal components analysis of Connor and Korajczyk [1988] before using the posterior odds ratio to test the APT restrictions. Geweke and Zhou [1996] (hereafter GZ) are the first to employ Markov Chain Monte Carlo (MCMC) methods in a one-step approach where estimation of the latent factors and testing of the APT implications are done simultaneously. Nardari and Scruggs [2007] extend this framework to allow for heteroscedasticity and time varying expected returns. Other studies evaluating the predictability of stock market returns and efficiency of multifactor models using Bayesian inference include Shanken [1987], Harvey and Zhou [1990], Armanov [2002] and Cremers [2006].

To the best of our knowledge, this is the first study to consider Bayesian analysis of an approximate factor model with observed risk factors. Although this paper is most closely related to Geweke and Zhou [1996] and Nardari and Scruggs [2007], our approach of evaluating competing models is different. Geweke and Zhou [1996] employ a measure of the pricing error to discriminate between four competing models, whereas Nardari and Scruggs [2007] use Bayes factors to measure the ability of the model to explain the entire distribution of returns and to compare models with five factors and four stochastic volatility specifications. The pool of observable candidate factors in our study includes economic and financial variables and the number of competing models is therefore potentially large. In our paper, it is  $2^{21} = 2,097,152$  models. The computational requirement of pairwise model comparison procedures like posterior odds ratios or Bayes factors is prohibitive.

Our paper employs Bayesian variable selection and Bayesian model averaging to investigate the role of measured economic variables in the pricing of securities. The statistical analysis is related to a large literature on model choice in linear regression that is based on probabilistic fit using latent mixture modeling. See for example Mitchell and Beauchamp [1988], George and McCulloch [1993, 1997], Chipman et al. [2001], Geweke [1996]. Bayesian variable selection employs latent variables (search variables) whose posterior density encapsulates the effectiveness of different explanatory regressors in explaining the dynamics of the response variables. Our approach adapts the multivariate framework of Brown et al. [1998] to an approximate factor model to estimate and test the APT implications. We use recent advances in MCMC algorithms to approximate the posterior distribution of the latent search variables and to efficiently search the space of competing models. We make statistical inference based on Bayesian model averaging (BMA). This approach enables the construction of posterior probability intervals that take into account the variability due to model uncertainty (Leamer [1978]) and gives more reliable prediction than using a single model (Madigan and Raftery [1994]). To rank the risk factors, we use the BMA estimates of their probabilities to be in the model as measures of their importance in the pricing of the excess returns.

We now briefly summarize our results. First, we find new evidence on the role of the number of cross sections in posterior inference. Under the empirical Bayes prior, more evidence is extracted from the data with a larger number of assets but not necessarily from a longer time series with a small cross-section of assets. This is consistent with the convergence result of Chamberlain and Rothschild [1983] and Ouyse [2006].

Second, the results provide strong evidence for market reward for “Economic” risk measured by unanticipated inflation, unemployment rate and changes in industrial production. Third, we find that the risk premium associated with economic risk is time-varying. Using the time series of the posterior mean of the risk premiums associated with the factors, we argue that the compensation for economic risk is higher and more volatile during recessions and times of financial distress. Finally, the data provide little support for the Fama and French [1993] size and book-to-market factors as potential sources of rewarded risk.

We use the following notation throughout the paper:  $E(\cdot|Z_t)$  is the conditional expectation given  $Z_t$  and  $E_t(\cdot)$  is the conditional expectation given the information at time  $t$ ,  $A'$  is the transpose of  $A$ ,  $vec(A)$  is the column vectorization of  $A$ , that is if  $A = (a_1, \dots, a_n)$  then  $vec(A) = (a'_1, \dots, a'_n)'$ ,  $tr(A)$  is the sum of the diagonal elements of  $A$ , the norm of  $A$  is  $\|A\| = [tr(A'A)]^{1/2}$ ,  $A \otimes B$  is the Kronecker product of  $A$  and  $B$ , i.e., for  $A = [a_{ij}]$ ,  $A \otimes B = [a_{ij}B]$ ,  $A^{-1}$  is the inverse of  $A$ ,  $\iota_m$  is an  $m$ -vector of ones,  $I_m$  is an  $m \times m$  identity matrix,  $diag(A)$  is the vector consisting of the diagonal elements of  $A$  and, by ‘‘vector’’ we mean column vector.  $\mathcal{N}(a, B)$  is a normal distribution with mean  $a$  and covariance  $B$  and  $\mathcal{W}(\eta, \Delta)$  (resp.  $\mathcal{IW}(\eta, \Delta)$ ) is a Wishart (resp. inverse-Wishart) distribution with scale parameter  $\Delta$  and degrees of freedom  $\eta$ .

## 2 Econometric model and the APT implications

### 2.1 The pricing model

Let  $\mathbf{y}_t$  be an  $N$ -vector of security returns in *excess* of the risk free rate in period  $t$ . We assume that the excess returns have a linear factor structure:

$$\mathbf{y}_t = \alpha_N + \Lambda_N F_t + e_t, \quad t = 1, \dots, T, \quad (1)$$

$$E(e_t|F_t) = 0, \quad (2)$$

where  $F_t$  is a  $K$ -vector of common risk factors with mean  $\mu_F$  and covariance  $\Sigma_F$ ,  $\alpha_N$  is an  $N$ -vector of intercept terms,  $\Lambda_N \equiv (\lambda_1, \dots, \lambda_i, \dots, \lambda_N)'$  is an  $N \times k$  matrix of factor betas (factor loadings),  $e_t$  is an  $N$ -vector of idiosyncratic returns and the subscript  $N$  indicates that the factor structure depends on  $N$ .

We assume an approximate factor structure in the sense of Chamberlain and Rothschild [1983] and Ingersoll [1984]. The covariance matrix of the idiosyncratic returns  $\Sigma_N = E(e_t e_t')$  is not required to be diagonal for the proof of the arbitrage pricing theory (APT) of Ross [1976]. Rather we assume that the eigenvalues of  $\Sigma_N$  are bounded as  $N$  tends to infinity and  $\lim_{N \rightarrow \infty} \Lambda_N' \Lambda_N / N$  is nonsingular. This structure allows for heterogeneity (i.e., different diagonal elements in  $\Sigma_N$ ) and limited amount of cross-sectional dependence.

We assume that the returns are independently and identically distributed through time, i.e.,  $\Sigma_N$  is time invariant and there is no time series dependence. A similar covariance matrix is used in previous studies of factor models and their applications to the APT; Chamberlain and Rothschild [1983], Ingersoll [1984], McCulloch and Rossi [1991], Connor and Korajczyk [1993], Bai and Ng [2002] and Armanov [2002].

Let  $\Psi_N$  be the covariance matrix for the  $N$ -vector of excess returns,  $\mathbf{y}_t$ . The factor structure (1) implies a variance decomposition of the form

$$\Psi_N = \Lambda_N \Sigma_F \Lambda_N' + \Sigma_N. \quad (3)$$

The APT assumes the universe of all assets traded in the market and therefore we should expect convergence as  $N$  tends to infinity. The existence and uniqueness of the approximate factor structure requires that the largest  $K$  eigenvalues of  $\Psi_N$  tend to infinity as  $N \rightarrow \infty$  and the remaining  $N - K$  eigenvalues are constant; see Chamberlain and Rothschild [1983], Brown [1989] and Connor and Korajczyk [1993].

The competitive equilibrium interpretation of the APT of Connor [1984] implies an exact multifactor pricing relationships between the expected excess returns  $\mu_N = E(\mathbf{y}_t)$  and the factor betas  $\Lambda_N$  of the form

$$\mu_N = \delta_{0,N} \iota_N + \Lambda_N \delta_N. \quad (4)$$

The intercept  $\delta_{0,N}$  measures the *mispricing* with respect to the  $K$ -factor model and  $\Lambda_N\delta_N$  is the factor-related component of risk. The APT therefore provides a decomposition of the risk premium of an asset into its exposure to the risk factors (factor betas) and the associated price of risk  $\delta_N$  (risk premiums). These constraints are nonlinear because both  $\Lambda_N$  and  $\delta_N$  are unknown.

Equation (4) and the econometric model (1) imply a set of restrictions on the estimates of the intercept term in the unrestricted econometric model. Testing the APT implications is equivalent to testing the set of nonlinear relationships between the expected excess return and the factor betas,

$$\alpha_N = \delta_{0,N}\iota_N + \Lambda_N(\delta_N - \mu_F). \quad (5)$$

Without loss of generality we can take the factors to have zero means, that is  $\mu_F = 0$ . Then

$$\mu_N = \alpha_N = \delta_{0,N}\iota_N + \Lambda_N\delta_N. \quad (6)$$

In the absence of risk free arbitrage opportunities, the pricing relationship in (4) holds approximately under weaker conditions (Ross [1976], Chamberlain and Rothschild [1983]). Rather than testing the exact pricing equality, Ingersoll [1984, Theorem 1] shows that based on the absence of asymptotic arbitrage opportunities, there exists a positive number  $V$  such that the weighted sum of the squared pricing errors is uniformly bounded,

$$(\alpha_N - \Lambda_N\delta_N)' \Sigma_N^{-1} (\alpha_N - \Lambda_N\delta_N) \leq V < \infty, \text{ for all } N. \quad (7)$$

## 2.2 Pricing errors and risk premiums

Geweke and Zhou [1996] measure the closeness of the pricing approximation using an average of squared pricing errors across assets,

$$Q_N^2 = \alpha_N' (I_N - \Lambda_N(\Lambda_N' \Lambda_N)^{-1} \Lambda_N') \alpha_N / N. \quad (8)$$

The framework of Geweke and Zhou [1996] is a static factor model where the matrix  $\Sigma_N$  is diagonal. In an approximate factor structure, Ingersoll [1984] shows that a sufficient condition for the pricing error  $Q_N^2$  to be valid is that the norm of  $\Sigma_N$  satisfies  $\|\Sigma_N\| < \infty$ ; which holds when  $N \rightarrow \infty$ . For fixed  $N$  however, achieving uncorrelated pricing errors requires using many factors than are actually needed for the pricing implication (7) to hold; Ingersoll [1984], Chamberlain and Rothschild [1983]. Our article therefore considers

$$\tilde{Q}_N^2 = (\alpha_N - \Lambda_N\delta_N)' \Sigma_N^{-1} (\alpha_N - \Lambda_N\delta_N) / N, \quad (9)$$

a covariance weighted measure of pricing error which follows directly from the no arbitrage condition (7) of Ingersoll [1984].

The APT pricing restriction (4) implies that any predictability of returns is driven by changes in the betas  $\Lambda_N$ , and changes in the expected risk premiums  $\delta_N$ . In the variance decomposition (3), the predictable variation in stock returns is related to the predictable variance captured by the factor model. Security returns are predictable only to the extent that expected returns are related to the risk factors (predictor variables). We follow Ferson and Korajczyk [1995] and use the variance ratio

$$\text{Var}1_i = \frac{\text{var}(\lambda_i' F_t)}{\text{var}(y_{it})}, \quad (10)$$

to measure the predictable variance in the excess returns that is attributed to the factor model.

The average risk premiums  $\delta_N$  minimize (9) given  $\alpha_N$  and  $\Lambda_N$ . The standard approach of running cross-sectional regressions (Shanken [1992], Robotti and Balduzzi [2008]) of the expected excess returns on the factor betas leads to the average risk premiums coefficients

$$\tilde{\delta}_N^* = \left( \tilde{\Lambda}'_N \Sigma_N^{-1} \tilde{\Lambda}_N \right)^{-1} \tilde{\Lambda}'_N \Sigma_N^{-1} \alpha_N, \quad (11)$$

where  $\tilde{\Lambda}_N = [\iota_N, \Lambda_N]$  and  $\tilde{\delta}_N^* = (\delta_{0,N}, \delta'_N)'$ . See Robotti and Balduzzi [2008] for a generalized method of moments interpretation of the risk premiums coefficients.

Equation (4) states the APT implications in terms of the unconditional expected returns. The pricing relation can be made flexible to allow for time-varying equilibrium expected returns. Stambaugh [1983] and Connor and Korajczyk [1989] derive the time-versions of the APT with conditioning information, in which the risk premiums are time-varying and the factor betas are fixed parameters. The no arbitrage condition at the equilibrium is derived with respect to the investor's information at time  $t$  and implies an approximate conditional pricing relation,

$$E_t(\mathbf{y}_{t+1}) \approx \delta_{0,N,t} + \Lambda_N \delta_{N,t}, \quad i = 1, \dots, N, t = 1, \dots, T-1, \quad (12)$$

where  $\delta_{N,t}$  and  $\delta_{0,N,t}$  are the realized market prices of systematic risks and the mispricing for the cross-section of  $N$  returns at time  $t$ , respectively.

For each month  $t$ , the time-varying risk premiums  $\delta_{N,t}$  are obtained from a cross-sectional projection of the *ex post* asset excess returns on the *ex ante* betas; Fama and MacBeth [1973], Ferson and Harvey [1991]. The assumption of constant factor betas in the data generating process (1), leads to the time-varying risk premiums coefficients

$$\tilde{\delta}_{N,t}^* = \left( \tilde{\Lambda}'_N \Sigma_N^{-1} \tilde{\Lambda}_N \right)^{-1} \tilde{\Lambda}'_N \Sigma_N^{-1} \mathbf{y}_t, \quad (13)$$

where  $\tilde{\delta}_{N,t}^* = (\delta_{0,N,t}, \delta'_{N,t})'$ .

Equations (11) and (13) are conditional on  $\Lambda_N$  and  $\Sigma_N$  known. Approaches to estimating these parameters include time series regressions of excess returns on the observed factors (Fama and MacBeth [1973], Shanken [1992]), generalized method of moments estimation applied to the conditional moments (1) and (6) (Robotti and Balduzzi [2008]), and maximum likelihood estimation (Campbell et al. [1997]).

For the rest of the paper, we drop the subscript  $N$  and use the notation  $\alpha$ ,  $\Lambda$ ,  $\delta_0$ ,  $\delta$  and  $\Sigma$  for ease of exposition.

### 3 Bayesian framework

Our methods are based on work on Bayesian variable selection and multivariate regression. See Evans [1965], Mitchell and Beauchamp [1988], George and McCulloch [1993, 1997], Smith and Kohn [1996], Fernandez et al. [2001] and Cripps et al. [2005]. Stricklanda et al. [2009] discusses Bayesian analysis in the related but different context of a multivariate state space model.

The econometric model in (1) can be rewritten as

$$\mathbf{y} = (I_N \otimes \iota_T) \alpha + (I_N \otimes F) \lambda + \epsilon, \quad (14)$$

where,  $\mathbf{y} = \text{vec}(\mathbf{Y})$ ,  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_t, \dots, \mathbf{y}_T)'$ ,  $\lambda = \text{vec}(\Lambda')$ ,  $F = (F_1, \dots, F_T)'$ ,  $\epsilon = \text{vec}((e_1, \dots, e_T)')$ ,  $\iota_T$  is a  $T$ -vector of ones and  $I_N$  is an  $N \times N$  identity matrix.

The factors entering the data generating process of the returns are unknown but are assumed to be elements of a finite set of potential variables  $\mathbf{X}$ . Let  $K$  be the total number of potential

variables represented by the columns of the matrix  $\mathbf{X}$  and assume that there exists a “true” factor structure with factors  $\mathbf{X}^0$  which defines the generating process of excess returns.

We express factor selection as a variable selection problem using a vector of indicator variables  $\gamma$ . Define the Bernoulli random variable  $\gamma_j$  as

$$\gamma_j = \begin{cases} 1 & \text{if } X_j \in X^0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore  $\gamma = \{\gamma_j, j = 0, 1, \dots, K\}$  is a selector vector over the columns of  $X = (X_0, X_1, \dots, X_K)$ , where  $X_0 = \iota_T$ . Let  $q_\gamma$  be the number of covariates included in the model,  $q_\gamma = \gamma_0 + \gamma_1 + \dots + \gamma_K$ . Adopting this notation, we can write (14) under model  $\gamma$  as

$$\underset{NT \times 1}{\mathbf{y}} = \underset{NT \times Nq_\gamma}{(I_N \otimes \mathbf{X}_\gamma)} \underset{Nq_\gamma \times 1}{\beta_\gamma} + \underset{NT \times 1}{\epsilon}, \quad (15)$$

where  $\mathbf{X} = (\iota_T, F)$ ,  $\beta = (\alpha', \lambda')'$  and the subscript  $\gamma$  indicates that only columns and elements with the corresponding  $\gamma$  element being 1 are included. Since  $\gamma$  is a binary sequence, the number of models to be evaluated is  $2^K$ .

Our article carries out model selection and model estimation simultaneously with inference about  $\gamma$  done with model parameters integrated out. The posterior density of  $\gamma$  conditional on the observed excess returns is

$$p(\gamma|\mathbf{y}, \mathbf{X}) = \frac{p(\mathbf{y}|\gamma, \mathbf{X})p(\gamma)}{\sum_\gamma p(\mathbf{y}|\gamma, \mathbf{X})p(\gamma)} \propto p(\gamma)p(\mathbf{y}|\gamma, \mathbf{X}), \quad (16)$$

where  $p(\gamma)$  is the prior on  $\gamma$  and  $p(\mathbf{y}|\gamma, \mathbf{X})$  is the marginal likelihood of the observed data under model  $\gamma$ , with

$$p(\mathbf{y}|\gamma, \mathbf{X}) = \int_\Sigma \int_\beta p(\mathbf{y}|\beta, \Sigma, \gamma, \mathbf{X})p(\beta, \Sigma, \gamma)d\beta d\Sigma. \quad (17)$$

In the rest of the paper we drop the fixed design matrix  $\mathbf{X}$  from the set of conditioning variables for ease of exposition.

### 3.1 Priors formulation

A hierarchical Bayes formulation of a variable selection prior is (George and McCulloch [1997])

$$p(\beta, \gamma, \Sigma) = p(\beta|\Sigma, \gamma)p(\Sigma|\gamma)p(\gamma). \quad (18)$$

A commonly used prior for  $\gamma$  is

$$p(\gamma) = \prod_{j=1}^K \pi^{\gamma_j} (1 - \pi)^{(1-\gamma_j)},$$

with  $\pi$  prespecified. The number of factors  $q_\gamma$  in the pricing relationship thus follows a binomial distribution. We follow Fernandez et al. [2001] and choose  $\pi = 0.5$  implying that  $p(\gamma) = 2^{-K}$ : so the expected model size is  $K/2$  and the standard deviation is  $\sqrt{K}/4$ . This prior allows each variable to be in or out of the model independently with the same probability  $1/2$ . If a smaller (bigger) value of  $\pi$  is prespecified, then smaller (larger) models are preferred a priori. To allow for mispricing in the APT restriction (6), we force the model to have an intercept by setting  $\gamma_0 = 1$ .

We can make the prior for  $\gamma$  more flexible by putting a prior probability on  $\pi$ . See for example, Brown et al. [1998], Nott and Kohn [2005] and Ley and Steel [2008].

We choose the following Normal inverse-Wishart conjugate prior for  $\beta$  and  $\Sigma$ ,

$$\beta|\Sigma \sim \mathcal{N}(\beta_0, \Sigma \otimes H), \quad (19)$$

$$\Sigma \sim \mathcal{IW}(m, \Phi), \quad (20)$$

where  $\Phi$  is an  $N \times N$  scale parameter,  $m > N + 1$  is a shape parameter and  $H$  is an  $K \times K$  nonsingular matrix.

To implement variable selection, it is necessary to specify the hyperparameters  $\beta_0$ ,  $H$ ,  $m$  and  $\Phi$ . Assuming that no subjective information about these parameters is available, their values are set to minimize their influence. For the parameters entering the prior on  $\Sigma$  we set  $m = N + 2$ , which reflects a minimum amount of prior information and  $\Phi = \hat{\Sigma} + s^2 I_N$ , where  $\hat{\Sigma}$  is the maximum likelihood estimator for  $\Sigma$  in the regression (1) and  $s^2$  is

$$s^2 = \frac{(\mathbf{y} - (I_N \otimes \mathbf{X})\hat{\beta})' (\mathbf{y} - (I_N \otimes \mathbf{X})\hat{\beta})}{NT};$$

$\hat{\beta}$  is the maximum likelihood estimate of  $\Sigma$  in the pooled regression (15) for  $\gamma = \iota_K$ . The term  $s^2$  is added to deal with rank deficient cases. We follow Brown et al. [2002] and choose  $\beta_0 = 0$ .

The covariance matrix  $\Sigma \otimes H$  in the prior (19) separates out the cross-sectional correlation from the correlation of the risk factors. The matrix  $H$  determines the amount of information in the prior and influences the covariance structure in the posterior distributions of the parameters of the model. Our choice is  $H = c(\mathbf{X}'\mathbf{X})^{-1}$  which is motivated by Brown et al. [1998] and is an extension of the conjugate g-prior of Zellner [1986] to the multivariate regression model.

In the Bayesian analysis of the univariate regression model, different values of  $c$  are recommended depending on the application and the choice of the optimality criterion. There is an asymptotic correspondence between fixed choices of  $c$  and the penalized sum-of-squares (classical) information criteria, see George and Foster [2000] and Chipman et al. [2001]. The case  $H = c(\mathbf{X}'\mathbf{X})^{-1}$  and  $c = T$  corresponds to the so called *unit information prior* which has the same amount of information about  $\beta$  as that contained in one observation. This prior leads to Bayes factors with asymptotic behavior similar to the Bayesian information criterion (BIC). The *risk information prior* (RIC) is obtained for  $c = K^2$  (Donoho and Johnstone [1994]). A conjugate g-prior with fixed  $c \cong 3.92$  corresponds asymptotically to Akaike's AIC. As  $c \rightarrow \infty$ , the penalty for dimension goes to infinity and the model size goes to zero, George and Foster [2000].

The larger the value of  $c$ , the more diffuse (flatter) is the prior over the region of plausible values of  $\beta$ . The value of  $c$  should be large enough to reduce the prior influence. However, excessively large values can generate a form of the Bartlett-Lindley paradox by increasing the probability on the null model as  $c \rightarrow \infty$ .

Our study considers three choices of  $c$ . The first choice  $c = 4$  approximately corresponds to the AIC. The second choice  $c = \max\{T, K^2\}$  is recommended by Fernandez et al. [2001] and is considered by Liang et al. [2008] as a bridging prior between the BIC and RIC. The third choice is the local empirical Bayes prior of George and Foster [2000],

$$\hat{c}_\gamma^{EB} = \max\{F_\gamma - 1, 0\}, \text{ where } F_\gamma = \frac{R_\gamma^2/q_\gamma}{(1 - R_\gamma^2)/(n - 1 - q_\gamma)},$$

and  $R_\gamma^2$  is the  $R$ -squared of the regression of  $\mathbf{y}$  on the covariates of the model  $\gamma$ , see Liang et al. [2008]. We adapt this definition to the multivariate case by using an F statistic for testing  $\beta_{l,\gamma} = 0, l = N + 1, \dots, Nq_\gamma$ ,

$$\hat{F}_\gamma = \frac{\mathbf{y}' \left( \hat{\Sigma}_\gamma^{-1} \otimes P_{\mathbf{X}_\gamma} \right) \mathbf{y} / N(q_\gamma - 1)}{\mathbf{y}' \left( \hat{\Sigma}_\gamma^{-1} \otimes (I_T - P_{\mathbf{X}_\gamma}) \right) \mathbf{y} / N(T - q_\gamma)}, \quad P_{\mathbf{X}_\gamma} = \mathbf{X}_\gamma (\mathbf{X}_\gamma' \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma', \quad (21)$$



and

$$\widehat{\Sigma}_\gamma = \frac{\mathbf{Y}'(I_T - P_{\mathbf{X}_\gamma})\mathbf{Y}}{T}, \quad (22)$$

is an estimate of  $\Sigma$  under model  $\gamma$ .

An empirical Bayes estimate  $\widehat{c}_\gamma^{EB}$  for  $c$  is required for each model  $\gamma$  which makes  $c$  model ( $\gamma$ ) dependent. For the remainder of the paper we use the notation  $c_\gamma$  to indicate that  $c$  may be dependent on model  $\gamma$ .

### 3.2 Posterior inference

Lemma 1 presents the results used in the MCMC sampling.

**Lemma 1** *Under the prior for  $\gamma$  with  $\pi = 0.5$  and the Normal inverse-Wishart priors (19) and (20), the full conditionals of the model parameters are*

1.  $\beta|\mathbf{y}, \Sigma, \gamma \sim \mathcal{N}(\widetilde{\beta}_\gamma, \Sigma \otimes D_\gamma^{-1})$  where,  $D_\gamma = (\mathbf{X}'_\gamma \mathbf{X}_\gamma + H_\gamma^{-1})$ ,  $H_\gamma = c_\gamma(\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1}$  and

$$\widetilde{\beta}_\gamma = (I_N \otimes D_\gamma^{-1} H_\gamma^{-1}) \beta_0 + (I_N \otimes D_\gamma^{-1} \mathbf{X}'_\gamma \mathbf{X}_\gamma) \widehat{\beta}_\gamma, \quad (23)$$

$$\widehat{\beta}_\gamma = [I_N \otimes (\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1} \mathbf{X}'_\gamma] \mathbf{y}. \quad (24)$$

2.  $\Sigma|\mathbf{y}, \gamma \sim \mathcal{IW}(m + T, S_\gamma + \Phi)$  where, for  $B_0$  such that  $\beta_0 = \text{vec}(B_0)$ ,

$$S_\gamma = (Y - B_0 W_\gamma)' (I_T - \mathbf{X}_\gamma D_\gamma^{-1} \mathbf{X}'_\gamma) (Y - B_0 W) + B'_0 (\Sigma^{-1} \otimes V_\gamma) B_0, \quad (25)$$

$$W_\gamma = [I_T - \mathbf{X}_\gamma D_\gamma^{-1} \mathbf{X}'_\gamma]^{-1} \mathbf{X} D_\gamma^{-1} H_\gamma^{-1} \quad (26)$$

and

$$V_\gamma = H_\gamma^{-1} - H_\gamma^{-1} D_\gamma^{-1} H_\gamma^{-1} - H_\gamma^{-1} D_\gamma^{-1} \mathbf{X}'_\gamma W_\gamma. \quad (27)$$

3.  $p(\gamma|\mathbf{y}) \propto |H|^{-\frac{N}{2}} |\mathbf{X}'_\gamma \mathbf{X}_\gamma + H_\gamma^{-1}|^{-\frac{N}{2}} |\Phi|^{\frac{m}{2}} |\Phi + S_\gamma|^{-\frac{(T+m)}{2}}$ .

The results are obtained similarly to Smith and Kohn [1996] and Brown et al. [1998]. For  $H_\gamma = c_\gamma(\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1}$  equation (23) becomes

$$\widetilde{\beta}_\gamma = (1 - \eta_\gamma) \beta_0 + \eta_\gamma \widehat{\beta}_\gamma, \quad \text{where } \eta_\gamma = \frac{c_\gamma}{1 + c_\gamma}. \quad (28)$$

The posterior mean  $\widetilde{\beta}_\gamma$  of model  $\gamma$  shrinks the maximum likelihood estimator  $\widehat{\beta}_\gamma$  of model  $\gamma$  towards  $\beta_0$ . The term  $\eta_\gamma$  can be interpreted as the relative importance or weight that is given to the sample information relative to the prior information. It is therefore important to assess the sensitivity of posterior inference to the specification of the hyperparameter  $c_\gamma$ , especially for the empirical Bayes prior where  $c_\gamma$  is estimated from the data.

If in addition we assume that  $\beta_0 = 0$ , then the conditional densities in Lemma (1) become

$$p(\gamma|\mathbf{y}) \propto (1 + c_\gamma)^{-\frac{Nq_\gamma}{2}} |\Phi|^{\frac{m}{2}} |\Phi + S_\gamma|^{-\frac{(T+m)}{2}}, \quad (29)$$

$$\Sigma|\mathbf{y}, \gamma \sim \mathcal{IW}(m + T, S_\gamma + \Phi), \quad (30)$$

$$\beta|\mathbf{y}, \Sigma, \gamma \sim \mathcal{N}\left(\widetilde{\beta}_\gamma, \frac{c_\gamma}{1 + c_\gamma} \Sigma \otimes (\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1}\right), \quad (31)$$

where

$$S_\gamma = Y' \left[ I - \frac{c_\gamma}{1 + c_\gamma} \mathbf{X}_\gamma (\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1} \mathbf{X}'_\gamma \right] Y, \quad \widetilde{\beta}_\gamma = \frac{c_\gamma}{1 + c_\gamma} \widehat{\beta}_\gamma. \quad (32)$$

### 3.3 Metropolis-Hastings sampling scheme

Markov Chain Monte Carlo (MCMC) methods are used to simulate from  $p(\gamma|\mathbf{y})$ . In Bayesian variable selection, Metropolis-Hastings algorithms are used to construct a Markov chain with stationary distribution  $p(\gamma|\mathbf{y})$ . We use the Metropolis-Hastings scheme proposed by Kohn et al. [2001] (Kohn-Smith-Chan) which efficiently handles the high-dimensionality of the state space and minimizes the algorithm's visits to useless predictors. The Kohn-Smith-Chan proceeds as follows.

1. Initialize  $\gamma^{(0)}$ .
2. For  $i = 1, \dots, M$ ,
  - (a) Generate a random permutation  $j = (j_1, \dots, j_k)$  of  $(1, \dots, K)$ .
  - (b) For  $l = 1, \dots, K$ ,
    - i. Generate a proposal value  $\gamma_{j_l}^P$  for  $\gamma_{j_l}$  from the conditional prior distribution for  $\gamma_{j_l}$ ,

$$p\left(\gamma_{j_l} | \gamma_{j_k, k < l}^{(i+1)}, \gamma_j^P, \gamma_{j_k, k > l}^{(i)}\right).$$

- ii. With  $\gamma^C = (\gamma_{j_k, k < l}^{(i+1)}, \gamma_j^i, \gamma_{j_k, k > l}^{(i)})$  and  $\gamma^P = (\gamma_{j_k, k < l}^{(i+1)}, \gamma_j^P, \gamma_{j_k, k > l}^{(i)})$ , set  $\gamma_{j_l}^{(i+1)} = \gamma_{j_l}^{(P)}$  with probability  $\min(1, \pi)$ , where

$$\pi = \frac{p(\mathbf{y}|\gamma^P)}{p(\mathbf{y}|\gamma^C)} \quad \text{and} \quad \log(\pi) = -\frac{N}{2} \log(1 + c) - \frac{(T + m)}{2} \log \frac{|S_{\gamma(1)} + \Phi|}{|S_{\gamma(0)} + \Phi|},$$

and set  $\gamma_{j_l}^{(i+1)} = \gamma_{j_l}^{(i)}$  otherwise.

This scheme generates iterates for  $\gamma^{(j)}, j = 1, \dots, M$ . Given  $\gamma^{(j)}$ , the iterates  $\beta_{\gamma^{(j)}}$  and  $\Sigma_{\gamma^{(j)}}$  are generated from their conditionals in (30) and (31) but are not part of the Metropolis-Hastings sampling scheme. That is, generating  $\beta$  and  $\Sigma$  does not affect the efficiency of the sampling scheme.

### 3.4 Bayesian model averaging

Determining model uncertainty is a complex problem. Bayesian model averaging (BMA) provides a formal way of handling inference in the presence of multiple competing models. In BMA the posterior distributions of quantities of interest are obtained as mixtures of the model-specific distributions weighted by the posterior model probabilities, Clyde [1999]. This approach enables construction of posterior probability intervals that take into account variability due to model uncertainty and gives more reliable prediction than using a single model (Madigan and Raftery [1994]).

Suppose that  $\theta$  is a quantity of interest that has similar interpretation in each model. The BMA posterior distribution of  $\theta$  is a weighted average of its model specific posterior distributions, where the weights are the posterior model probabilities

$$p(\theta|\mathbf{y}) = \sum_{\gamma} p(\theta|\mathbf{y}, \gamma) p(\gamma|\mathbf{y}), \quad (33)$$

and the BMA point estimate of  $\theta$  is

$$\hat{\theta}_{BMA} = E(\theta|\mathbf{y}) = \sum_{\gamma} E(\theta|\mathbf{y}, \gamma) p(\gamma|\mathbf{y}). \quad (34)$$

Implementation of (34) is difficult because the sum over the  $2^K$  possible models is impractical when  $K$  is large. One approach to get around this difficulty is to use MCMC and the simulated Markov chain from the posterior distribution  $p(\gamma|\mathbf{y}); \gamma^{(j)}, j = 1, \dots, M$ . Under suitable regularity conditions (Smith and Roberts [1993]), the posterior mean

$$\hat{\theta}_{pm} = \frac{1}{M} \sum_{j=1}^M E(\theta|\gamma^{(j)}, \mathbf{y}), \quad (35)$$

is a consistent estimate of  $E(\theta|\mathbf{y})$ . If the analytical expression of  $E(\theta|\gamma^{(j)}, \mathbf{y})$  is not available then (34) is approximated using

$$\hat{\theta}_{pm} = \frac{1}{M} \sum_{j=1}^M \theta^{(j)}, \quad (36)$$

where  $\theta^{(j)}$  is the quantity  $\theta$  under model  $\gamma^{(j)}$ . We use the posterior mean estimate

$$\hat{\beta}_{pm} = \frac{1}{M} \sum_{j=1}^M \beta^{(j)},$$

to approximate the BMA estimate of  $\beta$ . Similarly, the posterior mean estimate  $\hat{\Sigma}_{pm}$  is obtained as the sample mean of the MCMC draws  $\Sigma^{(j)}, j = 1, \dots, M$ . The BMA estimates  $Q_{pm}^2$  and  $\tilde{Q}_{pm}^2$  of the pricing errors (8) and (9), and any function of  $\gamma$  are obtained by calculating the appropriate function at each draw and averaging.

In particular, the posterior mean estimates of the average risk premiums  $\hat{\delta}_{pm}$  and the time-varying risk premiums  $\hat{\delta}_{t,pm}$  are obtained as the sample averages of the iterates  $\hat{\delta}^{(j)}$  and  $\hat{\delta}_t^{(j)}$  of the risk premiums coefficients defined in (11) and (13), respectively.

Bayesian model averaging can be used to rank the risk factors in order of their posterior probabilities to be in the factor structure and to estimate the number of risk factors in the pricing relation. The BMA estimate of the probability that a risk factor  $k$  is in the factor structure is

$$\hat{\pi}_k = \frac{1}{M} \sum_{j=1}^M \gamma_k^{(j)}, \quad (37)$$

and the posterior average dimension of the factor structure is

$$\hat{q}_{pm} = \frac{1}{M} \sum_{j=1}^M q_{\gamma}^{(j)}, \text{ where } q_{\gamma}^{(j)} = \sum_{k=1}^K \gamma_k^{(j)}. \quad (38)$$

We use the posterior probabilities  $\hat{\pi}_k$  to measure the importance of each risk factor in the pricing of the excess returns.

To assess the predictive ability of the BMA forecasts we evaluate their out-of-sample forecast accuracy. We divide the data series into two periods. The first is an *estimation period* with final time  $T$  and sample data  $X$  and  $\mathbf{y}$ . The second is a *prediction period* with forecast horizon  $s$ . For  $t = T + 1, \dots, T + s$ , let  $\mathbf{x}_t^{out}$  be the  $K$ -vector of observations on the risk factors and  $\mathbf{y}_t^{out}$  be the  $N$ -vector of excess returns observed for period  $t$ . Assume that at time  $T$ , we observe  $\mathbf{X}^{out} = [\mathbf{x}_{T+1}^{out}, \dots, \mathbf{x}_{T+s}^{out}]'$ . We generate forecasts of  $Y^{out} = [\mathbf{y}_{T+1}^{out}, \dots, \mathbf{y}_{T+s}^{out}]'$  conditional on the information available at time  $T$ . Under model  $\gamma$  and specification (14),

$$\mathbf{y}^{out} = (I_N \otimes \mathbf{X}_{\gamma}^{out}) \beta_{\gamma} + \epsilon^{out}, \quad (39)$$

where  $\epsilon^{out} \sim \mathcal{N}(\mathbf{0}, \Sigma \otimes I_s)$  and  $\mathbf{y}^{out} = \text{vec}(Y^{out})$ . The BMA estimate of the posterior predictive distribution of  $\mathbf{y}^{out}$ , conditional on  $\mathbf{y}, \mathbf{X}$  and  $\mathbf{X}^{out}$ , is (Brown et al. [1998])

$$p(\mathbf{y}^{out} | \mathbf{y}, \mathbf{X}, \mathbf{X}^{out}) = \sum_{\gamma} p(\mathbf{y}^{out} | \mathbf{y}, \mathbf{X}, \mathbf{X}^{out}, \gamma) p(\gamma | \mathbf{y}, \mathbf{X}). \quad (40)$$

The BMA estimate of  $\mathbf{y}^{out}$ , defined as the expected value of the density in (40), is

$$\hat{\mathbf{y}}^{out} = \sum_{\gamma} (I_N \otimes \mathbf{X}_{\gamma}^{out}) \tilde{\beta}_{\gamma} p(\gamma | \mathbf{y}, \mathbf{X}). \quad (41)$$

Because the forecast origin  $T$  is fixed, there is no recursive updating of the conditioning information in (40). Therefore, forecasts for time  $T + h, 1 \leq h \leq s$ , do not have future information on  $\mathbf{y}$  between  $T + 1$  and  $T + h - 1$ .

Table 1: Standard errors  $s_{\hat{\beta}_{pm,m}}$  and posterior standard deviations  $\hat{\sigma}_{\hat{\beta}_{pm,m}}$  of the BMA estimate  $\hat{\beta}_{pm,m}$  corresponding to the minimum, median and maximum inefficiencies  $\tau_{\hat{\beta}_{pm,m}}$ . The prior is  $c_\gamma = \hat{c}_\gamma^{EB}$ , the number of MCMC iterations is  $M = 50,000$  and the bandwidth  $L = 500$ .

		T=168			T=528		
N		43	93	136	43	93	136
minimum	$\tau_{\hat{\beta}_{pm,m}}$	1.56	2.12	1.51	1.77	1.50	1.62
	$\hat{\sigma}_{\hat{\beta}_{pm,m}}$	0.0102	0.0096	0.0187	0.0098	0.0096	0.0097
	$s_{\hat{\beta}_{pm,m}}$	5.69E-05	6.25E-05	0.0001	5.83E-05	5.25E-05	5.52E-05
median	$\tau_{\hat{\beta}_{pm,m}}$	4.88	7.91	8.13	6.38	5.05	3.44
	$\hat{\sigma}_{\hat{\beta}_{pm,m}}$	0.0110	0.0098	0.0194	0.0105	0.0098	0.0100
	$s_{\hat{\beta}_{pm,m}}$	1.08E-03	1.23E-03	2.47E-03	1.18E-03	9.84E-05	8.29E-05
maximum	$\tau_{\hat{\beta}_{pm,m}}$	18.37	9.60	9.93	17.36	31.63	27.62
	$\hat{\sigma}_{\hat{\beta}_{pm,m}}$	0.0120	0.0101	0.0205	0.0150	0.0101	0.0103
	$s_{\hat{\beta}_{pm,m}}$	2.30E-03	1.39E-03	2.88E-03	2.79E-03	2.54E-03	2.42E-03

Using the MCMC Markov chain  $\{\gamma^{(j)}, j = 1, \dots, M\}$ , the quantity in (41) is approximated by the posterior mean forecast

$$\hat{\mathbf{y}}_{pm}^{out} = \frac{1}{M} \sum_{j=1}^M (I_N \otimes \mathbf{X}_{\gamma^{(j)}}^{out}) \tilde{\beta}_{\gamma^{(j)}}, \quad (42)$$

where  $\tilde{\beta}_{\gamma^{(j)}}$  is determined using (32) and (24). The matrix  $\mathbf{X}_{\gamma^{(j)}}^{out}$  depends on the *estimation sample* through  $\gamma^{(j)}$  and is formed by selecting the columns of  $\mathbf{X}^{out}$  corresponding to  $\gamma_k^{(j)} = 1$ .

As a measure of the overall out-of-sample forecast performance, we define a mean-squared forecast error (MSFE) for the cross-section of  $N$  excess returns. For  $h = T + 1, \dots, T + s$ , the posterior mean forecast for the  $N$  vector of excess returns  $\mathbf{y}_h^{out}$  is

$$\hat{\mathbf{y}}_{pm,T}^{out} = \frac{1}{M} \sum_{j=1}^M (I_N \otimes \mathbf{x}_{\gamma^{(j)},h}^{out}) \tilde{\beta}_{\gamma^{(j)}}, \quad (43)$$

where  $\mathbf{x}_{\gamma^{(j)},h}^{out}$  is row  $h$  of  $\mathbf{X}_{\gamma^{(j)}}^{out}$ . We consider a weighted average of the forecast errors,  $\mathbf{y}_h^{out} - \hat{\mathbf{y}}_{pm,h}^{out}$ , and use the following measure of forecast performance,

$$MSFE = \frac{1}{Ns} \sum_{h=T+1}^{T+s} (\mathbf{y}_h^{out} - \hat{\mathbf{y}}_{pm,h}^{out})' \hat{\Sigma}_{pm}^{-1} (\mathbf{y}_h^{out} - \hat{\mathbf{y}}_{pm,h}^{out}). \quad (44)$$

MSFE is a weighted sum of squared forecast errors motivated by a generalized least squares principle for the pooled regression model (39). The forecast errors are normalized by the estimated error covariance matrix to account for cross-sectional correlation. In the empirical results, we report the values of  $RMSE = \sqrt{MSFE}$ .

### 3.5 Convergence of the MCMC sampler

In our application, we use an MCMC burn-in period of 200,000 iterations and a sampling period  $M = 50,000$ . Our approach is not strict in determining convergence and we are satisfied if four MCMC runs starting from different points arrive at broadly similar marginal distributions. The starting points for

Table 2: The candidate risk factors and the NBER business cycle dates

Economic Variables	Indices & size-BE/ME portfolios		Business Cycle Dates		
			Peak	Trough	
January Effect Dummy	JAN	Market portfolio	MARKET	August 1957	April 1958
Consumption	CG	Small Minus Big	SMB	April 1960	February 1961
Term Structure	UTS	High Minus Low	HML	December 1969	November 1970
Risk Premium	URP	Momentum Factor	MOMT	November 1973	March 1975
Expected Inflation	EI	Value weighted return	VWRET	January 1980	July 1980
Unexpected Inflation	UEI	Equally weighted return	EWRET	July 1981	November 198
Change in EI	DEI	Standard and Poor	SP500	July 1990	March 1991
Monthly Production Growth	MP			March 2001	November 2001
Annual Production Growth	YP			December 2007	
Unemployment rate	UNEM				
Change in Producer Price	PPI				
Change in Oil Price	OG				
Money Growth	GB				
Private savings	SAVE				

these runs are: (1)  $\gamma_0 = 1$  and all  $\gamma_k = 0$  for  $k = 1, \dots, K$ , (2) all  $\gamma_k = 1$ , (3) first half equal to 1 and the rest equal to 0 and (4)  $\gamma_0=1$ , the first half equal to 0 and the second half equal to 1. The results reported in the empirical section correspond to the first set of starting values. In the literature (Geweke [1992], Chib [2001]), the inefficiency factor

$$\tau = 1 + 2 \sum_{l=1}^L \left(1 - \frac{l}{M}\right) \hat{\rho}_l,$$

is used to assess how well the MCMC chain mixes. The term  $\hat{\rho}_l$  is the sample autocorrelation at lag  $l$  calculated from the MCMC sampled values. We choose a bandwidth  $L = 500$ . The inefficiency factor  $\tau$  is the factor by which we have to increase the run length of the MCMC sampler compared to independent and identically distributed (iid) sampling to obtain the same accuracy. The standard error of the BMA estimator  $\hat{\beta}_{pm,m}$  for the parameter  $\beta_m, m = 1, \dots, NK$  is estimated by

$$s_{\hat{\beta}_{pm,m}} = \hat{\sigma}_{\hat{\beta}_{pm,m}} \frac{\sqrt{\tau_{\hat{\beta}_{pm,m}}}}{\sqrt{M}},$$

where  $\hat{\sigma}_{\hat{\beta}_{pm,m}}$  is the posterior standard deviation of the iterates of  $\tilde{\beta}_{pm,m}^{(j)}, j = 1, \dots, M$ .

Table 1 reports the standard errors  $s_{\hat{\beta}_{pm,m}}$  and posterior mean standard deviations  $\hat{\sigma}_{\hat{\beta}_{pm,m}}$  corresponding to the BMA estimates  $\hat{\beta}_{pm,m}$  with the minimum, median and maximum inefficiencies  $\tau_{\hat{\beta}_{pm,m}}$ .

## 4 Data description

### 4.1 The Economic Risk factors

We use the variables listed in Table 2 and discussed below to proxy for the potential sources of economy wide risk. The data are monthly observations for two periods: January 1960 to December 2003 with  $T = 528$  and January 1990 to December 2003 with  $T = 167$ . For out-of-sample forecast analysis, we use the 24 monthly observations from January 2004 to December 2006.

*Measured economic variables.* Chen et al. [1986] argue that systematic risk factors are variables that change the discount factor and the expected cash flows. Real and nominal forces account for the changes in cash flows. Changes in the discount factor are related to changes in the marginal utility of wealth. The measured economic variables we use are the growth rate of real per capita personal

consumption expenditures for nondurable goods (CG), unanticipated changes in the price-level (UEI) and changes in the expected rate of inflation (DEI) which affects the nominal level of the expected cash flow and the interest rate. In addition, we use the average inflation rate (EI) to account for its effect on asset valuation through changes in relative prices. Industrial production is related to capacity utilization and is considered as a coincident indicator, meaning that changes in its level usually reflect similar changes in overall economic activity. Chen et al. [1986] use both monthly and yearly changes in the industrial production reflecting both the contemporaneous short term effect on stock returns and the long term anticipated changes in the industrial activity. Unanticipated changes in risk premiums reflecting movements in the risk of corporate default are measured by the difference between the return on corporate bonds rated Baa and the return on long term U.S. government bonds (URP). To capture unanticipated changes in the return on long government bonds, Chen et al. [1986] use the term structure which is the difference between the return on the long bonds and the risk free rate. See Chen et al. [1986] for definitions of variables and data sources. In addition to the variables used by Chen et al. [1986], we use unemployment rate (UNEM), changes in the producer price index (PPI), money growth (GB) computed as changes in the money base and the growth of private saving (SAVE). The data for these variables are from the St. Louis federal reserve bank economic data (FRED).

As a proxy for the market index we consider the three market indices, the equally weighted market return (EWRET), the value weighted market return (VWRET) and the return on the Standard and Poors index (SP500). The data for the market indices are from the Wharton Research Data Services (WRDS).

*Variables based on firm characteristics.* Fama and French [1993, 1995, 1996] argue that firm characteristics such as size (ME, stock price times the number of shares) and book to market equity (BE/ME, the ratio of the book value of a firm's common stock, BE, to its market value, ME), are related to the economic fundamentals and therefore can be used to proxy for economic market risk. They find evidence for both size and BE/ME effect and propose a three-factor model with the factors, (i) the excess return on a broad market portfolio, (ii) the difference between the return on a portfolio of high-BE/ME stocks and the return on a portfolio of low-BE/ME stocks (HML) and, (iii) the difference between the return on a portfolio of small size stocks and the return on a portfolio of large size stocks (SMB). Jegadeesh and Titman [1993] reports that stocks with higher returns in the previous 12 months tend to have higher future returns than stocks with lower returns in the previous months. The momentum factor is constructed to capture this predictable pattern in excess returns. The Fama and French size and book-to-market factors and the momentum factor are available from the Kenneth French data library at <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>.

The correlation matrix of the 21 factors shows that some factors are highly correlated. The correlation between expected inflation, unexpected inflation and, the change in expected inflation ranges from 0.93 to 0.99. These inflation variables are also highly correlated with the return on the value weighted portfolio (including dividends). There is also a relatively high correlation of 0.80 between monthly industrial production growth and the book-to-market factor HML, and a mild correlation of 0.66 between HML and consumption growth.

To study the time series dynamics of the risk premiums, we use the National Bureau of Economic Research (NBER) dating of the business cycle to compare the behavior of the estimated monthly risk premiums over the months of contraction and expansion. Panel C of Table 2 reports the dates of the peaks and the troughs of the business cycle.

## 4.2 The returns

We study monthly returns, in excess of the risk free rate, on three types of portfolios of NYSE, AMEX and NADAQ firms formed on the basis of (1) firm characteristics, (2) industrial classification and, (3) market capitalization.

1. *Size and book-to-market portfolios.* The 100 portfolios formed on the basis of firm characteristics are intersections of 10 common stock portfolios formed from sorting the stocks on size (ME) and 10 common stock portfolios formed from sorting the stocks on book-to-market equity (BE/ME). For example, the first portfolio contains the stocks in the low-ME group that are also in the

low-BE/ME group. See Fama and French [1993] for a detailed description of to construct the intersection portfolios.

2. *Industry portfolios.* The 12 and 48 industry portfolios of NYSE, AMEX, and NASDAQ firms are grouped by two-digit and four-digit standard industrial classification (SIC) respectively. Each NYSE, AMEX, and NASDAQ stock is assigned to an industry portfolio based on its COMPUSTAT SIC codes.
3. *Decile portfolios.* Ten common stock portfolios formed according to size deciles on the basis of market capitalization are from the Center for Research in Security Prices (CRSP). These “size” portfolios are value-weighted averages of the firms approximating a “buy and hold” strategy. See Ferson and Harvey [1991].

The data for the 100 size-BE/ME portfolios and the 12 and 48 industry portfolios are from the Kenneth French data library. After accounting for missing data points, we are left with 43 industry portfolios and 93 size-BE/ME portfolios. The cross-section sizes we consider are:  $N = 10$  for the decile portfolios,  $N = 12$  and  $N = 43$  for the industry sorted portfolios,  $N = 93$  for the size-BE/ME portfolios and  $N = 136$  for the combination of the last two cross sections.

## 5 Sensitivity of posterior inference to the priors and to the time and cross-sectional dimensions

Section 3.1 discusses some of the literature about the sensitivity of the inference to the choice of  $c$  in Bayesian univariate regression model. This section assesses the sensitivity of the posterior mean of model dimension to changes in the sample data, through increases in the number of cross sections and/or the number of observations, and to the choice of the hyperparameter  $c_\gamma$ .

Table 3 reports the BMA point estimates  $\hat{q}_{pm}$  of the model dimension and the estimate  $\hat{c}_{pm}^{EB}$  of the model empirical Bayes prior  $c_\gamma^{EB}$ . The shrinkage parameter for  $c_\gamma = c_\gamma^{EB}$  measures the importance given to the sample information over the prior information under model  $\gamma$  and is estimated by its average over the MCMC draws

$$\hat{\eta}_{pm}^{EB} = \frac{1}{M} \sum_{j=1}^M \hat{\eta}_\gamma^{(j)}, \text{ and, } \hat{\eta}_\gamma^{(j)} = \frac{\hat{c}_\gamma^{EB}}{1 + \hat{c}_\gamma^{EB}},$$

where  $\hat{c}_\gamma^{EB} = \max\{\hat{F}_\gamma, 0\}$  and  $\hat{F}_\gamma$  is defined in (21).

### 5.1 Sensitivity of $\hat{\eta}_{pm}^{EB}$ to the panel dimensions $N$ and $T$

Panel A of Table 3 reports the results for the empirical Bayes prior where we take  $c = \hat{c}_\gamma^{EB}$ . Two main patterns emerge from analyzing the results.

First, for fixed values of  $N$ , increasing the sample size from  $T = 168$  to  $T = 528$  results in large increases in the value of  $\hat{c}_{pm}^{EB}$  whereas it has no effect on  $\hat{\eta}_{pm}^{EB}$ . This suggests that the usefulness of the sample data relative to the prior is determined by the number of cross sections and not by the number of observations. Unless there are enough cross sections in the data, sample information from large time series is not sufficient to dissipate the relative importance of the prior information.

Second, for fixed values of  $T$ , increasing the number of cross sections  $N$  results in a significant increase in the value of  $\hat{\eta}_{pm}^{EB}$  for small and moderate  $N$  (from 34.6% to 96.8% when  $N$  increases from 12 to 43) and remains relatively stable afterwards. The relative importance of the data in identifying the “true” factor structure does not necessarily increase with the amount of data available but depends on the usefulness of the sample information.

These findings are consistent with the convergence result for  $N \rightarrow \infty$  in Chamberlain and Rothschild [1983]. Identification of the factors is sensitive to  $N$  and only when all available assets are included in the sample is it possible to identify the *correct* factor structure behind the APT. More evidence is extracted from the data with a large number of assets and not necessarily from long time series of a small cross-section of assets.

Table 3: The posterior mean estimate for the dimension of the model  $\hat{q}_{pm}$ , the empirical Bayes parameter  $\hat{c}_{pm}^{EB}$  and the shrinkage parameter  $\hat{\eta}_{pm}^{EB}$ . The RMSE is the square-root of the overall measure of the out-of-sample forecast performance MSFE of equation (44).

<b>Panel A:</b> Empirical Bayes prior $c_\gamma = \hat{c}_\gamma^{EB}$ for each model $\gamma$								
N	T=168				T=528			
	13	43	93	136	12	43	93	136
$\hat{c}_{pm}^{EB}$	0.53	30.05	136	68	0.57	180.3	468	323.9
$100\hat{\eta}_{pm}^{EB}$	34.6%	96.8%	99.3%	98.5%	36.3%	99.4%	99.8%	99.69%
$\hat{q}_{pm}$	10.83	5	4.60	6	11	8.16	5	5
RMSE	0.0663	0.0621	0.0538	0.1111	0.0646	0.0626	0.0486	0.0890

<b>Panel B:</b> Fixed prior $c_\gamma = 4$ , $100\eta_\gamma = 80\%$ for all $\gamma$								
N	T=168				T=528			
	12	43	93	136	12	43	93	136
$\hat{q}_{pm}$	4.70	8.41	9.33	20.03	4	8.16	9	10
RMSE	0.0663	0.0659	0.0541	0.1382	0.0885	0.0640	0.0372	0.0720

<b>Panel C:</b> Risk/unit information prior $c_\gamma = \max\{K^2, T\}$ for all $\gamma$								
N	$T = 168$ $c_\gamma = 441$ $100\eta_\gamma = 99.77\%$ for all $\gamma$				$T = 528$ $c_\gamma = 528$ $100\eta_\gamma = 99.81\%$ for all $\gamma$			
	12	43	93	136	12	43	93	136
$\hat{q}_{pm}$	1	4.58	4.76	5	1	5	5	5
RMSE	0.0997	0.0800	0.1102	0.1280	0.0879	0.0704	0.0964	0.1545

## 5.2 Sensitivity of $\hat{q}_{pm}$ to $c_\gamma$ and the cross-section dimension

The discussion about the effect of the data and the prior hyperparameter on the average size of the model follows directly from the discussion above. The average model size  $\hat{q}_{pm}$  reflects the size of the penalty. When the data does not have enough evidence to identify the factor structure, higher importance is given to the prior information. The prior expected number of factors is equal to  $K/2 = 10.5$ . From the results above, one expects that for low values of  $\hat{\eta}_{pm}^{EB}$ , the estimated value of  $\hat{q}_{pm}$  is closer to its prior value. Indeed for  $N = 12$ , the posterior model size is  $\hat{q}_{pm} = 10.83$  which reflects the small information in the sample and the high importance given to the prior.

The values of  $\hat{q}_{pm}$  are decreasing with  $N$  and are equal for panels with equal values of  $\hat{\eta}$  for the empirical Bayes prior (Panel A). As the number of assets increases, more evidence of common dynamics is found in the data and less importance is given to the prior information. However, this is reversed for the fixed prior  $c_\gamma = 4$  and unit/risk information prior  $c_\gamma = \max\{K^2, T\}$ . In Panels B and C, the parameter  $\eta_\gamma$  is *cross-section invariant* and  $\hat{q}_{pm}$  increases with  $N$ .

Ouysse [2006] finds that the penalty in the multivariate case needs to be a function of both  $N$  and  $T$  to achieve consistency. In small samples, the effect of increasing  $N$  depends on the correlation structure of the error term. For each model  $\gamma$  the empirical Bayes prior  $\hat{c}_\gamma^{EB}$  is equivalent to a data dependent penalty for model dimension in a penalized sum-of-squared information criterion. This penalty incorporates the evidence the data conveys about the factor structure in terms of the amount of information (the time and cross-sectional dimensions) and the cross-sectional correlation in the excess returns through the term  $\hat{F}_\gamma$  in (21).



Table 4: Posterior mean estimates and posterior probability interval for the pricing errors.  $Q_{pm}$  and  $\tilde{Q}_{pm}$  are the BMA estimates of  $Q_N$  and  $\tilde{Q}_N$ , respectively. The mean, standard deviation (std) and the 95% probability interval (95BI) are in percent.

<b>Panel A:</b> January 1990- December 2003, $T = 168$ .						
	$c_\gamma = \hat{c}_{pm}^{EB} = 68$		$c_\gamma = 4$		$c_\gamma = \max\{T, K^2\} = 441$	
	$Q_{pm}$	$\tilde{Q}_{pm}$	$Q_{pm}$	$\tilde{Q}_{pm}$	$Q_{pm}$	$\tilde{Q}_{pm}$
Mean(%)	0.3721	0.0257	1.34	0.2319	0.7202	0.1063
std(%)	0.03	0.003	0.19	0.03	0.1048	0.0174
95BI(%)	[0.2231, 0.2757][0.0049, 0.0127]		[1.0311, 1.7105][0.1721, 0.2935]		[0.5789, 0.8807][0.0839, 0.1396]	

<b>Panel B:</b> January 1960- December 2003, $T = 528$ .						
	$c_\gamma = \hat{c}_{pm}^{EB} = 323$		$c_\gamma = 4$		$c_\gamma = \max\{T, K^2\} = 528$	
	$Q_{pm}$	$\tilde{Q}_{pm}$	$Q_{pm}$	$\tilde{Q}_{pm}$	$Q_{pm}$	$\tilde{Q}_{pm}$
Mean(%)	0.2415	0.0179	10.92	2.79	0.4467	0.0685
std(%)	0.016	0.001	2.67	0.74	0.0424	0.0071
95BI(%)	[0.3392, 0.4311][0.0128, 0.0231]		[6.52, 15.05]		[0.3950, 0.5165][0.0589, 0.0815]	

### 5.3 Pricing errors and predictive performance

Table 4 reports the posterior mean and standard deviation for the square-root of the pricing errors,  $Q_N^2$  and  $\tilde{Q}_N^2$ , under the three choices of the hyperparameter  $c_\gamma$ . To further assess the pricing error, we provide the 95% posterior probability interval (or high density region) which states that there is a 95% probability that the pricing error is in this interval. The results are for the cross-section of combined industry and size-BE/ME excess returns with  $N = 136$ .

The results reported in Panel A and Panel B of Table 4 show that for the priors  $c_\gamma = \hat{c}_\gamma^{EB}$  and  $c_\gamma = \max\{T, K^2\}$ , the posterior mean and standard deviation of the pricing errors decrease when the sample size increases. The opposite outcome is observed for the data independent prior  $c_\gamma = 4$ . The posterior mean and standard deviation of  $Q_N$  are larger than those of  $\tilde{Q}_N$ . For example, for the case of  $c_\gamma = \hat{c}_\gamma^{EB}$  and  $T = 168$ , accounting for the cross-sectional correlation shrinks the posterior mean of the pricing error by 94% (from 0.3721% to 0.0255%). The pricing error  $Q_N^2$  measures the deviation from the APT restrictions under the assumption that the excess returns are homoscedastic and uncorrelated. If there is heteroscedasticity and cross-sectional correlation,  $Q_N^2$  underestimates the fit of the pricing model. Our results are consistent with recent empirical studies on the issue of information gain from increasing the number of cross sections. Boivin and Ng [2005] conduct a simulation study of the forecasting power of factors extracted from large macroeconomic panels. They find that the forecasts and the factor estimates are adversely affected by the cross-correlation in the residual components. Their study suggests that the forecasts are less efficient when the errors are cross correlated and/or have vast heterogeneity in the variances. Abundance of information can cause the correlation to grow larger than what is warranted by the approximate factor theory and therefore create a situation where more data might not be desirable. Boivin and Ng [2005] conclude that it is not simply  $N$  that determines estimation and forecast efficiency. The information that the data can convey about the factor structure is also important.

The results in Table 3 show that all models have relatively similar out-sample performance with values of RMSE that range from a minimum of 3.72% (for  $\hat{q}_{pm} = 9, T = 528, N = 93$  in Panel B) to 15.45% (for  $\hat{q}_{pm} = 20.03, T = 528, N = 136$  in Panel C).

The results in Table 3 and Table 4 suggest that what improves the pricing error does not necessarily improve the predictive performance of the model. Forecast performance is only concerned with the times series dynamics of the excess returns while the pricing error measures how well the factor structure explains the cross-section of returns. The smallest RMSE is achieved for  $c_\gamma = 4$  with large  $\hat{q}_{pm}$  while the smallest pricing errors are achieved under the empirical Bayes prior with moderate posterior mean of the model dimension.

Table 5: Comparison of the maximum likelihood estimate  $\hat{\alpha}_{MLE}^a$  and the Bayesian estimate  $\hat{\alpha}_{pm}^a$  of the annualized historical risk premium. The Bayesian estimates are calculated for the priors  $c_\gamma = 4$  and  $c_\gamma = \hat{c}_\gamma^{EB}$ .

	A: Industry Portfolios $N = 43$			B: Size-BE/ME Portfolios $N = 93$			C: Combined Returns $N = 136$		
	$\hat{\alpha}^a$	$\hat{\alpha}_{pm}^a$		$\hat{\alpha}^a$	$\hat{\alpha}_{pm}^a$		$\hat{\alpha}^a$	$\hat{\alpha}_{pm}^a$	
		$c_\gamma = 4$	$\hat{c}_\gamma^{EB}$		$c_\gamma = 4$	$\hat{c}_\gamma^{EB}$		$c = 4$	$\hat{c}_\gamma^{EB}$
$T = 168$	8.34%	8.60%	8.60%	8.99%	10.30%	8.99%	8.73%	5.54%	8.47%
$T = 528$	7.06%	6.80%	7.06%	5.91%	5.41%	5.91%	6.19%	6.15%	6.17%

An increase in the sample size affects the results in two ways. First, it changes the relative importance of the prior via  $\hat{\eta}_\gamma$ , which depends on the choice of the tuning parameter  $c_\gamma$ , therefore affecting the penalty for model selection. Secondly, it affects the fit of the regression and the predictive power of the model. The net effect depends on whether the structure of the pricing is time invariant or not. If the risk factors are constant through time as well as their premiums, then more observations will improve identification.

## 6 Empirical Results

### 6.1 Expected return and portfolio risk

If the investor's risk preference can be described by a quadratic utility function then only the expected return and its standard deviation matter to the investor's portfolio allocation. Consistent with the efficient market hypothesis, the standard deviation is a proxy for portfolio risk and expected return is an expectation on the future return. This section compares BMA estimates of risk and return to those calculated using maximum likelihood estimation (MLE).

Table 5 reports the point estimates for the *annualized* historical risk premium that investors require on average over the risk free rate for an investment with average risk for each factor. The cross-section average of the annualized risk premiums

$$\hat{\alpha}^a = \frac{1}{N} \sum_{i=1}^N \hat{\alpha}_i^a, \text{ where } \hat{\alpha}_i^a = (1 + \hat{\alpha}_i)^{12} - 1,$$

and  $\hat{\alpha}_i$  is the sample average of the monthly excess returns for asset  $i$ . The Bayesian estimate  $\hat{\alpha}_{pm}^a$  is computed similarly using

$$\hat{\alpha}_{pm}^a = \frac{1}{N} \sum_{i=1}^N \hat{\alpha}_{i,pm}^a, \text{ where } \hat{\alpha}_{i,pm}^a = (1 + \hat{\alpha}_{i,pm})^{12} - 1,$$

and  $\hat{\alpha}_{i,pm}$  is the BMA estimate  $\alpha_i$ .

The estimates of the risk premiums using cross-sectional sample averages range from an annualized rate of 5.91% to 8.99%, not far from values reported in previous studies. Dimson et al. [2003] show evidence of high risk premiums relative to both bonds and bills with an average of 7.5% per annum for the US equity premium relative to bills. Their results also support our finding of high risk premiums in the 90's which is probably driven by the high returns investors enjoyed prior to the "technology bust" in the year 2000. The estimates of the annualized risk premium are comparable using both Bayesian and MLE.

Panel A of Table 6 reports summary statistics for the point estimates of the monthly (realized) risk premiums for the 12 industry returns over the period of January 1960 to December 2003. The table reports the sample means of monthly excess returns  $\hat{\alpha}_i$  for  $i = 1, \dots, 12$ , their estimated standard errors  $s_{\hat{\alpha}_i}$

$$s_{\hat{\alpha}_i} = \frac{s_i}{\sqrt{T}}, \text{ where } s_i^2 = \frac{\sum_{t=1}^T (y_{it} - \hat{\alpha}_i)^2}{T - 1},$$

Table 6: Expected returns, BMA estimates of risk premium and posterior probability interval for the excess returns on the 12 industry excess returns over the period of January 1960 to December 2003. Panel A reports the monthly average excess return  $\hat{\alpha}_i$  for security  $i$ , its standard error  $s_{\hat{\alpha}_i}$  and its 90% confidence interval  $90CI(\hat{\alpha}_i)$ , the BMA estimate of the monthly excess returns  $\hat{\alpha}_{i,pm}$ , its posterior standard deviation  $s_{\hat{\alpha}_{i,pm}}$  and its 90% posterior probability interval  $90BI(\hat{\alpha}_{i,pm})$ . Panel B reports the risk factors with posterior probability  $\hat{\pi}_k > 0$ ,  $\hat{\delta}_{pm}$  are the BMA estimates of the average risk premiums,  $s_{\hat{\delta}_{pm}}$  are their posterior standard deviations,  $\hat{\delta}_{pm}^{CON}$ ,  $\hat{\delta}_{pm}^{EXP}$  are the BMA estimates of the average risk premiums over the months of economic contraction and expansion respectively,  $s_{\hat{\delta}_{pm}^{CON}}$  and  $s_{\hat{\delta}_{pm}^{EXP}}$  are their posterior standard deviations. The results are for the prior  $c_\gamma = 4$  and all the reported values are in percent except for the probability  $\hat{\pi}_k$ .

**Panel A:** Estimates of the realized risk premiums  $\alpha$ .

	$\hat{\alpha}_i$	$s_{\hat{\alpha}_i}$	$\hat{\alpha}_{i,pm}$	$s_{\hat{\alpha}_{i,pm}}$	$\hat{Var}1_i$	$90BI(\hat{\alpha}_{i,pm})$	$90CI(\hat{\alpha}_i)$
NoDur	0.63	0.20	0.59	0.58	51.99	[ -0.17 , 1.55 ]	[ 0.31 , 0.95]
Durbl	0.42	0.24	0.40	0.12	74.86	[ 0.25 , 0.60 ]	[ 0.02 , 0.82]
Manuf	0.43	0.22	0.46	0.07	83.82	[ 0.36 , 0.60 ]	[ 0.07 , 0.78]
Enrgy	0.55	0.22	0.66	0.21	92.79	[ 0.38 , 1.10 ]	[ 0.18 , 0.91]
Chems	0.42	0.21	0.32	0.24	92.29	[ -0.24 , 0.56 ]	[ 0.08 , 0.76]
BusEq	0.48	0.29	0.44	0.11	85.31	[ 0.27 , 0.61 ]	[ -0.00 , 0.96]
Telcm	0.41	0.20	0.54	0.21	92.84	[ 0.31 , 1.02 ]	[ 0.08 , 0.75]
Utils	0.35	0.18	0.36	0.10	89.88	[ 0.21 , 0.52 ]	[ 0.05 , 0.63]
Shops	0.57	0.23	0.54	0.31	96.72	[ -0.23 , 1.00 ]	[ 0.19 , 0.95]
Hlth	0.64	0.23	0.78	0.25	97.37	[ 0.43 , 1.35 ]	[ 0.26 , 1.00]
Money	0.60	0.22	0.57	0.07	96.97	[ 0.41 , 0.69 ]	[ 0.23 , 0.96]
Other	0.35	0.24	0.36	0.10	95.05	[ 0.18 , 0.50 ]	[ -0.04 , 0.73]

**Panel B:** Estimates of the factor specific risk premiums  $\delta$ .

Factors	$\hat{\pi}_k$	$\hat{\delta}_{pm}$	$s_{\hat{\delta}_{pm}}$	$\hat{\delta}_{pm}^{CON}$	$s_{\hat{\delta}_{pm}^{CON}}$	$\hat{\delta}_{pm}^{EXP}$	$s_{\hat{\delta}_{pm}^{EXP}}$	$90BI(\hat{\delta}_{pm})$
$\delta_0$	1	0.0044	0.0423	-0.0107	0.0431	0.0080	0.0424	[ 0.0032 , 0.0056]
JAN	0.5	0.0116	1.7453	-0.2317	1.8032	0.0884	1.7042	[ -0.0680 , 0.0962]
SMB	0.1	-0.0052	0.3121	-0.0247	0.3131	-0.0004	0.2867	[ -0.0522 , 0.0000]
CG	0.7	-0.1368	3.7798	-0.2584	4.5129	0.0429	3.1259	[ -0.4709 , 0.0000]
URP	0.1	0.0140	0.7791	-0.0654	0.7706	0.0101	0.7588	[ 0.0000 , 0.1398]
UNEM	0.9	-0.2500	4.5690	-1.3143	5.7435	0.1970	4.3154	[ -0.5286 , 0.0234]
EWRET	0.1	0.0055	0.6662	0.1449	0.7437	-0.0749	0.5889	[ 0.0000 , 0.0551]
MOMT	0.6	0.0113	3.2388	0.1485	3.4179	0.0513	3.1581	[ -0.1554 , 0.1992]

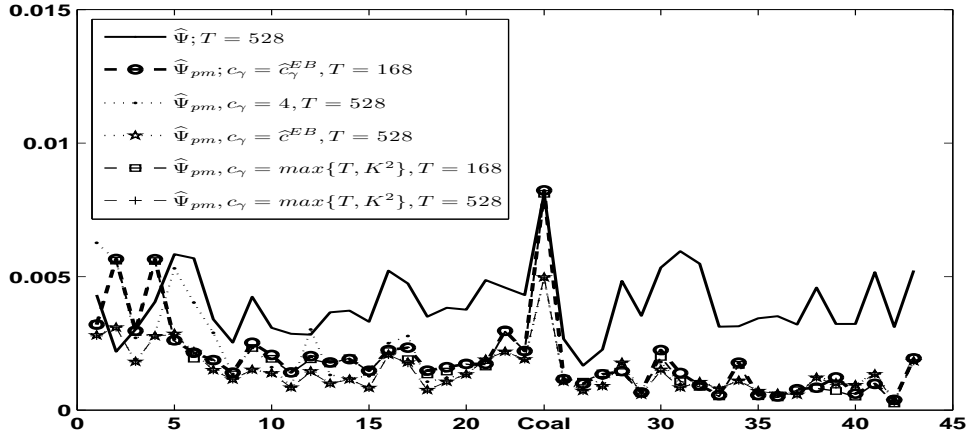
and the BMA estimates  $\hat{\alpha}_{i,pm}$  and the corresponding posterior standard deviations

$$\hat{s}_{\hat{\alpha}_{i,pm}} = \frac{\sum_{j=1}^M (\alpha_i^{(j)} - \hat{\alpha}_{i,pm})}{M - 1}.$$

The estimates of the historical risk premiums based on the sample means of the monthly excess returns  $\hat{\alpha}$  are quantitatively comparable to the Bayesian estimates  $\hat{\alpha}_{pm}$ . Although the standard errors  $s_{\hat{\alpha}}$  and the posterior standard deviations  $s_{\hat{\alpha}_{pm}}$  are relatively comparable for some securities, the posterior standard deviation suggests that risk is heterogenous across industries. These point estimates are not very useful unless accompanied by some probabilistic interval. Inference based on the posterior probability intervals may lead to different conclusions compared to those based on the asymptotic confidence intervals. For example, the posterior probability interval corresponding to the highest density region for  $\hat{\alpha}_{pm}$  for the chemicals industry (Chems) is  $[-0.24\% , 0.56\%]$  suggesting that the realized premium is very likely to be negative. However, the asymptotic confidence interval  $[ 0.08\% , 0.76\%]$  strongly suggests a positive monthly realized premium.

Using decomposition (3) of the covariance matrix of excess returns, we compute the BMA estimates

Figure 1: Comparison of the maximum likelihood and Bayesian estimates of risk for the cross-section of the industry sorted portfolios,  $N = 43$ .



of  $\Psi_N$ ,

$$\hat{\Psi}_{pm} = \frac{1}{M} \sum_{j=1}^M \Psi^{(j)}, \text{ where } \Psi^{(j)} = \hat{\Lambda}^{(j)} \Sigma_F^{(j)} \hat{\Lambda}^{(j)'} + \Sigma^{(j)},$$

and  $\Sigma_F^{(j)}$  is the covariance matrix of the design matrix  $F_{\gamma^{(j)}}$  and  $\gamma^{(j)}$ ,  $\Lambda^{(j)}$  and  $\Sigma^{(j)}$  are the iterates from the MCMC sampling scheme. Figure 1 plots the diagonal elements of  $\hat{\Psi}_{pm}$  and those of the sample covariance matrix of  $\mathbf{Y}$ ,  $\hat{\Psi} = \text{cov}(\mathbf{Y})$ . The results indicate that the estimates of risk for the industry portfolios are not significantly affected by the choice of  $c_\gamma$ . All the priors for  $c_\gamma$  lead to quantitatively similar measures of risk. The results also suggest that maximum likelihood tends to give higher estimates of portfolio risk. However both maximum likelihood and Bayesian estimates give qualitatively similar conclusions. For example, all the estimates in Figure 1 suggest that the Coal industry bears the highest level of risk with a standard deviation of 1.252% per month.

## 6.2 The January effect

Rozeff and Kinney [1976] report evidence of seasonal patterns in stock market returns. January returns appear to be more than eight times higher than a typical month. From 1904 through to 1974, the average stock market return during the month of January was 3.48 percent whereas the monthly return for the remaining months of the year was 0.42 percent.

Our analysis of the APT factor structure includes a January dummy (which takes one if the month is January and zero otherwise) as a likely risk factor. The market reward for the January effect is equal to the product of the premium per unit exposure  $\delta_{JAN}$  and the portfolio beta  $\beta_{i,JAN}$ . The January betas reflect the seasonal patterns in the time series dynamics of the excess returns while the per unit premium reflects the market reward for exposure to the seasonal risk.

Panel B of Table 6 reports the posterior mean estimates  $\hat{\pi}_k$  of the probability of a risk factor  $k$  to be in “true” factor structure. For the January Dummy, this probability is equal to its prior value of 1/2. The corresponding premium per unit of exposure to the seasonal risk is 0.0116%. The evidence from the data does not support the existence of a January effect except in the cross-section of size-BE/ME portfolio returns and their combination with the 43 industry portfolio returns. In Panel I of Table 7,  $\hat{\pi}_{JAN} = 0.03$  for the cross-section of industry returns over the two sample periods. The posterior probability intervals further supports the observation that the January seasonal risk is not priced in the market. In Table 8, the posterior mean of  $\hat{\pi}_{JAN} = 1$ . The January has a positive unit premium  $\hat{\delta}_{JAN,pm}$  of order 0.0181% with a

90% posterior probability interval of  $[-0.0320\%, 0.0282\%]$ . The estimated January pricing component for the combined returns over the period of 1990 to 2003 is 0.15% (Table 8). The results for the size-BE/ME returns show strong evidence for January effect. In Panel II of Table 7, the BMA estimate for the probability that the January is in the factor structure is 0.93 for the period of 1990-2003 and one for both the sample period 1960-2003. The premium per unit exposure to the January effect is positive and is equal to 0.0709% and 0.1490% for the period the two periods respectively. The posterior probability interval indicates that there is high probability the January premium is negative.

Haug and Hirschey [2006] find a persistent January effect for small-cap stocks in equally weighted returns for the period 1802-2004. They document an “anomalous” pattern of monthly returns for portfolios based on the Fama and French size-BE/ME factors and show that both factors contribute to the continuing January effect.

### 6.3 Time varying price for economic risk

This section investigates the time series dynamics of the estimated risk premiums. We use the NBER dating of the business cycle in Table 2 to define the periods of economic expansion and recession.

The BMA estimate  $\hat{\delta}_{t,pm}$  of the vector of monthly risk premiums associated to the  $K$  risk factors is

$$\hat{\delta}_{t,pm} = \frac{1}{M} \sum_{j=1}^M \hat{\delta}_t^{(j)}, t = 1, \dots, T, \quad (45)$$

where  $\hat{\delta}_t^{(j)}$  are the coefficients of time-varying risk premiums in (13) calculated under model  $\gamma^{(j)}$ . We define the BMA estimates  $\hat{\delta}_{pm}^{EXP}$  of the average risk premiums during the months of economic expansion as

$$\hat{\delta}_{pm}^{EXP} = \frac{1}{M} \sum_{j=1}^M \hat{\delta}_{EXP}^{(j)}, \quad (46)$$

$$\hat{\delta}_{EXP}^{(j)} = \frac{1}{T_{exp}} \sum_{t=1}^T \hat{\delta}_t^{(j)} I(t = exp), \quad (47)$$

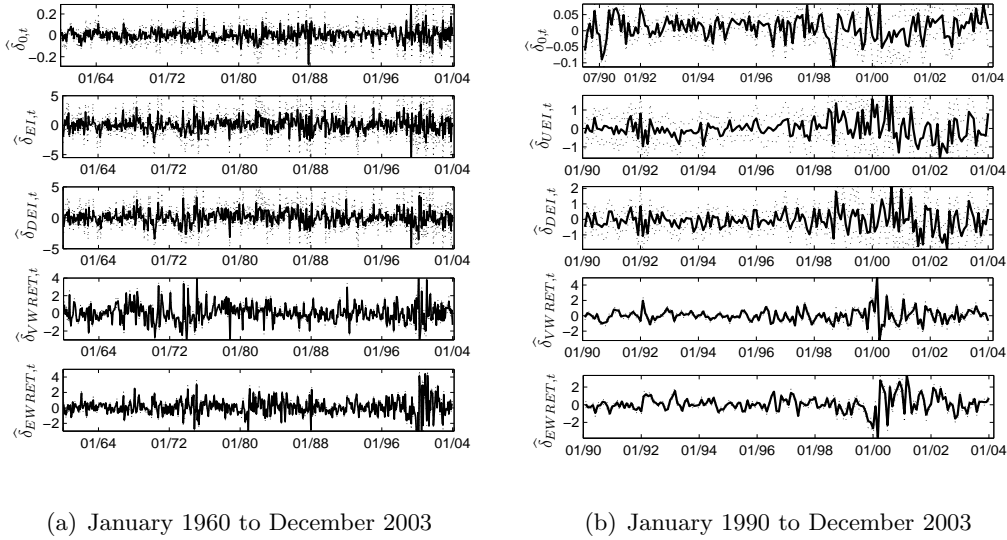
where  $I(t = exp)$  is an indicator function that takes 1 if  $t$  corresponds to a month in the expansion period, and  $T_{exp} = \sum_{t=1}^T I(t = exp)$ . The BMA estimates  $\hat{\delta}_{pm}^{CON}$  for the average risk premiums during the months of economic contraction are computed similarly by averaging over the contraction months. The posterior standard deviations of  $\hat{\delta}_{pm}^{EXP}$ ,  $\hat{\delta}_{pm}^{CON}$  are denoted by  $s_{\hat{\delta}_{pm}^{EXP}}$ ,  $s_{\hat{\delta}_{pm}^{CON}}$  respectively.

Our analysis of the monthly time series of risk premiums indicates that overall the market price for risk strongly fluctuates over the business cycle. Figure 2 plots the iterates of the time-varying risk premiums for risk factors with  $\hat{\pi}_k = 1$ . The market premium for exposure to risk is fluctuating around a long run average with episodes of high volatility. A noteworthy observation is the change in the dynamics of the series of risk premiums around the year 2000 with an apparent increase in its volatility.

The figure also shows the time series dynamics of  $\hat{\delta}_{0,t,pm}$ , the BMA estimate of the monthly mispricing. For the period 1960–2003, the ability of the APT model to explain the realized risk premium starts declining in the mid-1998 with a sharp incidence of mispricing in January 1988. The estimated monthly mispricing for the period 1990-2003 are significantly lower compared to those of the 1960-2003 sample period, with values within  $[-0.05\%, 0.05\%]$ . This suggests the factor structure itself could be time-varying or at least that the market’s valuation of risk has changed over time. This is an important finding which separates the time series dynamics of the returns from the pricing dynamics of the cross-section of returns. A factor might be “significant” in explaining the time series variation and its beta might be constant over time, however the market reward for the exposure to the underlying risk may be time-varying. A possible explanation is that the opportunities for the investors to hedge against the risk are changing overtime.

We also compute the sample average of the risk premiums estimates over phases of economic contraction and expansion. Three patterns emerges from the analysis of the results in Tables 6-8. First, the average premium over the months of economic recession is higher than the pooled sample average.

Figure 2: Time series of the posterior mean estimates of the monthly risk premium  $\hat{\delta}_{k,t,pm}$  for factor  $k$  and month  $t$  and the corresponding 95% probability interval (dotted line). In the plots we drop the subscript “pm” for ease of presentation. The plots are for the cross-section of combined industry and size-BE/ME portfolio returns. The results are for the *unit information* or *risk information* prior  $c_\gamma = \max\{T, K^2\}$ ,  $N = 136$ .



Second, the volatility of the posterior mean estimates of the monthly prices of risk is higher in periods of economic contraction compared to the pooled sample. Third, phases of expansion appear to have less volatility than in the full sample.

#### 6.4 Excess returns predictability and economic fundamentals

This section analyzes the “economic” composition of the posterior mean of the distribution of the Bayesian estimates of the risk premium  $\hat{\alpha}_{pm}$ . We consider  $\hat{\pi}_k$  the posterior mean estimate of the probability that a risk factor  $k$  is priced in the stock market,  $\hat{\delta}_{k,pm}$  the posterior mean estimate of the average risk premium (per unit exposure to the risk factor  $k$ ) and  $\widehat{\text{Var}}1_i \times 100$  the percentage of the systematic risk relative to the total risk defined in equation (10).

*Industry Portfolios.* The estimation results for the 43 industry returns for  $c_\gamma = 4$  in Panel I of Table 7 suggest that the economic risk is more likely to be driven by expected inflation EI, changes in expected inflation DEI, changes in the monthly industrial production, the value/equally weighted portfolio return and the momentum factor. For the period of January 1990 to December 2003, the premium on DEI is negative and equal to 0.0233 percent per month (equivalent to 0.28 percent annually). The posterior probability interval further suggests that given the observed excess returns, there is 0.95 probability that the per annum risk premium associated to exposure to changes in expected inflation lies between minus 0.84 percent and 0.54 percent. Chen et al. [1986] argue that the negative sign could be explained by investors hedging against the adverse influence on assets that are fixed in nominal terms. However, DEI does not appear as a likely risk factor for the period of January 1960 to December 2003. Expected inflation EI earns a high positive premium for both subperiods and ranges from a per annum average premium of 2.43 percent (monthly 0.2010%) for the period of 1960 – 2003 to 3.10 percent (0.2551% per month) for the period of 1990 – 2003. Risa [2001] finds that the 10 year inflation risk premium for the UK between 1983 and 1999 is mostly around 2% with an initial level of 4% and a sharp drop to a slightly

negative value in 1999. The posterior probability interval show high probability of both negative and positive EI risk premium.

We find strong evidence for high positive momentum premium for both subperiods. The per unit reward for exposure to the momentum factor ranges from 0.0512% per month for the period of 1990 – 2003 to 0.1377% per month for the period of 1960 – 2003. These significant posterior BMA estimates of the average momentum premium is This means that investors are rewarded for holding stocks that are “winners”, i.e., those which performed well in the past 12 months. L’Her et al. [2004] measure the momentum factor premium at 16.07% and find that the premium is larger in up markets and remains positive in down markets. Monthly changes in production growth MP earns positive premium for the 1960-2003 sample period with a per annum premium of 0.75%.

The predictable variation in the Bayesian historical premium is measured by the percentage of variation attributed to the estimated systematic risk relative to the Bayesian estimate of total risk. The values reported in the last column of Panel A of Table 6 are all above 74% with the exception of the nondurable goods industry (NoDur) which has an estimated 49% of the risk due to the idiosyncratic component.

Table 6 suggests that the growth rate of consumption of nondurables requires a risk premium of minus 1.65 percent per annum (0.1368% per month). This is in contrast to the higher and positive premium reported by Ferson and Harvey [1991] (0.32%) and to the non significant result found by Chen et al. [1986]. This result is however consistent with Bansal and Yaron [2004]. They find negative premium (about  $-0.4\%$  and  $-1.2\%$  for the returns on the 12 and the 36 months real bonds, respectively) suggesting that real bonds offer a consumption insurance to investors.

*Size-BE/ME returns.* The results for the excess returns on the size-BE/ME portfolios (Panel II of Table 7) further reinforce the role of unanticipated changes in inflation. The posterior average premiums for unexpected inflation and changes in expected inflation are negative and similar to those reported above for the industry returns. The magnitudes of the risk premiums have a wide range; from minus 0.27 percent per year for DEI in the 1990-2003 period to minus 0.19 percent per annum for UEI. The posterior mean  $\hat{\pi}$  for the unemployment rate indicator is equal to 0.48 for  $c_\gamma = 4$  and for the sample period 1990-2003 with an average premium of 0.0147 percent indicating that investors dislike (surprise) increases in the unemployment rate. The momentum factor again requires a positive and high per annum premium of 1.93 percent and 3.39 percent for 1990-2003 and 1960-2003 respectively, reflecting the strong momentum effect in the size returns.

*Combined returns.* For the combined cross sections of industry and size-BE/ME returns, the results in Table 8 further reinforce the existence of market price for the inflation variables and the monthly changes in industrial production. There is also evidence of positive premium for the January and the momentum effects.

## 7 Conclusion

Our article adapts recent advances in Bayesian multivariate analysis and variable selection to investigate the role of measured economic variables in the pricing of securities. The methodology provides a flexible framework for simultaneously dealing with model and parameter uncertainty. We use variable selection and Bayesian model averaging to estimate an overall model because model selection using Bayes factors is computationally intractable as there are a large number of models under consideration.

Our article contributes to understanding the role of the cross-sectional dimension in a Bayesian multivariate regression. For the empirical Bayes specification of  $c_\gamma$  in the g-prior, we find that the posterior mean of  $c_\gamma$  is an increasing function of the number of cross-sections which means an increased importance given to the cross-section information over the prior information. We also find that the posterior mean of the model dimension decreases with the number of cross-sections, which is consistent with the asymptotic argument about the existence of a factor structure in Chamberlain and Rothschild [1983]. If an approximate  $K$ -factor structure does exist, then as the number of cross-sections  $N$  grows large only  $K$  of the eigenvalues of the covariance matrix of the excess returns can be unbounded while the remaining  $N - K$  converge to zero.

In the case of a fixed, a *unit information* and a *risk information* prior for  $c_\gamma$ , the importance given to the cross-section information over the prior information is constant with respect to  $N$  and the posterior mean of the model dimension increases with the number of cross-sections. Our results also show that the posterior mean of the pricing error under the empirical Bayes prior is smaller indicating a higher ability to explain the cross-section of excess returns.

Our results provide strong evidence for market reward for “Economic” risk measured by unanticipated inflation, changes in expected inflation, unemployment rate and changes in industrial production. We demonstrate the time-varying nature of the factor risk premiums and show that their time series dynamics are tightly linked to the Business cycle. We find evidence of higher average risk premium with high volatility during times of economic contraction. These changes in the direction and magnitude of the economic risk premium can be attributed to changes in the hedging capability of alternative securities with no significant risk premium.

However, we find no robust evidence for the relevance of the Fama and French three factors in the pricing of the returns on industry and size-BE/ME sorted portfolios. This suggests that multiple betas from a multifactor model absorb the role of size and BE/ME equity and that the pricing of securities can ultimately be determined by systematic economic risks instead of firm-specific variables.

This finding is consistent with the recent empirical evidence in the finance literature. Cremers [2006] uses a Bayesian framework and an inefficiency metric based on the maximum correlation between the market portfolio and any multifactor-efficient portfolio to show that neither Fama and French factors nor the momentum factor improve pricing performance relative to the CAPM.

## References

- Armanov, D., 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* 64 (3), 423–458.
- Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–172.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59, 1481–1509.
- Boivin, J., Ng, S., 2005. Are more data always better for factor analysis? *Journal of Econometrics* 132, 169–194.
- Brown, P. J., Vannucci, M., Fearn, T., 1998. Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society, series B* 60 (3), 627–641.
- Brown, P. J., Vannucci, M., Fearn, T., 2002. Bayes model averaging with selection of regressors. *Journal of the Royal Statistical Society, series B* 64 (3), 519–536.
- Brown, S. J., 1989. The number of factors in security returns. *Journal of Finance* 44, 1247–1262.
- Campbell, J. Y., Lo, A. W., MacKinlay, C. A., 1997. *The econometrics of financial markets*. Princeton University Press, Princeton, New Jersey.
- Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* 51, 1305–1324.
- Chen, N. F., Roll, R., Ross, S. A., 1986. Economic forces and the stock market. *Journal of Business* 59, 383–403.
- Chib, S., 2001. Markov chain Monte Carlo methods: Computation and inference. In: J.J. Heckman and E. Leamer, Editors, *Handbook of Econometrics* 5, 3569–3649.



- Chipman, H., George, E. I., McCulloch, R. E., 2001. The practical implementation of Bayesian model selection. *IMS Lecture Notes- Monograph Series* 38, 65–134.
- Clyde, M. A., 1999. Bayesian model averaging and model search strategies. *Bayesian Statistics* 6, 157–185.
- Connor, G., 1984. A unified beta pricing theory. *Journal of Economic Theory* 34, 13–31.
- Connor, G., Korajczyk, R. A., 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics* 21, 255–289.
- Connor, G., Korajczyk, R. A., 1989. An intertemporal equilibrium beta pricing model. *Review of Financial Studies* 2 (3), 255–289.
- Connor, G., Korajczyk, R. A., 1993. A test for the number of factors in an approximate factor model. *Journal of Finance* XLVIII (4), 1263–1291.
- Cremers, K. J. M., 2006. Multifactor efficiency and Bayesian inference. *Journal of Business* 79, 2951–2998.
- Cripps, E., Carter, C., Kohn, R., 2005. Variable selection and covariance selection in multivariate regression models. *Handbook of Statistics: Bayesian Thinking: Modeling and Computation* Editors: Dipak K. Dey, University of Connecticut and C.R. Rao, Pennsylvania State University Publisher: Elsevier Science, Amsterdam, The Netherlands 25.
- Dimson, E., Marsh, P., Staunton, M., 2003. Global evidence on the equity risk premium. *Journal of Applied Corporate Finance* 15 (4), 08–19.
- Donoho, D. L., Johnstone, I. M., 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–456.
- Evans, I. G., 1965. Bayesian estimation of parameters of a multivariate normal distribution. *Journal of the Royal Statistical Society, Series B (Methodological)* 27 (2), 279–283.
- Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81 (3), 607–636.
- Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. *Journal of Finance* 47 (2), 427–465.
- 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., 1995. Size and book to market factors in earnings and returns. *Journal of Financial Economics* 50 (1), 131–156.
- Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. *Journal of Financial Economics* 51 (1), 55–84.
- Fernandez, C., Ley, E., Steel, M., 2001. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100, 381–427.
- Ferson, E. W., Harvey, R. C., 1991. The variation of economic risk premiums. *Journal of Political Economy* 99 (2), 385–415.
- Ferson, W., Korajczyk, R. A., 1995. Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business* 68 (3), 309–349.
- Galimberti, G., Montanari, A., Viroli, C., 2009. Penalized factor mixture analysis for variable selection in clustered data. *Computational Statistics & Data Analysis* 53 (12), 4301–4310.

- George, E., McCulloch, R. E., 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881–889.
- George, E., McCulloch, R. E., 1997. Approaches for Bayesian variable selection. *Statistica Sinica* 7, 339–373.
- George, E. I., Foster, D. P., 2000. Calibration and empirical Bayes variable selection. *Biometrika* 87 (4), 731–747.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid and A. Smith (eds), *Bayesian Statistics 4*, 169–193.
- Geweke, J., 1996. Variable selection and model comparison in regression. In *Bayesian Statistics 5* (eds J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith), 609–620.
- Geweke, J., Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies* 9 (2), 557–587.
- Harvey, C. R., Zhou, G., 1990. Bayesian inference in asset pricing tests. *Journal of Financial Economics* 26, 221–254.
- Haug, M., Hirschey, M., 2006. The January effect. *Financial Analyst Journal* 62 (5), 78–88.
- Hofmann, M., Gatu, C., Kontoghiorghes, E. J., 2007. Efficient algorithms for computing the best subset regression models for large-scale problems. *Computational Statistics & Data Analysis* 52 (1), 16–29.
- Ingersoll, J. E. J., 1984. Some results in the theory of arbitrage pricing. *Journal of Finance* 39, 1021–1039.
- Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance* 48, 65–91.
- Kapetanios, G., 2007. Variable selection in regression models using nonstandard optimisation of information criteria. *Computational Statistics & Data Analysis* 52 (1), 4–15.
- Kohn, R., Smith, M., Chan, D., 2001. Nonparametric regression using linear combinations of basis functions. *Statistics and Computing* 11, 313–322.
- Leamer, E., 1978. *Specification searches: Ad hoc inference with non experimental data*. John Wiley and Sons, Inc.
- Ley, E., Steel, M. F. J., 2008. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. World Bank, <http://mpra.ub.muenchen.de/3214>.
- L’Her, J. F., Masmoudi, T., Suret, J. M., 2004. Evidence to support the four-factor pricing model from the canadian stock market. *Journal of International Financial Markets, Institutions & Money* 14 (4), 313–328.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., Berger, J., 2008. Mixtures of g-priors for Bayesian variable selection. *Journal of the American Statistical Association* 103 (481), 410–423.
- Madigan, D., Raftery, A., 1994. Model selection and accounting for model uncertainty in graphical models using Occam’s window. *Journal of the American Statistical Association* 89, 1535–1546.
- McCulloch, R., Rossi, P. E., 1990. Posterior, predictive and utility based approaches to testing arbitrage pricing theory. *Journal of Financial Economics* 28, 7–38.
- McCulloch, R., Rossi, P. E., 1991. A Bayesian approach to testing the arbitrage pricing theory. *Journal of Econometrics* 49, 141–168.

- Mitchell, T. J., Beauchamp, J. J., 1988. Bayesian variable selection in linear regression. *Journal of the American Statistical Association* 83 (404), 1023–1032.
- Nardari, F., 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 (4), 857–891.
- Nott, d., Kohn, R., 2005. Adaptive sampling for Bayesian variable selection. *Biometrika* 92, 747–763.
- Ouyse, R., 2006. Consistent variable selection in large panels when factors are observable. *Journal of Multivariate Analysis* 97, 946–984.
- Reinganum, M. R., 1981. A new empirical perspective on the CAPM. *Journal of Financial and Quantitative Analysis* 16 (4), 439–462.
- Risa, S., 2001. Nominal and inflation indexed yields: Separating expected inflation and inflation risk premia. <http://ssrn.com/abstract=265588> or DOI: 10.2139/ssrn.265588.
- Robotti, C., Balduzzi, P., 2008. Mimicking portfolios, economic risk premia, and tests of multi-beta models. *Journal of Business and Economic Statistics* 26 (3), 354–368.
- Ross, S. A., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.
- Rozeff, M. S., Kinney, W. R., 1976. Capital market seasonality: The case of stock returns. *Journal of Financial Economics* 3 (4), 379–402.
- Shanken, J., 1987. A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics* 19, 195–215.
- Shanken, J., 1992. On the estimation of beta-pricing models. *Review of Financial Studies* 5 (1), 1–33.
- Smith, A., Roberts, G., 1993. Bayesian computation via Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 55, 3–24.
- Smith, M., Kohn, R., 1996. Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* 75, 317–343.
- Stambaugh, R. F., 1982. On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics* 10 (3), 237–268.
- Stambaugh, R. F., 1983. Arbitrage pricing with information. *Journal of Financial Economics* 12 (3), 357–369.
- Stricklanda, C. M., Turnera, I. W., Denhamb, R., Mengersena, K. L., 2009. Efficient Bayesian estimation of multivariate state space models. *Computational Statistics & Data Analysis* 53 (12), 4116–4125.
- Zellner, A., 1986. Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients. In Slottje, D., eds., *Advances in Econometrics* 5, 171–182.

Table 7: BMA estimates  $\hat{\pi}_k$  for the probability that risk factor  $k$  to be in the factor structure, BMA estimates  $\hat{\delta}_{k,pm}$  of the risk premium for exposure to risk factor  $k$  and the corresponding posterior probability interval  $95\text{BI}(\hat{\delta}_{k,pm})$ . The returns are the industry sorted portfolios in Panel I and the size-BE/ME sorted portfolios in Panel II. All the reported values are in percent except for the probability  $\hat{\pi}_k$ .

Panel I: Industry returns, N=43						
1960 : 01 – 2003 : 12, $T = 528$				1990 : 01 – 2003 : 12, $T = 168$		
$c_\gamma = 4, \hat{q}_{pm} = 3$				$c_\gamma = 4, \hat{q}_{pm} = 8.41$		
Factors	$\hat{\pi}_k$	$\hat{\delta}_{k,pm}$	$95\text{BI}(\hat{\delta}_{k,pm})$	$\hat{\pi}_k$	$\hat{\delta}_{k,pm}$	$95\text{BI}(\hat{\delta}_{k,pm})$
$\delta_0$	1	0.0055	[ 0.0041 , 0.0067]	1	0.0067	[ 0.0045 , 0.0093]
JAN	0.03	0.0030	[ 0 , 0 ]	0.03	0.0109	[ 0 , 0 ]
Market	0	-	-	0.87	0.0576	[-0.0769 , 0.0759]
CG	0	-	-	0.35	0.0165	[-0.0894 , 0.0019]
URP	0.12	0.0206	[-0.0167 , 0.0163]	0	-	-
UNEM	0.03	0.0053	[ 0 , 0 ]	0	-	-
MP	1	0.0634	[-0.0492 , 0.1064]	0	-	-
DEI	0	-	-	1	-0.0233	[-0.0700 , 0.0450]
EI	1	0.201	[-0.1880 , 0.5850]	1	0.2551	[-0.2513 , 0.7407]
VWRET	1	-0.0096	[-0.0232 , 0.0363]	0	-	-
EWRET	1	-0.0340	[-0.0692 , 0.0015]	1	0.0215	[-0.0719 , 0.0435]
MOMT	1	0.1377	[-0.0414 , 0.0630]	1	0.0512	[-0.0368 , 0.1033]

Panel II: size-BE/ME returns, N=93						
1990 : 01 – 2003 : 12, $T = 168$				1960 : 01 – 2003 : 12, $T = 528$		
$c_\gamma = 4, \hat{q}_{pm} = 9.33$				$c_\gamma = 4, \hat{q}_{pm} = 10$		
Factors	$\hat{\pi}_k$	$\hat{\delta}_{k,pm}$	$95\text{BI}(\hat{\delta}_{k,pm})$	$\hat{\pi}_k$	$\hat{\delta}_{k,pm}$	$95\text{BI}(\hat{\delta}_{k,pm})$
$\delta_0$	1	0.0077	[ 0.0055 , 0.0097]	1	0.0019	[ 0.0033 , 0.0064]
JAN	0.93	0.0709	[-0.0394 , 0.0962]	1	0.1490	[-0.0737 , 0.0672]
SMB	0.04	0.0078	[ 0 , 0 ]	0	-	-
CG	0.18	-0.0219	[-0.0024 , 0.0022]	0	-	-
UNEM	0.48	0.0147	[-0.0322 , 0.0871]	0	-	-
MP	0.09	0.0025	[ 0 , 0 ]	1	0.3264	[-0.0989 , 0.0332]
EI	0.30	-0.0045	[-0.0015 , 0.0013]	1	0.0642	[-0.0644 , 0.0208]
UEI	0.98	-0.0229	[-0.0986 , 0.0049]	0	-	-
DEI	0.80	-0.0164	[-0.0677 , 0.0048]	0	-	-
VWRET	1	0.0134	[-0.0489 , 0.0029]	1	0.0544	[-0.0413 , 0.0206]
EWRET	1	0.0347	[-0.0020 , 0.0576]	1	0.0625	[ 0.0352 , 0.0905]
MOMT	1	0.1688	[-0.0063 , 0.1471]	1	0.2785	[-0.0582 , 0.1107]

Table 8: BMA estimates of the risk premiums  $\widehat{\delta}_{k,pm}$  for the risk factor  $k$ , its 90 percent posterior probability interval  $90\text{BI}(\widehat{\delta}_{k,pm})$ , the BMA estimates of the average risk premiums for the months of economic expansion ( $\widehat{\delta}_{pm}^{EXP}$ ) and contraction ( $\widehat{\delta}_{k,pm}^{CON}$ ) and their posterior standard deviations  $s_{\widehat{\delta}_{k,pm}^{EXP}}$  and  $s_{\widehat{\delta}_{k,pm}^{CON}}$ . The returns are combined cross sections of size-BE/ME and industry portfolios,  $N = 136$ . All the risk factors reported in this table have  $\widehat{\pi}_k = 1$  and all the values are in percent.

Panel A: period of 1960-2003 with $c_\gamma = 4$						
<i>Factors</i>	$\widehat{\delta}_{k,pm}$	$90\text{BI}(\widehat{\delta}_{k,pm})$	$\widehat{\delta}_{k,pm}^{EXP}$	$s_{\widehat{\delta}_{k,pm}^{EXP}}$	$\widehat{\delta}_{k,pm}^{CON}$	$s_{\widehat{\delta}_{k,pm}^{CON}}$
mispricing	0.0027	[ 0.0029 , 0.0027]	0.0039	0.0464	-0.0031	0.0631
JAN	0.0181	[-0.0320 , 0.0282]	-0.0014	0.1132	-0.0172	0.1274
MP	0.0139	[-0.0319 , 0.0226]	-0.0047	0.0451	-0.0027	0.0538
EI	0.0332	[-0.0045 , 0.0242]	0.0080	0.1213	0.0043	0.1547
UEI	0.0302	[-0.0059 , 0.0241]	0.0078	0.1381	0.0086	0.1695
VWRET	0.0354	[ 0.0003 , 0.0355]	0.0161	0.2309	0.0262	0.2046
EWRET	0.0458	[ 0.0222 , 0.0538]	0.0407	0.6622	-0.0309	0.7696
MOMT	0.0262	[-0.0149 , 0.0443]	0.0122	0.1413	0.0222	0.1479

Panel B: period of 1990-2003 with $\widehat{c}_\gamma^{EB} = 68.52$						
<i>Factors</i>	$\widehat{\delta}_{k,pm}$	90BI	$\widehat{\delta}_{k,pm}^{EXP}$	$s_{\widehat{\delta}_{k,pm}^{EXP}}$	$\widehat{\delta}_{k,pm}^{CON}$	$s_{\widehat{\delta}_{k,pm}^{CON}}$
mispricing	0.0081	[ 0.0038 , 0.0107]	0.0071	0.0440	0.0134	0.0540
EI	-0.0614	[-0.1464 , 0.0774]	-0.0412	1.167	-0.3571	1.695
VWRET	-0.0599	[-0.1196 , 0.0358]	-0.0374	1.2257	-0.3070	1.8961
EWRET	0.0438	[ 0.0029 , 0.0782]	0.0463	1.0818	0.0724	0.9705