Bayesian Variable Selection of Risk Factors in the APT model

Robert Kohn
University of New South Wales
Sydney NSW, 2052

Rachida Ouysse
University of New South Wales
Sydney NSW, 2052

September 22, 2006

Abstract

PRELIMINARY AND INCOMPLETE

In this paper we use a probabilistic approach to risk factor selection in the arbitrage pricing theory model. The methodology uses a bayesian framework to select the pervasive risk factors and estimate the model. This will enable correct inference and testing of the implications of the APT model. Furthermore, we are able to make inference on any function of the parameters, in particular the pricing errors. We can also carry out tests of efficiency of the APT using the posterior odds ratio and bayesian confidence intervals. We investigate the macroeconomic risk factors of Chen, Roll, and Ross (1986) and the firm characteristic factors of Fama and French (1992,1993). We also augment the model to allow for both measured economic variables and latent factors. Using monthly portfolio returns grouped by size and book to market, we find that the economic variables have zero risk premia although some appear to have non zero posterior probability. The ”Market” factor is not priced. An APT model with factors mimicking size (SMB), book to market equity (HML), value-weighted portfolio and Standard and Poor, is supported by a conditionally independent prior and offers a significant decrease in the pricing error over a two-factor APT with SMB and HML. The posterior probability and cumulative distributions functions of the average risk premia and the pricing errors are compared to the normal distribution. The results show that under certain conditions the distortions are very small.
1 Introduction

In this paper, we address model selection in the context of a linear factor model with potentially measured and latent factors. The study proposes an exact statistical framework for the estimation and inference in a factor model. We use a bayesian framework to implicitly incorporate model uncertainty into the estimation of the parameters and model inference.

Since the inception of the arbitrage pricing theory (APT) by Ross (1976, 1977), there is an increased interest in the use of linear factor models in the study of Asset pricing. There is a growing evidence that high returns are driven by a multifactor model rather than the one factor capital pricing model (CAPM). The APT has the attractive feature of minimal assumptions about the nature of the economy. However, this tractability comes at the cost of certain ambiguities such as an approximate pricing relation and an unknown set of pervasive factors. In order to test the implications of the APT, one must specify the number and the identity of the factors.

There are two main streams in the literature of factor selection in the APT. The first view toward model determination uses latent (unobservable) factors as sources of common variations. These common factors are estimated from sample covariance matrices using statistical techniques like factor analysis and principal components. Bai and Ng (2000) develops an econometric approach to consistently determine the dimension of the model for large panels. Bai (2001) addresses the asymptotic properties of the estimated model. This literature addresses the asymptotic properties of the distributions and therefore is based on the model selection being consistent and therefore treated as deterministic.

The second alternative view suggests the use of observed economic variables as factors. There is no doubt that asset prices are intimately linked to macroeconomic activity and that the influences go in both directions. However, little is done to formalize the search for the set of significant influential variables. Chen, Roll and Ross (1986) (CRR) asserts “A rather embarrassing gap exists between the theoretically exclusive importance of systematic “state variables” and our complete ignorance of their identity”. CRR attempts to explore this identification terrain by combining a set of “likely” macroeconomic and financial candidates for pervasive risk in asset returns. The selection procedure consists of a series of t-test for the significance of average risk premia corresponding to each of the variables allowed into the regression. The authors identify five common risk factors that are significantly priced in the stock market. Using a different approach, based on firm characteristics as a proxy for the firms’ sensitivity to systematic risk in the economy, Fama and French (1993) shows that the variation in returns on stocks and bonds can be explained by five size and book-to-market based factors. However, these studies raise two fundamental critiques. First, the number of factors is often assumed and arbitrarily prespecified. Second, the set of potential pervasive factors is subjectively reduced to a few number of candidates and only a few specifications are tested. Hence, no statistical justification is provided to justify the selected set of variables. Ouysse (2005) proposes a formal econometric procedure to consistently

\footnote{For example, Fama and French (1996) indicates that market anomalies largely disappear in a three-factor model. Ferson and Korajczyk (1995) finds that a five factors model capture a large fraction of asset returns predictability.}
select the set of pervasive factors in panels with large dimensions. However, the study does not address the post-selection properties of the model estimates.

The APT implies nonlinear restrictions on the model parameters, which make it very difficult and complex to derive the asymptotic distribution of the restricted estimates, let alone, the distribution of post-selection restricted estimates. The bayesian approach enables exact inference by making it possible to implicitly incorporate model uncertainty and to derive post-selection distributions of any functions of the parameters.

Geweke and Zhou (1996) uses a bayesian framework to analyze the APT. The authors propose the use of the pricing error to test the implication the APT that the expected returns are approximately linear function of the risk premium on systematic factors. The authors use latent factors and do not perform a selection of the appropriate number of factors. They borrow the results from the asymptotic principal component analysis of Connor and Korajczyk (1986,1993) and propose the use of 1 to 4 factors.

The present study extends [17] to allow for both latent and measured factors. In particular, this method makes it possible to derive the exact post-selection posterior distributions for the measures of the pricing error and for the measure of the systematic risk and risk premia.

2 Methodology

2.1 Asset Pricing model

In the finance literature, the debate on what drives excess returns continues. A large number of studies use factor models to identify the common sources of systematic risk in expected returns. The factors used are classified into latent factors estimated through statistical methods as principal components and factor analysis and observable factors based on the sensitivity of stocks returns to economic and financial news. The use of observable variables to explain excess returns is particularly appealing. The estimated factor loadings have a meaningful interpretation. The estimated pricing relationship can be used to stimulate the financial markets through the pervasive economic and financial variables. The ability to predict excess returns with tangible factors can be useful for portfolio management and stock market investment decisions.

Unlike the Capital Asset Pricing model, the APT allows for multiple risk factors to enter the return generating process for asset returns. In a rational asset pricing model with multiple beta, expected returns of securities are related to their sensitivities to changes in the state of the economy. Consider a world where investors are broadly diversified, but there may be multiple sources of risk in the economy. Instead of everyone caring solely about the market portfolio, investors actually care about lots of things, including shifts in stock index levels, interest rates, inflation, changes in GNP or other broad macro-economic factors that are difficult to purge from your portfolio through diversification.

The Arbitrage Pricing Theory assumes that the return generating process being con-
sidered is an approximate multi-beta k-factor structure

\[
r_{it} = \alpha_i + \sum_{j=2}^{k_0} \beta_{ij} f_{jt} + \varepsilon_{it}
\]

where \( r_{it} \) is the return on asset \( i \) at time \( t \) and \( \alpha_i \) is the intercept of the factor model. For the system of \( N \) assets,

\[
R_t = \alpha + Bf_t + \varepsilon_t
\]

where, \( E[\varepsilon_t | f_t] = 0 \) and, \( E[\varepsilon_t \varepsilon'_t | f_t] = \Sigma E[f_t] = 0 \) and, \( E[f_t f'_t] = \Omega_f \)

In the system equation \( R_t \) is \((N \times 1)\) vector and the intercept \( \alpha = [\alpha_1 .. \alpha_i .. \alpha_N]' \), \( B \) is an \((N \times K)\) matrix with \( B = [\beta_1 .. \beta_i .. \beta_N]' \), and \( \varepsilon_t \) is an \((N \times 1)\) vector with \( \varepsilon_t = [\varepsilon_1t .. \varepsilon_{it} .. \varepsilon_{Nt}]' \). The risk factors are common across the assets and are assumed to account for the common variation in asset returns. The APT is based on the pricing relation for a countably infinite vector of returns to a countably infinite set of traded assets. Consequently, the disturbance term for large well diversified portfolios vanishes. This implies that the disturbance terms be sufficiently uncorrelated across assets. The assets risk can be divided into a common diversifiable risk due to the exposure to the \( K \) common risk factors \( f_t \), and an idiosyncratic non-diversifiable risk due to the idiosyncratic factor \( \varepsilon_t \). The betas \( \beta_{ij} \), or factor loadings of the \( j^{th} \) factor for asset \( i \), represent the amount of exposure to each risk factor. There are \( T \) time periods and \( N \) assets. The APT implies that the expected returns of an asset is approximately a linear function of the risk premiums on systematic factors in the economy.

In the absence of arbitrage, Ross (1976) shows that in large economies the APT structure implies

\[
\alpha \approx \iota_N \delta_0 + B \delta_K
\]

where \( \alpha \) is the \( N \times 1 \) vector of expected returns, \( \delta_0 \) is the model zero-beta parameter and is equal to the risk free return, \( \delta_K \) is \( K \times 1 \) vector of factor risk premia. The return on the T-bill will be used as a close approximation of the zero beta rate. Equation (3) represents an approximate linear relationship between the expected asset returns and their risk exposures. Exact arbitrage pricing obtains when (3) holds with equality.

Geweke and Zhou (1996) proposes a measure of pricing error given by the average of squared deviations from the restriction across assets. The pricing error is measured by

\[
Q_N^2 = \sum_{i=1}^{N} \left( \alpha_i - \delta_0 - \beta_{i1} \delta_1 - ... - \beta_{ik} \delta_k \right)^2 / N
\]

\( Q_N \rightarrow 0 \) as \( N \rightarrow \infty \)

2 An approximate factor structure, first introduced by Chamberlain and Rothschild (1983), relaxes the static factor model assumption of diagonal idiosyncratic covariance matrix to allow for a limited amount of cross-section dependence.

3 The riskless rate will be measured by the 30-day Treasury-bill rate that is known at the beginning of each period month.

4 Connor (1984) replaces the approximation with equality under the assumption of competitive equilibrium.
The authors argue that for Connor’s equilibrium APT, \( Q \) is equal to zero. For Ross’s asymptotic APT, the pricing error will converge to zero as the number of assets approaches infinity. Although for fixed \( N \), the pricing error is not necessarily small, the authors suggest the use of the pricing error to examine some of the testable implications of the APT.

Another measure of the pricing error which takes into account the cross section correlation structure in the idiosyncratic term is given by,

\[
\tilde{Q}_N^2 = \frac{\left( \alpha - \tilde{B}' \delta \right)' \Omega \left( \alpha - \tilde{B}' \delta \right)}{N}
\]

\[
\tilde{B} = (\iota_N, \mathbf{B}); \quad \delta = (\delta_0, ..., \delta_{k_0})'
\]

\[
\tilde{Q}_N \to 0 \text{ as } N \to \infty
\]

Stated differently, the pricing theory imposes a testable cross-equation restriction on the parameters of a multivariate regression of asset excess returns on the factors. The equilibrium version of the APT implies that

\[
E[R_t] = R_{ft} \iota_N + \mathbf{B} \delta_t
\]

where \( R_{ft} \) represents the return on a risk free asset. The pricing theory imposes a testable cross-equation restriction on the parameters of a multivariate regression of asset excess returns on the factors. It implies zero intercepts in a regression of asset excess returns on the factors. A test of miss-pricing is a test for non-zero intercept, i.e., \( \alpha = 0 \).

### 2.2 Review of the Classical Inference

There are mainly two approaches to estimating and testing the APT. First, traditional factor analysis uses a likelihood ratio to test the restrictions implied by the APT. This involves computing the maximum likelihood estimates under the nonlinear restrictions in (3). This is a very difficult task in practice and the model inference is non standard. Indeed, (1) shows that the asymptotic distributions of the model parameters estimates are very complex in factor analysis, and the constrained estimates should be even more complex. This complexity makes it difficult to derive the asymptotic distribution of the likelihood ratio tests.

The second approaches surmount the complexity of the estimation under nonlinear restrictions by a two-pass approach. In the first pass, either the factor loading or the factors are estimated. In order to estimate the factors betas (factor loading) of assets, excess returns are regressed against the common factors using the time series from \( t - 60 \) to \( t - 1 \) to get the conditional betas \( \hat{\beta}_{ik,t-1} \). The estimates \( \left\{ \hat{\beta}_{ik} \right\}_{k=1,...,K} \) are then used in the second pass.

Treating these estimates as the true variables, the APT restrictions in (3) become linear constraints on the coefficients of the multivariate regression. In fact, these restrictions imply zero intercepts and can be tested using the standard methods. To estimate the
risk premia, a cross sectional regression model is utilized for each time point to get the
time series of each risk premia. For each month $t$ of the next 12 months, perform cross
section regressions:

$$r_{it} = \delta_0 + \sum_{k=1}^{K} \beta_{ik,t-1} \delta_{kt} + \varepsilon_{it}$$

with $i = 1, \ldots, N$ and get an estimate
of the sum of risk premium $\hat{\delta}_{kt}$ for month $t$ associated to variable $k$, $t = 1, \ldots, 12$. The
two-pass steps are repeated for each time period in the sample.

Estimation technique: (i) Regress excess returns on the economic variables using the
time series from $t - 60$ to $t - 1$ to get the conditional betas $\hat{\beta}_{ik,t-1}$. (ii) For each month $t$
of the next 12 months, perform cross section regressions:

$$r_{it} = \delta_0 + \sum_{k=1}^{K} \beta_{ik,t-1} \delta_{kt} + \varepsilon_{it}$$

with $i = 1, \ldots, N$ and get an estimate of the sum of risk premium $\hat{\delta}_{kt}$ for month $t$ associated
to variable $k$, $t = 1, \ldots, 12$. (iii) Steps (i) and (ii) are repeated for each year in the
sample period. The time series means of the series of risk premium estimates associated
to each variable are then tested for their significance. A factor with statistically
significant risk premia is priced by investors in the market.

This method is based on the estimates from the first pass being consistent. However, in
small samples, this procedure suffers from errors in variables problem. The uncertainty
about the first pass estimates can lead to misleading inference. Another option is to
use maximum likelihood under the assumption of joint normality of the conditional
distributions of the asset returns. The betas and risk premia are estimated under the
constrained and unconstrained model and a test for the restrictions implied by the APT
is conducted using an asymptotic chi-square test. Deriving the exact distributions of
the test statistics under the nonlinear restrictions is very hard and the asymptotic test
may have very misleading results in a finite samples. In the case of non normal and non
IID returns, a Generalized Method of Moments framework can be used to estimate
the model. Tests for the APT implications of zero intercept can be conducted using a
Wald statistic type test. In this case, the problem of errors in variables is avoided.
The tests for the nonlinear restrictions will also have size and power problems in finite
sample.

Variable selection adds an extra source of uncertainty to the model and an extra di-
mension to the complexity of the first method and to the unreliability of the inference
in the second method. Indeed, the empirical literature so far used ad hoc methods
to select the variables to enter the APT model and failed to address the issue of post
variable selection inference.

3 Bayesian Inference

3.1 Preliminaries

In our analysis, it is convenient to work with the matrix form of the system of $N$ asset
returns with $T$ observations. The model in its multivariate representation

$$R = X\Theta + \epsilon$$  \hfill (5)

where $R$ is $(T \times N)$ matrix of returns with $R = (r_{1\bullet}, \ldots, r_{i\bullet}, \ldots, r_{N\bullet})$ and, $\Theta$ is a $(K+1) \times N$
matrix representing both the loadings on the risk factors and the intercept term,$\Theta =
(\alpha B)'$. The idiosyncratic term is a $(T \times N)$ matrix with $\epsilon = (\varepsilon_{1\bullet}, \ldots, \varepsilon_{i\bullet}, \ldots, \varepsilon_{N\bullet})$. The
regressors in this multivariate model represent the risk factors and the intercept variable. \( X = [\nu_T \, F] \) where \( F = [f_1 \, ... \, f_T]' \). Pooling the cross section returns leads to the following representation which will be used jointly with (5),

\[
\mathbf{r}_{NT \times 1} = (I_N \otimes X)_{NT \times Nk} \mathbf{\theta}_{Nk \times 1} + \mathbf{\varepsilon}_{NT \times 1}
\]

where \( \mathbf{r} = \text{vec}(\mathbf{R}) \), and \( \mathbf{\varepsilon} = \text{vec}(\mathbf{\epsilon}) \) and, \( \mathbf{\theta} = \text{vec}(\mathbf{\Theta}) \).

The idiosyncratic factors are assumed to be uncorrelated with the factors, \( X \). They are also assumed to have a normal distribution with mean zero and covariance matrix \( \Sigma \otimes I_T \).

The factors in the APT are unknown but are assumed to be elements of a finite set of potential variables. Let \( K \) be the total number of potential variables represented by the columns of the matrix \( X \). Further, let \( X^0 \) be the set of pervasive factors in the true data generating process. Define the Bernoulli random variable \( \gamma_j \) as:

\[
\gamma_j = \begin{cases} 
1 & \text{if } X_j \subseteq X^0 \\
0 & \text{otherwise}
\end{cases}
\]

Therefore, \( \gamma = \{\gamma_j\}_{j=1}^K \), is a selector vector over the column of \( X \). Let \( q_\gamma \) be the "Binomial" random variable representing the number of variables in the selected model, therefore

\[
q_\gamma = \sum_{i=1,...,K} \gamma_i
\]

### 3.2 Linear Factor model with measured factors

Based on the idea that asset prices react sensitively to economic news, [9] used economic forces to proxy for the systematic influences in stock returns. Using intuition and empirical investigation, the authors combined macroeconomic variables and financial markets variables to capture the systematic risk in asset returns and suggests a five factor APT model. To assess whether the risk associated to a given variable is rewarded in the market, the authors test the significance of the estimated risk premia using a \( t - \text{statistic} \) using 20 equally weighted portfolios constructed on the basis of firm size as dependent variables. Their results show evidence of five factors. CRR concludes that the spread between long and short interest rate (UTS), expected (EI) and unexpected inflation (UEI), monthly industrial production (MP) and the spread between high- and low-grade bonds (URP) are significantly priced. However, neither the market portfolios (EWNY, VWNY) nor consumption (CG) are priced separately.

Fama and French [14] argued that size and book to market equity are related to economic fundamentals. They suggested the use of firm characteristics, such as size (ME) and book to market equity (BE/ME), to construct factors portfolios proxy for sensitivity to common risk factors in returns. The authors used slopes and \( R^2 \) values to test whether these mimicking portfolios capture shared variation in stock and bond returns. Their results show that the portfolios constructed to mimic risk factors related to ME and BE/ME capture strong variations in stock returns. Using 25 stock portfolios as dependent variables, their results show evidence that a three factor model, using Market, SMB and HML as risk factors, captures the common variations in the cross section of stock returns.

---

\(^5\text{If there is serial correlation, the covariance matrix will be of the more general form, } E(\varepsilon \varepsilon') = \Sigma \otimes \Gamma.\)
3.2.1 Uninformative Prior

The standard diffuse prior for the multivariate regression model (5) is the following prior density on the parameters \( \beta \) and \( \Sigma \):

\[
p(\theta, \Sigma) \propto |\Sigma|^{-(N+1)/2}
\]

where \(|\Sigma|\) is the determinant of the covariance matrix \( \Sigma \). With this prior, the posterior density for \( \theta \), conditional on \( r \) and \( \Sigma \), is

\[
\theta | \Sigma, r, \gamma, X \sim N\left( \hat{\theta}_\gamma, \Sigma \otimes (X'_\gamma X_\gamma)^{-1} \right)
\]

where \( \hat{\theta}_\gamma = \left[ I_N \otimes (X'X)^{-1} X' \right] r \), and \( \Omega = \Sigma^{-1} \). The posterior density for the inverse covariance matrix \( \Omega \), conditional on the data is a Wishart with shape parameter \( \delta = T - q_\gamma + N - 1 \) and scale matrix \( S_{\gamma}^{-1} \), where \( S_{\gamma} = Y'(I - X_\gamma(X'_\gamma X_\gamma)^{-1}X'_\gamma)Y \),

\[
\Omega | r, \gamma, X \sim W_N(\delta, S_{\gamma}^{-1})
\]

The variable selection process is performed by sampling from the posterior density of the indicator variable \( \gamma \). Assuming a uniform prior, \( p(\gamma_i = 1) = 0.5 \), the posterior density is given by

\[
p(\gamma | r) = p(r | \gamma) p(\gamma) / p(r) \propto p(r | \gamma) \propto |X'_\gamma X_\gamma|^{-\frac{\delta}{2}} |S_{\gamma}|^{-\frac{\delta}{2}}
\]

3.2.2 The general Normal-Wishart prior

The diffuse prior given in (7) was first introduced into bayesian multivariate analysis by Geisser and Cornfield (1963). It is a prior of 'minimum prior information'. However, if one has prior information on \( \theta \), an informative prior should be used. Indeed, Monte Carlo integration makes it possible to work with many choices of priors. In this section we will derive the full conditionals under a general form of Normal-Wishart prior. The amount of prior information and its importance relative to the sample information will be determined by the covariance matrix of the prior density of \( \beta \).

Lemma 3.1 Consider the Normal-Wishart prior for \( \theta \) and \( \Sigma \), \( \theta | \Sigma \sim N(\theta_0, \Sigma \otimes H_\theta) \) and \( \Sigma^{-1} \sim W_N(m, \Phi^{-1}) \).

Under uniform priors on \( \gamma \), the full conditionals are given by,

1. \( \theta | r, \Omega, \gamma \sim N\left( \bar{\theta}_\gamma, \Sigma \otimes D^{-1} \right) \) where \( D_\gamma = (X'_\gamma X_\gamma + H_\theta^{-1}) \) and,

\[
\bar{\theta}_\gamma = \left( I_N \otimes D_\gamma^{-1} H_\theta^{-1} \right) \theta_0 + \left( I_N \otimes D_\gamma^{-1} X'_\gamma X_\gamma \right) \hat{\theta}_{GLS}
\]

\[
\hat{\theta}_{GLS} = \left[ I_N \otimes (X'_\gamma X_\gamma)^{-1} X'_\gamma \right] r
\]

\[\text{Wishart distribution with location parameter } \Phi \text{ and scale parameter, } m. \quad \text{The first moment of } \Sigma \text{ is } E(\Sigma) = \frac{\Phi}{m-N-1}, \quad \text{and therefore the scale parameter must satisfy } m > N + 1.\]
2. \( \Omega | \tau, \gamma \sim W_N(\lambda + \tau, (S(\gamma) + \Phi)^{-1}) \) where
\[
S(\gamma) = (Y - B_0W)' \left( I_T - x_\gamma D\gamma^{-1}x'_\gamma \right) (Y - B_0W) + B_0'LB_0
\]
\[
W_\gamma = \left[ I_T - XD^{-1}X' \right]^{-1} XD^{-1}H_\beta^{-1}
\]
\[
L = \Sigma^{-1} \otimes V_\gamma
\]
\[
V_\gamma = H_\beta^{-1} - H_\beta^{-1}D^{-1}H_\beta^{-1} - H_\beta^{-1}D^{-1}X'X W_\gamma
\]
\[
\theta_0 = \text{vec}(B_0)
\]

3. \( p(\gamma | \tau, X) \propto |H_\beta|^{-\frac{T}{2}} |X'X + H_\beta^{-1}|^{-\frac{N}{2}} |\Phi|^{\frac{m}{2}} |\Phi + S|^{-\frac{(T+m)}{2}} \)

In order to implement the selection process, the hyperparameters \( \theta_0 \) and \( H_\beta \) determining the on \( \theta \) need to be specified. Assuming that no subjective information about these parameters is available, their values will be set in order to minimize their influence. Two values of the prior mean are considered in this application. The default prior \( \theta_0 = 0 \) which reflects indifference between positive and negative values and, a somewhat more informative prior which centers the prior distribution around the generalized least squares estimator \( \bar{\theta}_0 = \theta_{GLS} \).

The covariance matrix \( H_\beta \) determines the amount of information in the prior and will influence the likelihood covariance structure. In the literature, the specification is simplified to \( H_\beta = cV_\beta \). The preset form \( V_\beta \), can be chosen to either replicate the correlation structure of the likelihood by setting \( V_\beta = (X'X)^{-1} \), this is also the \( g \)-prior recommended by Zellner (1986); or to weaken the covariance in the likelihood by setting, \( V_\beta = I_K \), which implies that the components of \( \theta \) are conditionally independent. The scalar \( c \) is a tuning parameter controlling the amount of prior information. The larger the value of \( c \), the more diffuse (more flat) is the prior over the region of plausible values of \( \theta \). 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 putting increasing probability on the null model as \( c \rightarrow \infty \). In the literature, different values of \( c \) were recommended depending on the application at hand. In this paper we will consider three values of \( c \in \{4, T, K^2\} \) which represent the most used values in the literature. As pointed out in Chipman, George and McCulloch (2001), there is an asymptotic correspondence between these choices of \( c \) and the classical information criteria AIC, BIC and RIC respectively when \( V_\beta = (X'X)^{-1} \). Given that \( D = (X'X + H_\beta^{-1}) \), one can see that under \( V_\beta = I_K \) the posterior correlations will be less than those of the design correlation. The posterior correlations are however identical to those of the design matrix under the prior \( V_\beta = (X'X)^{-1} \).

The conditionally independent prior for \( \theta \),

\[
\theta | \Sigma \sim N \left( \theta_{GLS}, \Sigma \otimes cI_K \right) \quad (9)
\]

is equivalent to the prior \( B \sim N(\Sigma, cI_K) \), a matrix variate representation used by Brown et al. (1998), where \( cI_K \) is the covariance matrix of \( B | \Sigma \). Therefore, the columns of \( B \) in [5] are independent under this prior.

\[ ^7 \text{In a simulation study of the effect of the choice of } c \text{ on the posterior probability of the true model, Fernandez, Ley and Steel (2001) found that the effect depends on the true model and noise level and they recommend the use of } c = \max \{ T, K^2 \}. \]
Given this prior, the posterior densities for the parameters and the covariance matrix are given by

$$
\theta | \Sigma, r, \gamma, X \sim N\left( \hat{\theta}_\gamma, \Omega \otimes \left( \frac{1}{c} I_q + X'_\gamma X_\gamma \right)^{-1} \right)
$$

(10)

where

$$
\hat{\theta}_\gamma = \left[ I_N \otimes (X'_\gamma X_\gamma)^{-1} X'_\gamma \right] r
$$

(11)

The posterior density of the inverse covariance matrix, conditional on the sample information,

$$
\Omega | r, \gamma, X \sim W_N \left( m + T, (S(\gamma) + \Phi)^{-1} \right)
$$

(12)

where the location matrix depends on the sample data through the sum of squares residuals

$$
S(\gamma) = Y' \left[ I - X_\gamma (X'_\gamma X_\gamma)^{-1} X'_\gamma \right] Y.
$$

Given the same uniform prior on the indicator variable, the posterior density is adjusted in light of the informative prior in (9). The posterior density for \( \gamma \), conditional on the data, is

$$
p(\gamma | r, X) \propto c^{-(Nq/2)} \left| X'_\gamma X_\gamma + \frac{1}{c} I_q \right|^{-\frac{N}{2}} |\Phi|^{\frac{m}{2}} |S(\gamma) + \Phi|^{-\frac{m}{2}}
$$

(13)

Now let’s consider a different informative prior for the parameters of the model,

$$
\theta | \Sigma \sim N \left( 0, \Sigma \otimes c(X'X)^{-1} \right)
$$

(14)

The parameter \( c \) measures the amount of information in the prior relative to the sample. Setting \( c = 50 \) gives the prior the same importance as 2% of the sample. Given this new prior and given the Wishart prior, the full conditional densities for the unknown parameters of the model are given by,

$$
\theta | \Sigma, r, \gamma, X \sim N \left( \tilde{\theta}_\gamma, \frac{c}{1 + c} \Sigma \otimes (X'X)^{-1} \right)
$$

(15)

where

$$
\tilde{\theta}_\gamma = \frac{c}{1+c} \left((X'_\gamma X_\gamma)^{-1} X'_\gamma \otimes I_N \right) r.
$$

The full posterior density for the covariance matrix is a wishart,

$$
\Omega | r, \gamma, X \sim W_N \left( m + T, \left( \tilde{S}(\gamma) + \Phi \right)^{-1} \right)
$$

(16)

where

$$
\tilde{S}(\gamma) = Y' \left[ I - \frac{c}{1+c} X_\gamma (X'_\gamma X_\gamma)^{-1} X'_\gamma \right] Y.
$$

Finally, given the same uniform prior on the indicator variable, the posterior density for \( \gamma \), conditional on the data is as follows

$$
p(\gamma | r, X) \propto (1 + c)^{-\frac{Nq}{2}} |\Phi|^{\frac{m}{2}} |\Phi + \tilde{S}(\gamma)|^{-\frac{(T+m)}{2}}
$$

(17)

### 3.3 APT with Observed and Latent Factors

The model representation for the system of \( N \) returns with a combination with \( q \) observed economic factors \( X \), and \( l \) latent factor \( F \),
\[
\begin{align*}
\mathbf{r}^{(NT \times 1)} &= (I_N \otimes X) \theta^{(NT \times Nk)} (N_q \times 1) + (I_N \otimes F) \lambda^{(NT \times NI)} + \varepsilon \\
\mathbf{r}^{(NT \times 1)} &= (I_N \otimes Z) \theta^{(NT \times Nq)} (N_q \times 1) + \varepsilon; \quad q = k + r; \quad Z = [X, F]
\end{align*}
\]

There are only \(N(N+1)/2\) distinct elements in \(\Psi = V(\mathbf{r})\). The number of unknown parameters in the factor model is: \(Nq = N(q+l)\) plus the \(N(N+1)/2\) elements of the errors covariance matrix \(\Sigma\). Obviously we need to put some restrictions on the parameters to be able to estimate the model.

Let’s assume that the covariance matrix of the error terms is diagonal. That is there are only \(N\) elements to estimate. In this case, the total number of parameters is: \(N(q+l) + N\). It is necessary to have: \(N(N+1)/2 − N(q+l) − N \geq 0\). We need to have

\[
q + l \leq \frac{N - 1}{2}
\]

The latent factors and the dependent variables are assumed to be jointly normally distributed. Therefore

\[
\begin{align*}
\mathbf{f}_t &\sim N(0, \mathbf{I}_t) \\
\mathbf{r} &\sim N[(I_N \otimes X) \mathbf{\theta}, (tr(\lambda\lambda') + \Sigma) \otimes \mathbf{I}_T]
\end{align*}
\]

**Issue of Identification:** Because the latent factors are identified only up to an orthogonal rotation. We need to put some restrictions on the loading matrix \(\Lambda\). Let’s assume that \(\Lambda\) has full rank \(l\) and that the matrix composed by the first \(l\) rows is nonsingular: \(\Lambda^{(l)}\). There exists a unique orthogonal matrix \(\mathbf{P}\) such that \(\Lambda^{(l)} \mathbf{P}'\) is lower triangular with positive diagonal elements. Therefore, in order to identify the factors, we will assume that:

\[
\Lambda = \begin{bmatrix} \Lambda^{(l)} & \Lambda^{(N-l)} \end{bmatrix}
\]

where \(\Lambda^{(r)} = \begin{bmatrix} \lambda_{11} & \lambda_{21} & \ldots & \lambda_{1l} \\
0 & \lambda_{22} & \ldots & \lambda_{2l} \\
\ldots & 0 & \lambda_{ii} & \ldots & \lambda_{ii} \\
0 & 0 & 0 & \ldots & \lambda_{ll} \end{bmatrix} \)

where \(\lambda_{ii} > 0, i = 1, \ldots, l\). In terms of \(\lambda = vec(\Lambda)\). These restrictions decrease the number of free parameters to estimate in the model to \(N(N+1)/2 − N(q+l) − N + l(l-1)/2 ≥ 0\). That is the upper bound for the number of parameters in the model now is higher. In terms of the elements of \(\lambda\), these restrictions become:

**Lemma 3.2** Consider the Normal-Wishart prior for \(\theta\) and \(\Sigma\), \(\theta | \Sigma \sim N(\theta_0, \Sigma \otimes \Psi)\) and \(\Sigma^{-1} \sim W_N(m, \Phi^{-1})\). The covariance matrix of the prior distribution on the unobservable parameters is block diagonal\(^8\), \(\Psi = \begin{bmatrix} H_\theta & 0 \\
0 & H_\lambda \end{bmatrix}\). Under uniform priors on \(\gamma\), the full conditionals are given by.

\(^8\)The prior distribution assumes that the loadings on the latent factors are uncorrelated to the loading of the observed economic variables.
1. $\theta | r, \Omega, F, \gamma \sim N\left(\bar{\theta}, \Sigma \otimes \overline{H}_\theta^{-1}\right)$ where $\overline{H}_\theta = (Z'Z + \Psi^{-1})$ and

$$\bar{\theta} = \left( I_N \otimes \overline{H}_\theta^{-1} \right) \theta_0 + \left( I_N \otimes \overline{H}_\theta^{-1} Z'Z \right) \hat{\theta}_{GLS} = \left( \bar{\theta}^*, \bar{\lambda}^* \right)'$$

(18)

$$\hat{\theta}_{GLS} = \left[ I_N \otimes (Z'Z)^{-1} Z' \right] r$$

2. $\theta | r, \Omega, F, \gamma \sim N\left(\bar{\theta}, \Sigma \otimes \overline{H}_\theta^{-1}\right)$ where $\overline{H}_\theta = X'X + H_\theta^{-1} - X'F (F'F + H_\lambda^{-1}) F'X$ and $\bar{\theta}$ represents the $Nk$ first elements of $\theta$ in (18).

3. $\lambda | r, \Omega, F, \gamma \sim N\left(\bar{\lambda}, \Sigma \otimes \overline{H}_\lambda^{-1}\right)$ where $\overline{H}_\lambda = F'F + H_\lambda^{-1} - F'X (X'X + H_\theta^{-1}) X'F$ and $\bar{\lambda}$ represents the $Nr$ last elements of $\theta$ in (18).

4. Identifying restrictions:

(a) For $i = 1, \ldots, l$, $\lambda_{i-1,N+1} > 0$ and $i = 2 : l$, $\lambda_{i-1,N+1} = 0$

(b) Let $\lambda_i^* = \left\{ \lambda_{i-1,r+1:(i-1)r+i} \right\}$ for $i = 1 : r$ and $F[i]$ is the matrix of latent factors including the first $i$ columns of $F$. The posterior density for $\lambda_i^*$:

$$\lambda_i^* | r, \Omega, F, \gamma \sim N\left(\bar{\lambda}_i^*, \Sigma \otimes \overline{H}_\lambda^{-1}\right) \times 1(\lambda_{i-1,r+1} > 0)$$

$$\overline{H}_\lambda = F[i]' r[i] + \left( H[i]^{-1}_\lambda \right)^{-1} - F[i]' X' (X'X + H_\theta^{-1}) X'F[i]$$

$$H[i]_\lambda = H_{\lambda[1:i,1:i]} \tilde{\lambda}_i^* = \left\{ \tilde{\lambda}_{i-1,r+1:(i-1)r+i} \right\}$$

(c) Let $\lambda_{i+r}^* = \left\{ \lambda_{i-1,r+1:(i-1)r+r} \right\}$ for $i = r + 1 : N$,

$$\lambda_{r+1}^* | r, \Omega, F, \gamma \sim N\left(\bar{\lambda}_{i+r}, \Sigma \otimes \overline{H}_\lambda^{-1}\right)$$

$$\overline{H}_\lambda = F'F + \left( H_\lambda^{-1} \right)^{-1} - F'X (X'X + H_\theta^{-1}) X'F$$

$$H[i]_\lambda = H_{\lambda[1:i,1:i]}$$

$$\tilde{\lambda}_{i+r}^* = \left\{ \tilde{\lambda}_{i-1,r+1:(i-1)r+r} \right\}$$

5. $\Omega | r, F \sim W_N(m + T, (S_F(\gamma) + \Phi)^{-1})$ where,

$$S = (Y - \Theta_0 W)' \left( I_T - Z\overline{H}_\theta^{-1} Z' \right) (Y - \Theta_0 W) + \Theta_0' L \Theta_0$$

$$W = \left[ I_T - Z\overline{H}_\theta^{-1} Z' \right]^{-1} X\overline{H}_\theta^{-1} \Psi^{-1}$$

$$V = \Psi^{-1} - \Psi^{-1} \overline{H}_\theta^{-1} \Psi^{-1} - \Psi^{-1} \overline{H}_\theta^{-1} Z'W$$

$$L = \Sigma^{-1} \otimes V$$

$$\theta_0 = vec(\Theta_0)$$

6. $p(r | F, \gamma) \propto |\Psi|^{-\frac{m}{2}} |Z'Z + \Psi^{-1}|^{-\frac{m}{2}} |\Phi|^{-\frac{(T+m)}{2}}$
4 Empirical Results

4.1 Measured Economic and Firm Characteristics Variables

Based on the idea that asset prices react sensitively to economic news, Chen, Roll and Ross (1986) (CRR) uses economic forces to proxy for the systematic influences in stock returns. Using intuition and empirical investigation, CRR combines macroeconomic variables and financial markets variables to capture the systematic risk in asset returns and suggests a five factor APT model. Using a set of 20 equally weighted portfolios constructed on the basis of firm size as dependent variables, the authors apply Fama and MacBeth (1973) two-step estimation procedure to estimate the average risk premia associated to the variables included in each regression. To assess whether the risk associated to a given variable is rewarded in the market, they test the significance of the estimated risk premia using a \( t \)-statistic. Results show evidence of five factors: CRR concludes that the spread between long and short interest rate (UTS), expected (EI), unexpected inflation (UEI), and change in expected inflation (DEI). Other economic variables included are, monthly industrial production (MP) and the spread between high- and low-grade bonds (URP) are significantly priced. However, neither the market portfolio (EWNY, VWNY) nor consumption (CG) are priced separately. Furthermore, they found no evidence that oil prices (OG) are rewarded separately. The APT relation is derived under the assumption that there is an infinite number of assets. In finite economies, a bound is placed on the deviations of the individual assets expected returns from the APT pricing in (3), see Dybvig (1983), Grinblatt and Titman (1983). As a result, the set of pervasive factors needs to be exhaustive and complete as possible. Following the work of McElroy and Burmeister (1988) and the subsequent work of De la Calle (1991), a return on a well-diversified portfolio is employed to mimic the series of potentially missing factors.

Fama and French (1992) documents that size and book to market equity are related to economic fundamentals. This motivates the use of these firm characteristics to construct factors portfolios. Fama and French (1993) suggests that variables that are related to average returns, such as size (ME) and book to market equity (BE/ME) must proxy for sensitivity to common risk factors in returns. The authors use slopes and \( R^2 \) values to test whether these mimicking portfolios capture shared variation in stock and bond returns. Their results show that the portfolios constructed to mimic risk factors related to ME and BE/ME capture strong variations in stock returns. Using 25 stock portfolios as dependent variables, their results show evidence that a three factor model, using Market, SMB and HML as risk factors, captures the common variations in the cross section of stock returns.

In this application we consider the monthly value weighted returns for the intersections of 10 ME portfolios and 10 BE/ME portfolios from Fama and French. The portfolios are constructed at the end of June. ME is market cap at the end of June. BE/ME is book equity at the last fiscal year end of the prior calendar year divided by ME at the end of December of the prior year. The sample period considered

---

The authors use portfolios instead of individual stocks in order to control for the errors-in-variables problems that arises from the use of the two steps estimation technique.
in the variable selection is $1960:10 - 2000:12$. We also report results on the sub-period $1980:01 - 2000:12$. These two periods are used estimation. In a later section, we will assess the performance of the most promising models over the period $2000:01 - 2005:12$ to explain the cross section of expected returns.

Given the Normal-Wishart priors described in section 3.1.2, we use a Gibbs sampler to draw an ergodic Markov chain sequence for $\gamma$, $\theta$ and $\Omega$. These parameters are drawn from their full conditional distributions as described in Lemma 1. The posterior distribution of the average pricing errors are then easily obtained from these samples of the model parameters. The following are the quantities of interest computed from the MCMC sequences,

- The posterior mean of the indicator variable $\gamma$, denoted $\gamma_r$. The elements in which $\gamma_r$ will represent the posterior probability of each variable being in the true data generating process.

- Iterates for the risk premia, $\delta_i$ for each variable $i$ with nonzero posterior probability. In addition of reporting their posterior mean,$\delta_r$, and standard deviation, $\text{Std}(\delta_r)$, we plot their histogram and a kernel density estimator of the posterior distribution. To assess significance of the risk premia, we also compute the bayesian confidence intervals for 90, 95 percent confidence level.

- Since variable selection will depend on the loss function considered, we report the model estimates for a selection based on the highest posterior model probability $p(r|\gamma)$ as well as the results for a selection based on minimizing the pricing errors. As a notation, we use $\delta_{\text{max}}p(r|\gamma)$ and $\delta_{\text{min}}Q$ respectively.

- Iterates for the pricing errors, $Q^2$ and $\tilde{Q}^2$, their histogram and kernel density estimate of their posterior distribution as well as the bayesian confidence intervals.

- It is of interest to examine the expected returns, the systematic risks, and the unsystematic risks in the APT model. We provide bayesian point estimates of the expected asset returns computed as the posterior means of the

<table>
<thead>
<tr>
<th>Measured Variables</th>
<th>Measured Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption</td>
<td>CG</td>
</tr>
<tr>
<td>Term Structure</td>
<td>UTS</td>
</tr>
<tr>
<td>Risk Premium</td>
<td>URP</td>
</tr>
<tr>
<td>Inflation</td>
<td>I</td>
</tr>
<tr>
<td>Expected Inflation</td>
<td>EI</td>
</tr>
<tr>
<td>Unexpected Inflation</td>
<td>UEI</td>
</tr>
<tr>
<td>Change in EI</td>
<td>DEI</td>
</tr>
<tr>
<td>Prod Growth (M)</td>
<td>MP</td>
</tr>
<tr>
<td>Prod Growth (A)</td>
<td>YP</td>
</tr>
<tr>
<td>Oil Prices</td>
<td>OG</td>
</tr>
<tr>
<td></td>
<td>UNEP</td>
</tr>
<tr>
<td></td>
<td>Private Savings</td>
</tr>
<tr>
<td></td>
<td>Money Growth</td>
</tr>
<tr>
<td></td>
<td>Missing Factor</td>
</tr>
<tr>
<td></td>
<td>Value-Weighted return</td>
</tr>
<tr>
<td></td>
<td>Standard and Poor</td>
</tr>
<tr>
<td></td>
<td>Market portfolio</td>
</tr>
<tr>
<td></td>
<td>Small Minus Big</td>
</tr>
<tr>
<td></td>
<td>High Minus Low</td>
</tr>
</tbody>
</table>


\[ \alpha_i, \text{i.e.} \alpha_{pm}^{[j]} = \frac{1}{M} \sum_{j=1}^{M} \alpha_i^{[j]} \]. The posterior mean, \( \Sigma_{pm} \), of the \( \Sigma \) iterates represents the bayesian estimate for the idiosyncratic risks. The bayesian estimate for the total risk is given as the sum of the estimate for the systematic risk \( \Theta_{pm}'X'_ \theta \Theta_{pm} \) and the estimate for the non-diversifiable \( \Sigma_{pm} \). We report the proportion of systematic risk to the total risk denoted \( D_1 \) and given the ratio of the diagonal elements of the systematic risk to the diagonal elements of the total risk, \( D_1 = \frac{\text{diag}(\Theta_{pm}'X'X\Theta_{pm})}{\text{diag}(\Theta_{pm}'X'X\Theta_{pm} + \Sigma_{pm})} \). We also report the proportion of bayesian total risk to the sample covariance of observed returns, \( D_2 = \frac{\text{diag}(\sigma_{pm}^2X'X\Theta_{pm} + \Sigma_{pm})}{\text{diag}(V(Y))} \), where \( \Theta_{pm} \) is posterior mean of the iterates of \( \Theta \), which is a bayesian estimate for the factors betas \( \Theta \).

- To assess the convergence of the Gibbs sampler, we plot the autocorrelation function for the risk premia and the pricing errors iterates.

- Finally, we use kernel density to estimate the posterior probability distribution of the risk premia and pricing errors. These densities along with the cumulative distributions are then plotted against the standard normal distribution.

We set the warming period to 10000 iterations and the sampling period to 4000. The initial values for the indicator variable are one for the constant term and zero for all other variables. The results were unchanged with different starting values. The initial value for the idiosyncratic covariance matrix is \( 0 \times I_N \). The location matrix for the Wishart distribution is \( \Phi = I_N \) and the scale parameter \( m = N + 2 \). The following points summarize the main results:

1. The most favored model using the \( g \)-prior \( H_{\theta} = c (X'X)^{-1} \) with \( c = \max\{T, K^2\} \) is the APT with the two factors \( \{SMB, HML\} \). Table 1 shows that the two factors have a posterior probability of one to be in the DGP. The point estimate for the risk premium for SMB is 0.01% while the risk premium associated to the factor HML is about 1.9%. None of the measured economic variables nor the financial variables were pervasive. Table 2 represents the results for the conditionally independent prior \( H_{\theta} = c I_K \). The most favored factors are \( \{VWRET, SP500, SMB, HML\} \) all with 100% posterior probability to be in the DGP. The money growth, \( GB \) has only a 0.05% posterior odds to be part of the pervasive set of factors, while the monthly production growth, \( MP \), appears to have a 0.05% probability. Both \( GB \) and \( MP \) have zero risk premia. The posterior means of the risk premia for SMB and HML are negative and are equal to \(-0.0022\) and \(-0.0014\) respectively. The \( VWRET \) has a negative risk premia of \(-0.006\), which is slightly smaller than point estimate for the risk premia on the \( SP500 \) which is equal to \(-0.0042\).

2. In order to compare the two favored models using the two different priors we first look at the pricing errors. Table 3 is a summary of the posterior means, standard deviation and confidence intervals for the pricing errors for the two types of priors. With the conditionally independent prior, the mean pricing error for the most favored model with the factors \( \{VWRET, SP500, SMB, HML\} \) is 0.0056% for \( \hat{Q}^2 \) (resp. 0.0347% for \( Q^2 \)) compared to 0.0071% (resp. 0.055% for...
$Q^2$) for the most favored model under the g-prior with factors $\{SMB, HML\}$. The 4-factor APT shrinks the pricing error by about 21% (resp. 36% for $Q^2$). To assess the economic importance of the pricing errors, we will follow Geweke & Zhou (1996) argument and compare the magnitude of the mean of $Q$ with the monthly expected returns. For the sample period 1960-2000, the means range from 0.7015 to 1.6584 percent per month. One can regard the above means of the pricing errors as economically negligible. To further assess the pricing error we provide the 90% (95%) Bayesian confidence interval, which state that there is 90% (95%) probability that the pricing errors in the interval. The smaller the confidence interval, the more heavily concentrated the posterior density of the average pricing error near its mean and the more informative are the data on the pricing error.

Table 3. Pricing Errors for the two types of g-priors. Case of combined set of measured factors, $K = 22, N = 45$ and $c = T$.

<table>
<thead>
<tr>
<th>Panel A: Sample Period 1960 : 01 – 2000 : 12, $T = 480$</th>
<th>$cI_{Kc}$</th>
<th>$c(X'X)^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q$</td>
<td>$Q$</td>
</tr>
<tr>
<td>$E(Q</td>
<td>r)$</td>
<td>0.0347</td>
</tr>
<tr>
<td>$Std$</td>
<td>0.0012</td>
<td>0.0005</td>
</tr>
<tr>
<td>90% CI</td>
<td>[0.0324; 0.0360]</td>
<td>[0.0049; 0.0064]</td>
</tr>
<tr>
<td>95% CI</td>
<td>[0.0317; 0.0361]</td>
<td>[0.0047; 0.0065]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Sample Period 1980 : 01 – 2000 : 12, $T = 240$</th>
<th>$Q$</th>
<th>$Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(Q</td>
<td>r)$</td>
<td>0.0292</td>
</tr>
<tr>
<td>$Std$</td>
<td>0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td>90% CI</td>
<td>[0.0693; 0.0816]</td>
<td>[0.0123; 0.0187]</td>
</tr>
<tr>
<td>95% CI</td>
<td>[0.0672; 0.0818]</td>
<td>[0.0117; 0.0194]</td>
</tr>
</tbody>
</table>

3. The average proportion $D1$ for the 4-factor model is 99.96% and an average of 93.34% for the proportion $D2$. The two-factor model has an average proportion of 99.74% for $D1$ and 49.02% for $D2$. Figure 1 and Figure 2 show that the 4-factor has both a higher $D1$ and $D2$ compared to the two-factor model for all asset returns in the sample. The higher the proportion $D1$, the smaller is the idiosyncratic risk relative to the systematic risk. Both models have proportions above 99.5%. However, the gap between the sample variances of the asset returns and the estimated total risk is far greater in the two-factor compared to the 4-factor APT.

4. In terms of matching the average returns, Figure 1 shows that the design matrix prior appears to produce estimates that are very similar to the classical approach. The independent prior. produces estimates for the expected returns that are quantitatively very low compared to the data dependent prior and the sample means.
Figure 1: Plot of the alphas for both priors against the average returns. Sample period 1960-2000 with $c=T$. 
### Table 1. Risk Premium for the combined set of measured factors, 1960 – 2000.

$T = 480$, $K = 22$, $N = 43$ and $c = T$.

| Panel A: $H_0 = cI_K$ | \( \gamma | r \) | \( E(.|r) \) | Std | \( \delta_{max}(p|r) \) | \( \delta_{min}(Q) \) | 90% CI | 95% CI |
|-------------------------|-----------------|-----------------|-----|-----------------|-----------------|-------|-------|
| \( \delta_0 \)         | 1               | 1.8378          | 0.0943 | 1.9130          | 1.8985          | [1.6891; 1.9940] | [1.6576; 2.0391] |
| \( \delta_{GB} \)      | 0.0005          | -0.000          | 0.000  | -0.000          | -0.000          | NA    | NA    |
| \( \delta_{VWRET} \)   | 1               | -0.0060         | 0.0009 | -0.0069         | -0.0064         | [-0.0076; -0.0046] | [-0.0080; -0.0043] |
| \( \delta_{SP500} \)   | 1               | -0.0042         | 0.0010 | -0.0051         | -0.0048         | [-0.0059; -0.0027] | [-0.0063; -0.0023] |
| \( \delta_{MP} \)      | 0.0003          | -0.000          | 0.000  | -0.000          | -0.000          | NA    | NA    |
| \( \delta_{HML} \)     | 1               | -0.0014         | 0.0015 | -0.0013         | -0.0081         | [-0.0038; 0.0009] | [-0.0045; 0.0014] |
| \( \delta_{SMB} \)     | 1               | -0.0022         | 0.0011 | -0.0007         | -0.0020         | [-0.0040; -0.0004] | [-0.0045; -0.0000] |

| Panel B: $H_0 = c(X'X)^{-1}$ | \( \gamma | r \) | \( E(.|r) \) | Std | \( \delta_{max}(p|r) \) | \( \delta_{min}(Q) \) | 90% CI | 95% CI |
|-----------------------------|-----------------|-----------------|-----|-----------------|-----------------|-------|-------|
| \( \delta_0 \)             | 1               | 1.2625          | 0.0113 | 1.2532          | 1.3027          | [1.2441; 1.2813] | [1.2390; 1.2853] |
| \( \delta_{HML} \)         | 1               | 0.0019          | 0.0027 | 0.0108          | 0.0018          | [-0.0026; 0.0064] | [-0.0034; 0.0074] |
| \( \delta_{SMB} \)         | 1               | 0.0001          | 0.0027 | 0.0035          | 0.0123          | [-0.0004; 0.0045] | [-0.0052; 0.0055] |

### Table 2. Risk Premium for the combined set of measured factors, 1980 – 2000.

$K = 22$, $T = 240$, $N = 45$ and $c = K^2$.

| Panel A: $H_0 = cI_K$ | \( \gamma | r \) | \( E(.|r) \) | Std | \( \delta_{max}(p|r) \) | \( \delta_{min}(Q) \) | 90% CI | 95% CI |
|------------------------|-----------------|-----------------|-----|-----------------|-----------------|-------|-------|
| \( \delta_0 \)         | 1               | 2.0215          | 0.0753 | 2.0557          | 1.9707          | 1.8925; 2.1405] | 1.8608; 2.1675 |
| \( \delta_{GB} \)      | 1               | -0.0014         | 0.0002 | -0.0014         | -0.0011         | [-0.0016; -0.0011] | [-0.0017; -0.0010] |
| \( \delta_{VWRET} \)   | 1               | -0.0069         | 0.0007 | -0.0073         | -0.0063         | [-0.0081; -0.0056] | [-0.0083; -0.0053] |
| \( \delta_{SP500} \)   | 1               | -0.0049         | 0.0008 | -0.0053         | -0.0047         | [-0.0061; -0.0036] | [-0.0063; -0.0033] |
| \( \delta_{MP} \)      | 1               | -0.0029         | 0.0001 | -0.0029         | -0.0027         | [-0.0031; -0.0027] | [-0.0031; -0.0027] |
| \( \delta_{I} \)       | 1               | -0.0067         | 0.0003 | -0.0067         | -0.0063         | [-0.0071; -0.0063] | [-0.0072; -0.0061] |
| \( \delta_{UEI} \)     | 1               | -0.0058         | 0.0003 | -0.0059         | 0.0045          | [-0.0063; -0.0054] | [-0.0065; -0.0053] |
| \( \delta_{HML} \)     | 1               | -0.0020         | 0.0015 | -0.0028         | -0.0014         | [-0.0044; 0.0006] | [-0.0049; 0.0011] |
| \( \delta_{SMB} \)     | 1               | -0.0047         | 0.0014 | -0.0056         | -0.0085         | [-0.0069; -0.0024] | [-0.0073; -0.0019] |

| Panel B: $H_0 = c(X'X)^{-1}$ | \( \gamma | r \) | \( E(.|r) \) | Std | \( \delta_{max}(p|r) \) | \( \delta_{min}(Q) \) | 90% CI | 95% CI |
|-----------------------------|-----------------|-----------------|-----|-----------------|-----------------|-------|-------|
| \( \delta_0 \)             | 1               | 0.3915          | 0.0224 | 1.3976          | 1.3209          | 1.3537; 1.4283] | 1.3469; 1.4367 |
| \( \delta_{VWRET} \)       | 1               | 0.0012          | 0.0044 | 0.0018          | 0.0130          | [-0.0059; 0.0085] | [-0.0073; 0.0099] |
| \( \delta_{HML} \)         | 1               | 0.0039          | 0.0059 | -0.0001         | 0.0266          | [-0.0058; 0.0133] | [-0.0078; 0.0155] |
| \( \delta_{SMB} \)         | 1               | -0.0007         | 0.0045 | -0.0051         | 0.0019          | [-0.0081; 0.0066] | [-0.0095; 0.0079] |

### 5 Conclusion

In this article, we propose a fully bayesian framework for selecting the risk factors and examining their risk premia and the pricing restrictions implied by the APT. This a one step approach which integrates the uncertainty behind model selection and the
estimation of the different functions of the parameters. In contrast to existing studies, we do not fix a priori the number of measured variables allowed to enter the pricing relationship. The number of measured variables and statistical factors is endogenously determined. This process is performed simultaneously with the estimation of the factor betas and their risk premia. Hence, this method avoids the errors in variables problem encountered in the main stream two-pass approach of Fama-MacBeth. Because, the bayesian approach evaluates the exact posterior distribution of the estimated parameters and any other function of the parameters, we are able to produce bayesian confidence intervals for the risk premia to gage if the market does price a certain factor. Inference is also done on the average pricing errors in order to evaluate the extent to which the APT restrictions deviate from the data. In an APT with only measured economic variables are allowed along with Fama and French three factors, the choice of the prior on the factor betas influences the posterior distribution of the promising factors. In the case of Zellner g-prior where the prior covariance matrix is a replica of the design matrix, the pervasive factors are Fama and French size and book to market risk factors $SMB$ and $HML$. However, using the conditionally dependent prior, in addition to $SMB$ and $HML$, some economic variables appear to be priced by the market. More specifically, inflation, unexpected inflation, return on value weighted portfolio and return on the standard and poor portfolio.

References


6 APPENDIX:

6.1 Data Appendix

- Fama and French Portfolio Factors.

First the stock returns are ranked on size (prices times shares). The median size is then used to split the stocks into two groups, small and big (S and B). The returns are also broken into three book-to-market groups based on the bottom 30% (L), middle 40% (M) and the top 30% (H). Six portfolios (S/L, S/M, S/H, B/L, B/M, B/H) are then constructed from the intersection of the two size groups and the three BE/ME groups. Two additional portfolios are constructed from these intersections. HML (high minus low) meant to mimic the risk factor in return related to book to market equity BE/ME. It is defined as the difference between the average return on the two high-BE/ME portfolios and the average returns on the two low-BE/ME. The portfolio SMB (small minus big) meant to mimic the risk factor in return related to size. It is the difference between small and big stocks with about the same book to market equity. Finally, a value-weighted portfolio on the six size-BE/ME portfolio to proxy for market factor. To simplify notations, the set of factors used in Fama and French (1993) will be denoted $F^F$.

- Chen, Roll and Ross macroeconomic factors

From Roll and Ross (1986), the following set of variables is constructed: - Consumption growth $CG$: growth rate in real per capita consumption constructed by dividing the series of seasonally adjusted real personal consumption (excluding durables) by the population. The series are from $FRED$ (Federal reserve bank of St. Louis).
- Term structure of interest rate $UTS$: the spread between the return on a long term government bond and the lagged return on one month bills. The two series are from $CRSP$ US Treasury and Inflation Indices.
- Risk premium $URP$: the spread between the return on low grade bonds (Moody’s seasonally adjusted $Baa$ corporate bond yield) and a long term government bond.
- Monthly growth of industrial production $MP(t)$: measured the change in industrial production lagged by one month.
- Annual growth of industrial production $YP(t)$.
- Unexpected inflation $UEI(t) = I(t) - E(I(t)|t - 1)$.
- Change in expected inflation $DEI(t) = E(I(t + 1)|t) - E(I(t|t - 1))$.
- Financial market indices: the return on value weighted $VWRET$ and equally weighted $EWRTD$ portfolios of $NYSE$ listed stocks.

To these variables, an additional set of potential sources of variation are added:
- The growth rate of money base GB: Currency component of money stock plus demand deposits seasonally adjusted (FRED).
- Private saving rate PSAVE: Percent, seasonally adjusted. All the variables used in CRR are collected in a set denoted by FRR.

6.2 Gibbs Sampler

The Gibbs sampler generates an ergodic Markov chain,

$$\gamma^{(0)}, \gamma^{(1)}, \Omega^{(1)}, \theta^{(1)}, \ldots, \gamma^{(j)}, \Omega^{(j)}, \theta^{(j)}, \ldots, \gamma^{(M)}, \Omega^{(M)}, \theta^{(M)}$$

Except for $$\gamma^{(0)}$$ which is initialized as $$\gamma^{(0)} = (1, 0, 0, \ldots, 0)$$, the subsequent values of $$\gamma^{(j)}, \Omega^{(j)}, \theta^{(j)}$$ are obtained by successively simulating values according to the following iterated scheme.

1. Given $$\gamma^{(j-1)}$$, the next iterate $$\gamma^{(1)}$$ is obtained by sampling from $$p(\gamma | r, X)$$

$$p_k = P(\gamma_k = 1 | \gamma_k, r, X) = \frac{L}{1 + L}$$

$$L \propto \left( \frac{|H_\theta(1)|}{|H_\theta(0)|} \right)^{-\frac{N}{2}} \left( \left| \frac{X'_\gamma(1)X_\gamma(1) + H_\theta(1)^{-1}}{X'_\gamma(0)X_\gamma(0) + H_\theta(0)^{-1}} \right| \left( \frac{|\Phi + S(1)|}{|\Phi + S(0)|} \right)^{-\frac{(T+m)}{2}}$$

Draw $$u \sim U(0, 1)$$

if $$u < p_k$$ then $$\gamma_k^{(j)} = 1; \gamma^{(j)} = \gamma^{(j)}$$ and $$S^{(j)} = S(1)$$

(a) else $$\gamma_k^{(j)} = 0; \gamma^{(j)} = \gamma^{(j)}$$ and $$S^{(j)} = S(0)$$ end else

2. Given $$\gamma^{(j)}$$, draw $$\Omega^{(j)}$$ by sampling from the Wishart distribution $$\Omega^{(j)} | r, \gamma^{(j)} \sim W_N(m + T, (S(\gamma^{(j)}) + \Phi)^{-1})$$, where

$$S(\gamma^{(j)}) = Y' \left( I_T - X_{\gamma^{(j)}} D_{\gamma^{(j)}}^{-1} X'_{\gamma^{(j)}} \right) Y$$ if we take a prior with $$\theta_0 = 0$$

$$D_{\gamma^{(j)}} = \left( X'_{\gamma^{(j)}} X_{\gamma^{(j)}} + H_\theta^{-1} \right)$$

3. Given $$\gamma^{(j)}, \Omega^{(j)}$$, draw $$\theta^{(j)}$$ by random sampling from $$\theta^{(j)} | r, \Omega^{(j)}, \gamma^{(j)} \sim N \left( \tilde{\theta}_{\gamma}, \Sigma^{(j)} \otimes D_{\gamma^{(j)}}^{-1} \right)$$

where $$\Sigma^{(j)} = (\Omega^{(j)})^{-1}$$ and

$$\tilde{\theta}_{\gamma^{(j)}} = \left( I_N \otimes D_{\gamma^{(j)}}^{-1} H_\theta^{-1} \right) \theta_0 + \left( I_N \otimes D_{\gamma^{(j)}}^{-1}, X_{\gamma^{(j)}}' X_{\gamma^{(j)}} \right) \hat{\theta}_{GLS}$$
4. Compute the risk premia iterates \( \delta^{(j)} = \left( \widetilde{B}^{(j)r} \widetilde{B}^{(j)} \right)^{-1} \widetilde{B}^{(j)r} \alpha^{(j)} \) and the pricing errors iterates

\[
\begin{align*}
\tilde{Q}^{2(j)}_N &= \frac{\alpha^{(j)} \left( I_N - \widetilde{B}^{(j)} \left( \widetilde{B}^{(j)r} \widetilde{B}^{(j)} \right)^{-1} \widetilde{B}^{(j)r} \right) \alpha^{(j)}}{N} \\
\tilde{B}^{(j)} &= \left( I_N - \widetilde{B}^{(j)} \left( \widetilde{B}^{(j)r} \widetilde{B}^{(j)} \right)^{-1} \widetilde{B}^{(j)r} \right) \alpha^{(j)} \\
Q^{2(j)}_N &= \frac{\alpha^{(j)} \left( I_N - \widetilde{B}^{(j)} \left( \widetilde{B}^{(j)r} \widetilde{B}^{(j)} \right)^{-1} \widetilde{B}^{(j)r} \right) \alpha^{(j)}}{N}
\end{align*}
\]

6.3 Reverse Jump MCMC

Reversible jump MCMC is a generalization of the Metropolis Hastings algorithm which introduces moves between parameter spaces of different dimensionality while retaining the detailed balance property, which is required for convergence within each type of move. Basically, there are within model moves (the usual MCMC to update parameters of the model given a dimension) and between models moves which update the model dimension.

Let \( r \) be the number of latent factors in the model: \( r \leq r_{\text{max}} \). Assume a uniform prior over the dimension space, that is:

\[
p(r) = \frac{1}{r_{\text{max}}}
\]

The acceptance probabilities (from \( x \) to \( r \)) in the M-H algorithm are defined as follows:

\[
\alpha(x \rightarrow r) = \min \left\{ 1, \frac{\pi(r) q(x | r)}{\pi(x) q(r | x)} \right\}
\]

where \( \pi(\cdot) \) is the target density and \( q(r, x) \) is the proposal density for a move from \( r \) to \( x \). The RJMCMC proposes an acceptance probability for interdimension moves as follows:

\[
\alpha((\lambda_r, r) \rightarrow (\lambda_{r'}, r')) = \min \left\{ 1, \frac{p(\lambda_{r'}, r'| r) q(\lambda_{r'}, r | \lambda_r, r) J(r' \rightarrow r)}{p(\lambda_r, r | r') q(\lambda_r, r' | \lambda_{r'}, r) J(r \rightarrow r')} \right\}
\]

\[
p(\lambda_{r'}, r'| r) = p(\lambda_{r'} | r, r') p(r' | r) = p(\lambda_{r'} | r', r) p(\lambda_r | r')
\]

\[
= \min \left\{ 1, \frac{p(\lambda_r | r, r') p(\lambda_{r'} | r') q(\lambda_r | \lambda_{r'}, r, r) J(r' \rightarrow r)}{p(\lambda_r, r) p(\lambda_r | r) q(\lambda_{r'} | \lambda_r, r', r) J(r \rightarrow r')} \right\}
\]

Transition probability from state \( k \) to state \( k' \)

We can start by assuming that the proposed move does not depend on the state information other than its dimension.
\[ q(r' | \lambda_r, r, r) = q(r' | r) = J(r \rightarrow r') \]

If the current dimension is \( r \), the choice of model move is determined by a transition density \( J(r \rightarrow r') \); the latter is a conditional probability of moving to state \( r' \) given current state is \( r \). Lets assume a discretised Laplacian density

\[ J(r \rightarrow r') \equiv p(r' | r) \propto \exp \left( -\eta |r' - r| \right) \]

**Proposal densities**

Following Lopes and West (2004), we will use a preliminary set of parallel MCMC that are run over a set of prespecified values \( r \) of the number of factors. These chains will produce within model posterior samples for \((\lambda_r, F_r)\) that approximate the distributions \((\lambda_r, F_r | r, r)\). From these samples we compute the posterior means and use these to guide the choice of analytically specified distributions to be used to generate candidate parameter values. We use the posterior means and posterior variance from these preliminary MCMC as parameters of the proposal densities.

**RJMCMC Algorithm** Lets assume the number of potential observed variables is known \( K \). There is uncertainty on the potential number of latent factors \( r = \{0, 1, 2, ..., r_{\text{max}}\} \)

1. • Estimate the proposal densities using a preliminary set of parallel MCMC
   • Initialize the model. Start value for \( \{\eta_r(0)\} \)
   • Set the current values of \( \lambda_r^{(0)} \) to a draw from the posterior (proposal density) \( q_r(\lambda_r) = p(\lambda_r | r^{(0)}, r) \). This step will also produce a set of values for the latent factors \( F_r^{(0)} \)

2. iteration \( j = 1 : M \)
   a. i. Draw a candidate value of \( r' \) from the proposal distribution \( J(r^{(j-1)} \rightarrow r') \)
      ii. Draw a set of parameters values from \( q_r(\eta_r) = p(\eta_r | r', r) \), the proposal density, \( \eta_r = (\lambda, \sigma_r^2) \)
      iii. sample \( u \sim U(0, 1) \)
      iv. if \( u < \alpha \left( (\lambda_r^{(j-1), r^{(j-1)}}, r^{(j-1)}) \rightarrow (\lambda_r^{(j)}, r^{(j)}) \right) \)
         \[ = \min \left\{ 1, \frac{p(\eta_r | \eta_r', r') p(\eta_r | r') q_r(\eta_r') J(r' \rightarrow r)}{p(\eta_r | \eta_r', r) p(\eta_r | r) q_r(\eta_r) J(r \rightarrow r')} \right\} \]
         where \( q_r(\eta_r) = p(\eta_r | r, r) \).
      v. else remain in current state \( r^{(j)} = r^{(j-1)} \), end if
      vi. given the state \( r^{(j)} \) run an MCMC step to update and generate iterates for the model parameters and the latent factors.
         A. Update \( F : f_{\theta} [x | \eta_r^{(j)}, \theta^{(j-1)}] \)
         B. Update \( \Omega^{(j)} \) and \( \lambda_r^{(j)} \): Draw the covariance iterate \( \Omega^{(j)} \) from \( \Omega | r, F_r^{(j)}, r^{(j)} \)
         C. Update \( \lambda_r^{(j)} : p \left( \lambda_r | r, \Omega^{(j)}, F_r^{(j)} \right) \)
D. Draw the iterates $\theta^{(j)}$ (coefficient for the observed factors) using
$\theta|F^{(j)}, r^{(j)}, \Omega^{(j)}, r$
Use a Gibbs sampler:
$\gamma = \{\gamma_l\}_{l=1:K}$

**FOR** $l = 1 : K$

Draw $u \sim U(0, 1)$

Compute the posterior odds of $\gamma_l = \{0, 1\}$

Let $\gamma_{(l)} = \gamma|\gamma_l = 1, \gamma_{-l} = \gamma_{(j-1)}$

Now, in our model, we only have the density of $r$ conditional on the latent factors.

\[
p(r|F, \gamma) \propto c^{-\frac{Nq}{2}} |X'X|^{-\frac{N}{2}} |M_{Z_l}|^{-\frac{N/2}{2}} |\Pi^-|^{-\frac{2(N+T+2)}{2}}
\]

\[
k_{l} = \sum_{i=1,...,k} \delta_i \quad \text{and} \quad r_{l} = \sum_{i=1,...,r} \tau_i
\]

where $\delta_i = 1$ if $X_i$ in the model
and $\tau_i = 1$ if $F_i$ in the model

\[
P\left(\gamma_l = 1|r, F^{(j)}, \gamma_{1:l-1}, \gamma_{l+1:K}\right) \propto p\left(\gamma_l = 1|F = F^{(j)}, \gamma_{1:l-1}, \gamma_{l+1:K}\right)
\]

\[
\log(P\left(\gamma_l = 1|r, F^{(j)}\right)) \propto -\frac{(Nq_{\gamma^{(1)}})}{2} \log(c)
\]

\[
\log(P\left(\gamma_l = 0|r, F^{(j)}\right)) \propto -\frac{(Nq_{\gamma^{(0)}})}{2} \log(c)
\]

\[
\log L = \log(P\left(\gamma_l = 1|r, F^{(j)}\right)) - \log(P\left(\gamma_l = 0|r, F^{(j)}\right)) \propto -\frac{N}{2} \log(c) - \frac{N}{2} \times \left[\log |M_{Z_{\gamma^{(1)}}}| - \log |M_{Z_{\gamma^{(0)}}}|\right]
\]
1. (a) i. If $u < P$ then $\gamma^{(j)} = 1; \gamma^{(j)} = \gamma_{(i)}$ and $S^{(j)} = S^{(i)}$
else $\gamma^{(j)} = 0; \gamma^{(j)} = \gamma_{(0)}$ and $S^{(j)} = S^{(0)}$ end else
Sample $\theta^{(j)}|F^{(j)}, r^{(j)}, \Omega^{(j)}, r$ from $N(\tilde{\theta}_{\delta}, V)$

A. $\tilde{\theta}_{(j)} = \left[\Omega^{(j)} \otimes X_{\gamma^{(j)}} \left(\left(M_{F^{(j)}} + \frac{1}{c} I_T\right) X_{\gamma^{(j)}}\right)^{-1} \Omega^{(j)} \otimes X_{\gamma^{(j)}} M_{F^{(j)}}\right] r$
$V = \left[\Omega^{(j)} \otimes X_{\gamma^{(j)}} \left(M_{F^{(j)}} + \frac{1}{c} I_T\right) X_{\gamma^{(j)}}\right]$
$M_{F^{(j-1)}} = I_T - F^{(j)} \left(F^{(j)'} F^{(j)} + \frac{1}{c} I_{(j)}\right)^{-1} F^{(j)'}$
$l = l + 1$
end FOR 1 : K

(b) $j = j + 1$ Back to step (2)

Output of the RJMCMC:
$\{r^{(1)}, ..r^{(j)}, ..r^{(M)}\}$
$\{\Omega^{(1)}, ..\Omega^{(j)}, ..\Omega^{(M)}\}; \{\lambda^{(1)}, ..\lambda^{(j)}, ..\lambda^{(M)}\}; \{\theta^{(1)}, ..\theta^{(j)}, ..\theta^{(M)}\}$

6.4 GRAPHICS

Figure 2: Risk Premia Iterates and Histogram for the combined set of measured factors for the sample period 1960 – 2000. The g-prior used is $c(X’X)^{-1}$. $N = 43, T = 480, K_{max} = 19$. $\gamma|\mathbf{r} = \{\gamma_{cons}, \gamma_{SMB}, \gamma_{HML}\}$
Figure 3: Empirical pdf and cdf for Risk Premia Iterates for the economic variables. Case of combined set of measured factors with g-priors $c(X'X)^{-1}$, $c = T = 480$, $N = 43$ and $K_{\text{max}} = 19$. Sample period 1960 – 2000
Figure 4: Autocorrelation Function for Risk Premia Iterates and the pricing error. Case of combined set of measured factors with g-priors $c(X'X)^{-1}$, $c = T = 480$, $N = 43$ and $K_{\text{max}} = 19$, lags = 300. Sample period 1960-2000.