JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 42, No. 4, Dec. 2007, pp. 857–892

Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions

Federico Nardari and John T. Scruggs*

Abstract

We analyze a new class of linear factor models in which the factors are latent and the covariance matrix of excess returns follows a multivariate stochastic volatility process. We evaluate cross-sectional restrictions suggested by the arbitrage pricing theory (APT), compare competing stochastic volatility specifications for the covariance matrix, and test for the number of factors. We also examine whether return predictability can be attributed to time-varying factor risk premia. Analysis of these models is feasible due to recent advances in Bayesian Markov chain Monte Carlo (MCMC) methods. We find that three latent factors with multivariate stochastic volatility best explain excess returns for a sample of 10 size decile portfolios. The data strongly favor models constrained by APT pricing restrictions over otherwise identical unconstrained models.

I. Introduction

We analyze a new class of linear factor models in which the factors are latent and the covariance matrix of returns follows a multivariate stochastic volatility (MSV) process. Linear factor models play a prominent role in theoretical and empirical asset pricing. They have been a parsimonious means of describing the covariance matrix of returns since the single-index model of Sharpe (1963). The arbitrage pricing theory (APT) of Ross (1976), (1977) assumes that asset returns follow a static linear factor model in which factors are latent. In the spirit of the APT, we assume that factors are latent and organic to the return generating process. Rather than prespecifying factors, we extract the latent factors from the test assets themselves.[1] We extend the traditional static linear factor model to

*Nardari, federico.nardari@asu.edu, Arizona State University, W. P. Carey School of Business, Department of Finance, PO Box 873906, Tempe, AZ 85287; Scruggs, jscruggs@terry.uga.edu, University of Georgia, Terry College of Business, Department of Banking and Finance, Brooks Hall, Athens, GA 30602. The paper has benefited from comments by Stephen Brown (the editor) and Martijn Cremers (the referee). We thank Sid Chib, Alex David, Phil Dybvig, Pedro Santa-Clara, Tim Simin, Guofu Zhou, and seminar participants at UCLA, the University of Houston, Washington University in St. Louis, the 2002 All-Georgia Finance Forum, the 2003 Western Finance Association meetings, and the 2003 European Finance Association meetings for useful comments. Sriram Villupuram provided excellent research assistance. Scruggs acknowledges financial support from a Terry-Sanford research award. The usual disclaimer applies.

[1]Our approach is similar to Geweke and Zhou (1996) who consider homoskedastic models.

allow for two important features of the data: i) a time-varying covariance matrix of returns in which factor and idiosyncratic shocks exhibit volatility shocks and persistence (i.e., clustering), and ii) time-varying expected returns.

Our interest in this class of models is twofold. First, we are interested in the features of the model that provide the best fit for the data. To that end, we compare models with different specifications for the covariance matrix of returns and different numbers of factors. Motivated by the empirical evidence on predictable returns, we also consider models in which expected returns vary over time. Second, we are interested in evaluating the linear pricing restriction implied by the APT. We evaluate this implication by comparing unconstrained models to models constrained by APT pricing restrictions.

We examine excess returns on 10 NYSE/AMEX/NASDAQ market capitalization (i.e., size) decile portfolios for the period from January 1952 to December 2003. We report a number of new and interesting empirical results. MSV is clearly an important feature of the data. Models in which both factor and idiosyncratic shocks follow stochastic volatility processes are strongly favored over models in which factor shocks and/or idiosyncratic shocks are homoskedastic. We find that three latent factors best explain the variation in market capitalization decile portfolio excess returns. We interpret these latent factors as: i) a stock market factor, ii) a size factor, and iii) a small stock factor. We find that all three of these latent factors are "priced" in the sense that exposure to them is rewarded with constant risk premia. We find that the data strongly support models constrained by APT pricing restrictions over otherwise identical unconstrained models. In other words, models in which expected returns are linear in exposures to latent factors are favored over models in which expected returns are independent of factor loadings. This is the central empirical prediction of the APT. None of our conclusions are altered when expected returns are allowed to vary over time. We find that models in which predictability is related to exposures to latent factors are favored over models in which predictability is unconstrained. Using an extensive battery of sensitivity checks, we demonstrate that our conclusions are remarkably robust to changes in prior specification.

We adopt a Bayesian approach and exploit recent advances in Markov chain Monte Carlo (MCMC) econometrics to estimate and compare models. The Bayesian MCMC approach is attractive for a number of reasons. First, MCMC methods provide exact finite-sample posterior densities for model parameters and for functions of interest. Popular classical methods for testing asset pricing models (GMM, Fama-MacBeth, etc.) rely on asymptotic distributions to make inferences. Second, MCMC techniques are especially well suited for estimating models with latent variables. In our model, both the factor shocks and the stochastic volatilities are latent. In order to estimate the model, one must employ a technique to integrate the latent stochastic volatilities out of the likelihood function. Recent advances in Bayesian statistical methods make estimation of high dimensional MSV models feasible.[2] Our MCMC algorithm samples the latent factor shocks and latent stochastic volatilities conditional on the data and the other parameters of the model. This enables us to effectively integrate over these "nuisance param-

[2]Examples include Jacquier, Polson, and Rossi (1994), Kim, Shephard and Chib (1998), and Chib, Nardari, and Shephard (2002), (2006).

eters" and obtain exact marginal posterior densities for the parameters of interest. Third, the MCMC approach enables the estimation of latent factor shocks, factor loadings, risk premia, stochastic volatilities, and other parameters in a single stage. This alleviates the errors-in-variables problem inherent in the two-pass procedures typically employed to test the APT.[3]

At the heart of our analysis are model comparisons based on Bayes factors. We compare models with up to five latent factors, with four different covariance matrix specifications, with and without time-varying expected returns, and with and without APT pricing restrictions. A Bayes factor summarizes the support provided by the data in favor of one model relative to an alternative model.[4] Since Bayes factors are constructed from marginal likelihoods, they measure the model's ability to explain the entire distribution (i.e., not just first moments) of returns.[5] Bayes factors also embed an implicit penalty for model complexity. The ubiquitous likelihood ratio test with fixed significance level tends to favor less parsimonious models as the sample size grows (see Kass and Raftery (1995), O'Hagan (1994)).[6] The risk of overfitting the data is greatly reduced in the Bayesian framework. Bayes factors also permit the simultaneous comparison of multiple models, regardless of whether the models are nested. Several non-Bayesian model selection criteria need to be modified in nontrivial ways to accommodate multiple and/or non-nested models (see Berger and Pericchi (1996)). Several previous papers use Bayes factors or posterior-odds ratios to evaluate asset pricing models. Shanken (1987) and Harvey and Zhou (1990) use posterior-odds ratios to evaluate portfolio efficiency. McCulloch and Rossi (1991) rely on posterior-odds ratios to test APT pricing restrictions for a set of size portfolios and Connor-Korajczyk factors. More recently, Avramov and Chao (2006) use Bayes factors to compare competing, non-nested asset pricing models. Ours is the first paper to compare linear factor models with latent factors and stochastic volatility for factor/idiosyncratic shocks.

The remainder of this paper is organized as follows. Section II describes the model. Our Bayesian MCMC estimation methodology is discussed in Section III. Technical details are provided in the appendices. We describe the dataset in Section IV. We employ Bayes factors to compare models in Section V. Section

[3]Papers that employ two-pass procedures to test the APT include Roll and Ross (1980), Connor and Korajczyk (1986), (1988), Lehmann and Modest (1988), and Jones (2001). An alternative to the two-pass approach is the systems estimation approach of Brown and Weinstein (1983), Burmeister and McElroy (1988), McElroy and Burmeister (1988), and Brown and Otsuki (1993). These papers employ a restricted nonlinear multivariate regression framework to test APT pricing restrictions on the linear factor model. The APT-constrained model can be estimated simultaneously using the iterated nonlinear seemingly unrelated regression (ITNLSUR) procedure (see, for example, McElroy and Burmeister (1988)).

[4]Bayes factors are related to posterior odds. In non-technical terms, the posterior-odds ratio is equal to the product of the prior-odds ratio and the Bayes factor. If the prior odds are 1:1 (i.e., each model is equally likely a priori), then the Bayes factor is equivalent to the posterior-odds ratio.

[5]Marginal likelihoods are computed by taking the expectation of the likelihood function (or sampling density) with respect to the joint prior density of the parameters. For a detailed discussion, see Chib (1995) or Kass and Raftery (1995).

[6]The Akaike information criterion and its corrections may help overcome some of the drawbacks of the likelihood ratio test. The Schwartz information criterion can be viewed as a rough, asymptotic approximation to the Bayes factor. The performance of these information criteria in the context of the complex dynamic structures we consider is not known.

VI discusses the empirical results for models with constant risk premia. We examine models with time-varying expected returns in Section VII. Section VIII summarizes our conclusions.

II. The Model

Our empirical model consists of i) a linear factor model for excess returns, ii) a cross-sectional pricing restriction motivated by the equilibrium APT, iii) an MSV model for factor shocks and idiosyncratic shocks, and iv) a linear instrumental variables model for time-varying expected returns or risk premia.

A. A Linear Factor Model for Excess Returns

Let y_t denote the N-vector of excess asset returns and f_t denote the K-vector of latent (or unobservable) factor shocks in period t. In our linear factor model, the realized excess return on an asset is the sum of its expected return, K systematic shocks, and an idiosyncratic shock. In matrix notation, the linear factor model for the excess return vector is

$$y_t = \mu + Bf_t + \epsilon_t, \tag{1}$$

where μ is an N-vector of asset expected returns and B is a $N \times K$ matrix of factor loadings (i.e., risk exposures or betas). We assume $E[f_t] = 0$, $E[\epsilon_t] = 0$, and $E[f_t \epsilon_t'] = 0$. We further assume that the factor loading matrix B is time invariant.[7] We discuss cross-sectional restrictions imposed on μ by the APT in the next subsection.

Our model differs from the familiar static factor model in two significant ways. First, we assume that the factors are latent (unobservable). Second, we assume that the covariance matrix of returns, Ω_t, is time varying. Since factor shocks are uncorrelated with idiosyncratic shocks by definition, Ω_t can be decomposed into systematic and idiosyncratic components:

$$\Omega_t = BD_tB' + V_t, \tag{2}$$

where D_t is a diagonal $K \times K$ covariance matrix of factor shocks and V_t is a diagonal $N \times N$ covariance matrix of idiosyncratic shocks. The model for D_t and V_t is described in detail in subsection C below.

B. APT Pricing Restriction

The linear factor model described in equation (1), which we refer to as the unconstrained model, is a statistical model of asset returns. Asset pricing models impose cross-sectional restrictions on equation (1). The APT predicts that

[7]Engle, Ng, and Rothschild (1990) provide theoretical justification for the constant beta assumption in the context of the consumption CAPM. Connor and Korajczyk (1989) discuss the strong assumptions on preferences and distributions required to deliver an intertemporal variant of the APT with constant betas. Aguilar and West (2000) also assume a time-invariant factor loading matrix.

expected returns are linear functions of exposures to systematic (i.e., undiversifiable) risk factors. We follow the competitive equilibrium version of the APT (e.g., Connor (1984)) and impose the following exact pricing relation:

$$\mu = B\gamma, \tag{3}$$

where B is the previously defined $N \times K$ matrix of factor loadings and γ is the K-vector of time-invariant factor risk premia. Substituting (3) into (1) delivers the constrained linear factor model:

$$\begin{aligned} y_t &= B\gamma + Bf_t + \epsilon_t \\ &= B(\gamma + f_t) + \epsilon_t. \end{aligned} \tag{4}$$

We interpret γ as the vector of factor risk premia where the factor shocks are mean zero (i.e., $E[f_t] = 0$). Alternately, we could interpret γ as a vector of ex ante factor means (i.e., $E[f_t] = \gamma$). These two interpretations are statistically indistinguishable. It is important to note that, conditional on contemporaneous portfolio returns, factor shocks are not mean zero (i.e., $E[f_t|y_t] \neq 0$). It follows that ex post average factor returns may not equal ex ante factor risk premia for a given finite sample. Likewise, portfolio average returns may not equal portfolio expected returns for a given finite sample.

Note that equation (4) does not include a separate intercept term. Many papers in the empirical asset pricing literature (e.g., Gibbons, Ross, and Shanken (1989)) test the hypothesis that the intercepts (i.e., pricing errors or alphas) are zero on average. This approach is not available when factors are latent and the model is estimated in a single stage. This is because intercept terms, factor shocks, and factor risk premia are not separately identified. We evaluate the hypothesis $\mu = B\gamma$ by using Bayes factors to compare unconstrained models to models constrained by APT pricing restrictions. Bayes factors are discussed in Section III.

Although the cross-sectional restrictions we impose are motivated by the APT, we hesitate to characterize the constrained model as a dynamic variant of the APT.[8] It would be difficult to justify constant risk premia in a dynamic version of the APT. However, the constant risk premia assumption is a reasonable starting point for our analysis.[9] We consider an extension of the model with time-varying expected returns in subsection E below.

C. Multivariate Stochastic Volatility

An innovative feature of our model is the specification of the time-varying covariance matrix. Previous papers that examine linear factor models with time-varying covariance matrices typically assume some form of multivariate GARCH model.[10] We assume that both factor shocks and idiosyncratic shocks follow in-

[8] Dynamic versions of the APT are derived by, among others, Bossaerts and Green (1989) and Connor and Korajczyk (1989). See Connor and Korajczyk (1995) for a review of these models.

[9] It should be noted that the prespecified factors or Connor-Korajczyk mimicking portfolios examined in other papers have heteroskedastic returns. Rarely do these papers model time-varying factor risk premia. Ferson and Korajczyk (1995) is a rare exception.

[10] Papers that examine multivariate GARCH models include Engle, Ng, and Rothschild (1990), Ng, Engle, and Rothschild (1992), and King, Sentana, and Wadhwani (1994). These papers all resort to multiple-stage estimation procedures due to model complexity.

dependent stochastic volatility processes. Stochastic volatility models are an attractive alternative to models from the GARCH family because they allow for exogenous volatility shocks.[11] Since D_t and V_t are diagonal, any covariance between assets is due to the factor loading matrix. Our MSV model is based on Chib, Nardari, and Shephard (2006).

We assume that the factor shocks and idiosyncratic shocks in (1) and (4) follow a multivariate Gaussian distribution,

$$\begin{pmatrix} \epsilon_t \\ f_t \end{pmatrix} \sim \mathcal{N}_{n+k}\left\{0, \begin{pmatrix} V_t & 0 \\ 0 & D_t \end{pmatrix}\right\}, \tag{5}$$

where

$$\begin{aligned} V_t &= V_t(h_t) &&= \operatorname{diag}\{\exp(h_{1,t}),\ldots,\exp(h_{N,t})\}, \\ D_t &= D_t(h_t) &&= \operatorname{diag}\{\exp(h_{N+1,t}),\ldots,\exp(h_{N+K,t})\}, \end{aligned}$$

and h_t denotes the $N+K$-vector of time-varying latent log-variances. We further assume that the log-variances, $h_t = (h_{1,t},\ldots,h_{N+K,t})$, each follow an independent stochastic volatility process:

$$h_{j,t} = \kappa_j + \phi_j(h_{j,t-1} - \kappa_j) + \sigma_j \eta_{j,t}, \tag{6}$$

where $\eta_{j,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $E[\eta_t \eta_t'] = I_{N+K}$.

We refer to this covariance specification as the MSV model. In addition to the MSV model, we estimate three constrained variants. Our goal is to determine whether intertemporal variation in the covariance matrix of returns is driven by time-series heteroskedasticity in factor shocks, in idiosyncratic shocks, or in both. In the stochastic volatility-in-errors (SVE) model, idiosyncratic shocks follow independent SV processes while factor shocks are homoskedastic ($D_t = D$). In the SVE model, each asset's idiosyncratic shock has its own SV process. In the second constrained model, the stochastic volatility-in-factors (SVF) model, factor shocks follow independent SV processes while idiosyncratic shocks are homoskedastic ($V_t = V$). In the final model, the classic factor (CF) model, both factor and idiosyncratic shocks are homoskedastic. The CF model is comparable to the model in Geweke and Zhou (1996).

D. Identification Issues

We must impose further structure on our model to deal with two identification issues. The first issue arises when we move from a static factor model to a dynamic model. In the most general dynamic model, both the covariance matrix of the factors and the factor loading matrix could be time varying (e.g., one could generalize equation (2) such that $\Omega_t = B_t D_t B_t' + V_t$). In practice, econometricians typically choose to either estimate a model with dynamic factor loadings and static (i.e., homoskedastic) factors, or estimate a model with static factor loadings and dynamic (i.e., heteroskedastic) factors. We adopt the latter approach and assume that B is constant. Since our empirical work examines the returns

[11]Taylor (1994) and Ghysels, Harvey, and Renault (1996) review SV models.

on characteristic-sorted portfolios (e.g., market capitalization decile equity portfolios), we do not believe this assumption is too onerous.[12] Ferson and Harvey (1991) and Evans (1994) attribute most of the predictable variation in expected returns to time-varying risk premia rather than time-varying risk exposures. Furthermore, Ghysels (1998) finds that models with time-varying factor loadings can lead to larger pricing errors than models with constant factor loadings.

The second identification issue is the familiar rotational indeterminancy problem. Since the K factors (f_t) are latent, we must impose $(K^2+K)/2$ restrictions on the $N \times K$ factor loading matrix B to fix the rotation. Let b_{ij} denote an element of the matrix B where i is the row index and j is the column index. We impose the constraints $b_{ij}=1$ for $i=j$ and $b_{ij}=0$ for $i<j$. Aguilar and West (2000) refer to this as a hierarchical structural constraint.[13] This leaves $NK-(K^2+K)/2$ free parameters to be estimated in B. The resulting loading matrix is:

$$B = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_{N1} & b_{N2} & \cdots & b_{NK} \end{pmatrix}. \tag{7}$$

E. Time-Varying Expected Returns

We also investigate whether linear factor models with latent factors, MSV, and APT pricing restrictions can explain predictable variation in stock returns. Many interpret stock return predictability as evidence of time-varying expected returns. If this is the case, then the predictability of asset returns should be proportional to systematic risk exposures. We address this issue by comparing models with unconstrained return predictability to models in which return predictability is due to time-varying latent factor risk premia.

Our specification for time-varying expected returns is simple, but widely employed in empirical tests of conditional asset pricing models. We assume that the vectors of portfolio expected returns (μ_t) or factor risk premia (γ_t) are linear functions of predetermined state variables.[14] Let z_{t-1} be a vector of instrumental variables (with a one in the first row) known at the beginning of period t. For unconstrained models, we assume $\mu_t = Mz_{t-1}$ where M is a conforming matrix of coefficients. Substituting μ_t into (1) provides the unconstrained linear factor model with time-varying expected returns:

$$y_t = \mu_t + Bf_t + \epsilon_t = Mz_{t-1} + Bf_t + \epsilon_t. \tag{8}$$

If predictable stock returns are due solely to time-varying factor risk premia, then predictability should be proportional to factor loadings.[15] We investigate this

[12]Other papers that assume constant factor loadings (or betas) in a dynamic model include Campbell (1987), Harvey (1989), Kirby (1998), Avramov (2004), and Avramov and Chao (2006).

[13]Geweke and Zhou (1996), who assume that both factor shocks and idiosyncratic shocks are homoskedastic, constrain $b_{ij}>0$ for $i=j$ and assume that $E[f_t f_t'] = I_K$.

[14]Papers employing related models include Campbell (1987), Harvey (1989), Shanken (1990), Ferson and Harvey (1991), Ferson and Korajczyk (1995), Kirby (1998), Avramov (2004), and Avramov and Chao (2006).

[15]Gibbons and Ferson (1985) is an early example of this type of test.

by allowing factor risk premia to be linear functions of predetermined information variables. For the constrained model, we assume $\gamma_t = Gz_{t-1}$ where G is a conforming matrix of coefficients. We modify the exact pricing relation in (3) to allow for time-varying factor risk premia:

$$\mu_t = B\gamma_t, \tag{9}$$

where B is the previously defined $N \times K$ matrix of factor loadings and γ_t is the K-vector of time-varying factor risk premia. Substituting (9) into (8) delivers the constrained linear factor model with time-varying risk premia:

$$\begin{aligned} y_t &= B\gamma_t + Bf_t + \epsilon_t \\ &= B(\gamma_t + f_t) + \epsilon_t \\ &= B(Gz_{t-1} + f_t) + \epsilon_t. \end{aligned} \tag{10}$$

We interpret γ_t as the vector of conditional factor risk premia where the factors are conditionally mean zero (i.e., $E[f_t|z_{t-1}]=0$). Alternately, we could interpret γ_t as a vector of conditional factor means (i.e., $E[f_t|z_{t-1}] = \gamma_t$). These two interpretations are statistically equivalent.

For the constrained model in (10), all return predictability is explained by exposures to latent factors. In contrast, predictability is independent of loadings on latent factors for the unconstrained model in (8). We employ Bayes factors to determine which of these non-nested, polar views is supported by the data.[16]

Models with time-varying expected returns introduce new degrees of freedom and thus require further restrictions to ensure identification. The parameters in the first K rows of the factor loading matrix B (see equation (7)) must be fixed. Many latent factor models (e.g., Zhou (1994)) assume that the first K rows of the B matrix are a $K \times K$ identity matrix. This approach ensures identification, but implicitly assumes no covariance between the first K assets under consideration. We take a different approach. We fix the off-diagonal elements in the first K rows of B, but we do not assume that they are all zero.[17] This approach ensures identification, but allows for nonzero covariance between the first K assets (i.e., the reference assets).

III. Empirical Methodology

A. Bayesian Estimation

We employ Bayesian MCMC methods to estimate the models described in the previous section in a single stage. Bayesian analysis requires three elements: the data, a likelihood function (i.e., sampling distribution) dictated by the model specification, and prior beliefs about the parameters. For notational convenience, let β denote the free elements of the loading matrix B and $\theta_j = (\kappa_j, \phi_j, \sigma_j)$

[16]In addition to these polar views, Avramov (2004) considers intermediate views. Avramov (2004) examines the portfolio allocation decisions of investors with varying prior beliefs regarding asset pricing models and return predictability. He finds that when prior beliefs deviate from a dogmatic faith in an asset pricing model, optimal portfolio allocations can change dramatically.

[17]In our empirical work, we fix the off-diagonal elements in the first K rows of B at or near the posterior estimates of the otherwise identical model with constant expected returns.

https://doi.org/10.1017/S0022109000003422 Published online by Cambridge University Press

denote the vector of parameters for the jth stochastic volatility process. Let $\psi = (\gamma, \beta, \theta_1, \ldots, \theta_{N+K})$ denote the complete parameter vector for the constrained linear factor model with constant risk premia. Let y_t denote the N-vector of asset excess returns in period t and $y = (y_1, \ldots, y_T)$ denote the full dataset. Following Bayes rule, the joint posterior density of the parameters is proportional to the product of the likelihood function and the prior density on the parameters:

$$\pi(\psi|y) \quad \propto \quad p(y|\psi)\pi(\psi).$$

Bayesian inference is accomplished by examining the joint posterior density of the parameters of interest.

The likelihood function for the models described in the previous section is complicated because the factors and log-variances are latent. Let $\mathcal{F}_{t-1}$ denote the history of the $\{y_t\}$ process up to time $t-1$, and let $p(h_t|\mathcal{F}_{t-1}, \psi)$ denote the density of the latent log-variances (h_t) conditioned on $(\mathcal{F}_{t-1}, \psi)$. The likelihood function for the model in (4), (5), and (6) is given by

$$\begin{aligned} p(y|\psi) &= \prod_{t=1}^{T} \int p(y_t|h_t, \gamma, B)p(h_t|\mathcal{F}_{t-1}, \psi)dh_t \\ &= \prod_{t=1}^{T} \int \phi_N(y_t|B\gamma, \Omega_t)p(h_t|\mathcal{F}_{t-1}, \psi)dh_t, \end{aligned} \tag{11}$$

where the notation $\phi_N(.|\mu, \Sigma)$ is used to indicate an N-variate Gaussian density with mean vector μ and covariance matrix Σ, and where $\Omega_t = \Omega_t(h_t) = BD_t(h_t)B' + V_t(h_t)$ is the time-varying covariance matrix of asset excess returns.

Estimation of this model by maximum likelihood methods is impractical because the high dimensional integral in (11) is analytically intractable. Our MCMC approach does not attempt to maximize the likelihood function in (11). Rather, we construct a Gibbs sampler with a limiting distribution equal to the joint posterior density $\pi(\psi|y)$. Inference is based on summary statistics (e.g., mean, standard deviation, critical values) describing the sample draws of parameters of interest and test statistics constructed from the sample draws. The additional complexity in the present case is that the joint posterior density is also very high dimensional. We essentially augment the parameter space ψ with the latent factors and the latent log-variances. One can think of the latent variables as nuisance parameters that are "integrated out" by the Gibbs sampler. We analyze returns for 10 portfolios and up to five factors. The Gibbs sampler for the unconstrained five-factor MSV model, the most parameter-rich specification with constant expected returns, yields 90 parameters, five time series of latent factor shocks (624 observations each), and 15 time series of latent log-variances (also 624 observations each). The strategy for sampling from this joint posterior density has to be carefully designed. We follow the approach suggested by Chib, Nardari, and Shephard (2006); the essential steps are summarized in Appendix A.

B. Priors

It is possible to model prior beliefs on the parameters by adopting virtually any reasonable distributional form. The choice is, by nature, subjective. Our goal is to minimize the impact of the priors on the posterior estimates and, especially, on the model comparisons. Accordingly, we employ relatively uninformative (i.e., imprecise) priors. As a result, the posterior densities of the parameters are driven primarily by the sample data. We employ only proper prior densities (i.e., densities that integrate to one over the parameter space). Furthermore, we assume that the distributions of the factor loadings, risk premia, and SV parameters are a priori independent. For example, the joint prior density for the parameters of the constrained MSV model can be factored

$$\pi(\psi) \quad = \quad \pi(\gamma)\pi(\beta)\pi(\kappa)\pi(\phi)\pi(\sigma),$$

where $\pi(\gamma) = \pi(\gamma_1)\cdots\pi(\gamma_K)$, $\pi(\kappa) = \pi(\kappa_1)\cdots\pi(\kappa_{N+K})$, etc.

In this subsection, we briefly describe our uninformative or "base" priors. Our prior for each of the free elements in B is normal, $b_{ij} \sim \mathcal{N}(b_{ij}^0, B_{ij}^0)$. For the base prior, the mean and variance hyperparameters are $b_{ij}^0 = 0$ and $B_{ij}^0 = 9$, respectively. For APT-constrained models with constant risk premia, our priors on the factor risk premia are also normal, $\gamma_j \sim \mathcal{N}(\gamma_j^0, G_j^0)$, where $\gamma_j^0 = 0$ and $G_j^0 = 1$. Similarly, for unconstrained models we assume $\mu_i \sim \mathcal{N}(\mu_i^0, M_i^0)$ and set $\mu_i^0 = 0$ and $M_i^0 = 1$. Since we estimate the models with portfolio returns in decimal form, these distributions represent extremely diffuse prior beliefs. For models with time-varying expected returns, we need to specify priors for the G or M matrices. Our priors on the elements of G and M are normal with mean zero and a variance of 10.

For κ_j, the average log-variance for both factor and idiosyncratic shocks, our prior is $\kappa_j \sim \mathcal{N}(\kappa_j^0, K_j^0)$, where $\kappa_j^0 = -8$ and $K_j^0 = 25$. At the prior mean, this implies an average volatility of about 1.83% per month. Again, this prior is very uninformative. Within one standard deviation of the mean, the prior allows average volatilities to vary between 0.15% and 22% per month. For our prior on ϕ_j, we follow Kim, Shephard, and Chib (1998) and Chib, Nardari, and Shephard (2002). ϕ_j is constrained to the interval $[-1, 1]$ (i.e., the region of stationarity) by implementing a change of variable, $\phi_j = 2\phi_j^* - 1$, and assuming

$$\phi_j^* \quad \sim \quad \text{Beta}\left(\phi_j^{(1)}, \phi_j^{(2)}\right).$$

The hyperparameters $\phi_j^{(1)} = 20$ and $\phi_j^{(2)} = 1.5$ imply that ϕ_j has a mean of 0.86 with standard deviation of 0.11. This moderately informative prior is consistent with well-established empirical evidence of volatility persistence for stock returns.

The prior for σ_j (i.e., the volatility of the log-variance) is inverse gamma, $\sigma_j \sim \mathcal{IG}(\varsigma_j^0, \xi_j^0)$. For the base prior, the hyperparameters $\varsigma_j^0 = 2.39$ and $\xi_j^0 = 0.347$ imply that σ_j is distributed with a mean of 0.25 with standard deviation of 0.4. For SVF, SVE, and CF models, the factor shocks and/or the idiosyncratic shocks are homoskedastic. For these models, our priors on the diagonal elements of the D and/or V matrices are inverse gamma. We choose the hyperparameters ς_j^0 and ξ_j^0 so that the prior densities for D and V roughly match the steady-state values for D_t and V_t implied by our priors on κ_j. This choice of prior aims to "level the playing field" when comparing homoskedastic and heteroskedastic models.

https://doi.org/10.1017/S0022109000003422 Published online by Cambridge University Press

C. Bayesian Model Comparison

We compare model specifications on a number of dimensions. Model comparisons of constrained ($\mu = B\gamma$) versus unconstrained (μ unrestricted) models permit us to evaluate cross-sectional restrictions suggested by the APT. We evaluate the relative merits of the MSV, SVF, SVE, and CF specifications. And we compare models with different numbers of factors.

We employ Bayes factors to compare models.[18] In Bayesian analysis, Bayes factors provide a unified way to compare the relative support that the data provide for competing model specifications. Unlike classical test statistics, such as the likelihood ratio statistic, Bayes factors can compare non-nested models. Bayes factors also have the appealing property of implicitly penalizing models for additional parameters.[19]

The Bayes factor comparing model $\mathcal{M}_i$ to model $\mathcal{M}_j$ is defined as

$$BF_{ij} = \frac{m(y|\mathcal{M}_i)}{m(y|\mathcal{M}_j)}, \tag{12}$$

where $m(y|\mathcal{M}_i)$ is the marginal likelihood of the data (y) given $\mathcal{M}_i$. The marginal likelihood is obtained by integrating the likelihood function with respect to the prior density of the parameters. Since Bayes factors are constructed from marginal likelihoods, they compare the relative abilities of competing models to explain the distribution of asset returns. The computation of marginal likelihoods is discussed in more detail in Appendix B.

In our empirical work, we report log Bayes factors ($\log_{10}(BF_{ij})$). Note that $\log_{10}(BF_{ij}) = \log_{10}(m(y|\mathcal{M}_i)) - \log_{10}(m(y|\mathcal{M}_j))$. Following Kass and Raftery (1995), we evaluate the significance of a Bayes factor using a base 10 logarithmic scale:

$\log_{10}(BF_{ij})$	BF_{ij}	Evidence against $\mathcal{M}_j$
0 to ½	1 to 3.16	Not worth more than a bare mention
½ to 1	3.16 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Bayes factors are closely related to posterior-odds ratios. The posterior-odds ratio comparing $\mathcal{M}_i$ to $\mathcal{M}_j$ is defined

$$K_{ij} = \frac{\pi(\mathcal{M}_i)}{\pi(\mathcal{M}_j)} \cdot \frac{m(y|\mathcal{M}_i)}{m(y|\mathcal{M}_j)},$$

where $\pi(\mathcal{M}_i)$ is the prior probability on $\mathcal{M}_i$, and $\pi(\mathcal{M}_i)/\pi(\mathcal{M}_j)$ is the prior-odds ratio. When the prior-odds ratio is 1:1 (i.e., equal prior probabilities), the Bayes factor is equivalent to the posterior-odds ratio. In our empirical work, we interpret Bayes factors under the assumption that all models are equally likely a

[18]See Kass and Raftery (1995) for an excellent review of Bayes factors.

[19]See O'Hagan (1994) for a discussion.

priori. This is, by far, the most common approach (e.g., Shanken (1987), Harvey and Zhou (1990), Connolly (1991), and Avramov and Chao (2006)).[20] If the reader has different prior odds for a particular pair of models, it is simple to compute the corresponding posterior odds for a given Bayes factor. McCulloch and Rossi (1991) employ posterior-odds ratios to test APT pricing restrictions between nested models.

Bayes factors are also closely related to posterior probabilities on models. Say we want to compare L models simultaneously, and let $\pi(\mathcal{M}_l)$ denote the prior probability of $\mathcal{M}_l$. The posterior probability for $\mathcal{M}_i$ is

$$\pi(\mathcal{M}_i|y) = \frac{\pi(\mathcal{M}_i)m(y|\mathcal{M}_i)}{\sum_{l=1}^{L}\pi(\mathcal{M}_l)m(y|\mathcal{M}_l)}. \tag{13}$$

For example, imagine that we are comparing two models, $\mathcal{M}_i$ and $\mathcal{M}_j$, with equal prior probabilities (i.e., $\pi(\mathcal{M}_i) = \pi(\mathcal{M}_j) = 0.5$). If $\log_{10}(BF_{ij}) = 2$, then $\pi(\mathcal{M}_i|y) = 100/101 \approx 0.99$. Posterior probabilities are a convenient way to summarize model comparison results when the number of models is greater than two. Avramov and Chao (2006) use posterior probabilities in this manner.

It is well known that Bayes factors may be sensitive to the choice of priors on parameters (see, for example, Klein and Brown (1984), McCulloch and Rossi (1991), and Kass and Raftery (1995)). This sensitivity is a potential problem even if samples are large and priors are uninformative. In the case of improper diffuse priors (i.e., an extreme form of uninformative priors), Bayes factors are actually undefined. We avoid this issue by employing only proper priors. Ideally, priors should be "neutral" in the sense that they do not unduly bias the Bayes factor in favor of a particular model or class of models.[21] However, since non-nested models may have very different parameters, it is virtually impossible to specify "neutral" priors. We address the issue by carefully checking the sensitivity of Bayes factors to prior specification. We perform two types of sensitivity analysis. First, we replace the uninformed base priors with priors "informed" by analysis of pre-sample data (i.e., training samples). We employ training samples of different lengths to vary the informativeness of the priors. Second, we analyze the effects of perturbing the uninformative priors by parameter type. We reports the results of this sensitivity analysis in Section V.

IV. Data

A. Size Decile Portfolio Returns

We examine monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization decile portfolios for the sample period 1952:1 to 2003:12 (624 months). All returns are simple returns in excess of the one-month U.S. Treasury bill yield. All data are provided by the Center for Research in Security

[20]Klein and Brown (1984) explicitly choose prior model probabilities that minimize appropriate measures of prior information. The paper derives conditions on the prior precision matrices under which equating prior probabilities leads to well-behaved posterior odds. Unfortunately, the Klein-Brown framework cannot reasonably be extended to the classes of models (i.e., multivariate, latent factor, stochastic volatility) considered in the present study.

[21]Kass and Raftery (1995) make a similar point.

Prices (CRSP) at the University of Chicago. Many previous tests of the APT (e.g., Connor and Korajczyk (1988), McCulloch and Rossi (1991), Ferson and Korajczyk (1995), and Geweke and Zhou (1996)) have employed market capitalization decile portfolios as test assets. Extracting latent factors from size-sorted portfolio returns is further motivated by Moskowitz (2003), who finds that "a size factor has significant explanatory power for contemporaneous and future return second moments" (p. 419).[22]

The descriptive statistics for portfolio excess returns reported in Table 1 exhibit some well-known patterns. There is a strong inverse relation between mean excess returns and market capitalization. Mean excess returns range from 53.3 basis points (bp) per month for the largest size decile portfolio (*Cap10*) to 1.321% per month for the smallest size decile portfolio (*Cap1*). The relation between standard deviation and size mirrors the inverse relation found for the means. Standard deviations range from 4.241% per month for *Cap10* to 7.480% per month for *Cap1*. Principal components analysis of the sample correlation matrix suggests that one or two factors appear to be driving the covariance matrix of portfolio returns. The first five eigenvalues of the sample correlation matrix are 8.91, 0.80, 0.14, 0.04, and 0.03. These five eigenvalues explain (cumulatively) 89.1%, 97.1%, 98.6%, 99.0%, and 99.3% of the common variation in portfolio excess returns.

TABLE 1

Descriptive Statistics for Portfolio Return Data

Table 1 reports descriptive statistics for monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). All returns are simple returns (in percent) in excess of the one-month Treasury bill yield.

Portfolio	Mean (%)	Std. Dev.	Skewness	Excess Kurtosis
Cap1	1.321	7.480	1.241	7.224
Cap2	1.016	6.505	0.423	3.786
Cap3	0.845	6.154	0.072	3.628
Cap4	0.821	5.915	0.056	3.893
Cap5	0.760	5.748	−0.260	3.607
Cap6	0.751	5.642	−0.304	3.065
Cap7	0.715	5.476	−0.468	3.309
Cap8	0.705	5.176	−0.559	3.018
Cap9	0.681	4.878	−0.563	3.028
Cap10	0.533	4.241	−0.390	1.694

B. Instrumental Variables for Models with Time-Varying Expected Returns

Our choice of instrumental variables for time-varying expected returns is motivated by previous empirical studies of conditional asset pricing models.[23] We

[22]Moskowitz (2003) also reports that the market portfolio is the most important factor for explaining the covariance matrix of returns. The explanatory power of book-to-market and momentum factors is weak and negligible, respectively.

[23]Examples of this literature include Campbell (1987), Harvey (1989), Shanken (1990), Ferson and Harvey (1991), Ferson and Korajczyk (1995), Kirby (1998), and Avramov (2004). Avramov (2002) reviews and compares the forecasting power of many instrumental variables.

employ five variables: a term structure yield spread, a default risk yield spread, a short-term interest rate, the dividend yield on a broad equity index, and a January dummy variable.[24] *Term* is the difference between the yields of 10-year and one-year U.S. Treasury bonds. The yields are from the fixed term indices of the CRSP Monthly U.S. Treasury Database. *Qual* is the difference between the yields of BAA- and AAA-rated commercial debt. The data are provided by Moody's Investor Service and were obtained from the Federal Reserve Bank of St. Louis. *Bill* is the yield on the U.S. Treasury bill with approximately one month to maturity. *Div Yld* is the dividend yield on the CRSP value-weighted index of NYSE/AMEX/NASDAQ stocks. Following Fama and French (1988), the dividend yield is the sum of the dividends paid on the index in the previous year divided by the level of the index. All yields and spreads are annualized. All instrumental variables are known at the beginning of the period in which asset returns are measured.

V. Model Comparisons: Models with Constant Expected Returns

We estimate a total of 40 models with constant expected returns: five choices for K (i.e., one to five factors) for each of four volatility specifications (i.e., MSV, SVF, SVE, and CF), with and without imposing cross-sectional restrictions suggested by the equilibrium APT. We employ Bayes factors to compare models. Table 2 reports Bayes factors ($\log_{10}(BF_{ij})$) comparing each model to a reference model: the three-factor MSV model constrained by the APT pricing restriction (MSV3f). The Bayes factor for a given cell compares the model for the corresponding row ($\mathcal{M}_i$) to the reference model ($\mathcal{M}_j$). For example, the Bayes factor comparing the unconstrained three-factor MSV model (i.e., MSV3f*) to the APT-constrained three-factor MSV model (i.e., MSV3f) is -26.45. This indicates that the constrained model is decisively favored over the unconstrained model. Recall that $\log_{10}(BF_{ij}) = \log_{10}(m(y|\mathcal{M}_i)) - \log_{10}(m(y|\mathcal{M}_j))$. It follows that Bayes factors comparing any other two models can be easily computed from those provided in Table 2.[25] Models with time-varying expected returns are considered in Section VII.

We find strong evidence of stochastic volatility in both factors and residuals. For a given number of factors, the data decisively favor models with MSV volatility specifications over other volatility specifications. MSV models have the highest marginal likelihoods, followed, in order, by SVF, SVE, and CF specifications. The dominance of SVF models over SVE models suggests that SV is a more important feature for factor shocks than for residual shocks. We conclude that intertemporal variation in the covariance matrix of returns is driven by heteroskedasticity in both factor shocks and idiosyncratic shocks.

[24] Studies that report the forecasting power of these variables include Campbell (1987) for the term structure yield spread; Fama and French (1989) for the default yield spread; Fama and Schwert (1977) and Ferson (1989) for short-term interest rates; Shiller (1984) and Fama and French (1988) for the dividend yield; and Keim (1983) for the January dummy.

[25] The Bayes factor comparing any other pair of models is simply the difference in Bayes factors from that column. For example, the Bayes factor comparing MSV3f* to CF3f* is $-26.45 - (-138.85) = 112.40$.

TABLE 2

Model Comparisons: Models with Constant Expected Returns

Table 2 presents Bayes factors ($\log_{10}(BF_{ij})$) comparing models with constant expected returns for the period 1952:1 to 2003:12. Priors on parameters are uninformative. Each cell compares the model identified with the row ($\mathcal{M}_i$) to the model identified with the column ($\mathcal{M}_j$). The model nomenclature indicates the stochastic volatility specification (MSV, SVF, SVE, or CF), the number of factors, and whether the model is constrained or unconstrained (*) by APT pricing restrictions. We interpret Bayes factors using the scale in Kass and Raftery (1995).

$\mathcal{M}_i$	$\mathcal{M}_j$ MSV3f	$\mathcal{M}_i$	$\mathcal{M}_j$ MSV3f
MSV1f	−700.41	MSV1f*	−711.51
MSV2f	−46.36	MSV2f*	−64.97
MSV3f	0.00	MSV3f*	−26.45
MSV4f	−9.15	MSV4f*	−20.15
MSV5f	−23.81	MSV5f*	−28.17
SVF1f	−1023.24	SVF1f*	−1052.61
SVF2f	−100.99	SVF2f*	−118.21
SVF3f	−53.46	SVF3f*	−68.46
SVF4f	−55.09	SVF4f*	−58.22
SVF5f	−55.98	SVF5f*	−65.44
SVE1f	−858.96	SVE1f*	−863.46
SVE2f	−102.92	SVE2f*	−123.78
SVE3f	−68.08	SVE3f*	−80.98
SVE4f	−72.19	SVE4f*	−80.47
SVE5f	−68.98	SVE5f*	−71.79
CF1f	−1037.79	CF1f*	−1054.03
CF2f	−186.62	CF2f*	−199.25
CF3f	−118.87	CF3f*	−138.85
CF4f	−119.05	CF4f*	−137.96
CF5f	−120.57	CF5f*	−136.18

We find that the APT-constrained three-factor MSV model (i.e., MSV3f) provides the best fit of the data. Of the 40 models with constant expected returns compared in Table 2, the three-factor MSV model with APT pricing restrictions (MSV3f) has the highest marginal likelihood. This is evident since all of the Bayes factors in Table 2 are negative. If we assign equal prior probabilities to each of the 40 models compared in Table 2, the posterior probability of model MSV3f is 1.00.

For a given stochastic volatility specification, models with three or four factors are strongly favored over models with one, two, or five factors. Models with three factors are decisively favored among constrained MSV and SVE specifications. For unconstrained MSV and SVF specifications, models with four factors are decisively favored. For the remaining volatility specifications, the data favor three- and four-factor models, but does not strongly distinguish between them.

Our test of the APT is based on Bayes factors comparing unconstrained models (μ unrestricted) to models constrained by the APT pricing restriction ($\mu = B\gamma$). Without exception, APT-constrained models are decisively favored over unconstrained models holding the number of factors and the stochastic volatility specification constant. This suggests that the pricing restrictions implied by the APT are supported by the data.

A. Sensitivity of Model Comparisons to Prior Specification

Given a large enough sample, posterior densities are relatively insensitive to the specification of priors. In such cases, Bayesian estimation of parameters is rather robust to prior specification. In contrast, model comparisons using Bayes factors (or marginal likelihoods) may be sensitive to prior specification even when the sample evidence is very strong (see, for example, O'Hagan (Ch. 8, (1994)), Kass and Raftery (1995)). In order to evaluate the robustness of the model comparison results reported in Table 2, we must analyze the sensitivity of marginal likelihood computations to the specification of priors. We employ two types of sensitivity analysis.

Our first approach is to replace the uninformative base priors with priors "informed" by the results from a training sample. We employ pre-sample data to construct training samples. Specifically, we use excess returns on 10 CRSP NYSE size decile portfolios for 1927:1 to 1951:12. We examine sensitivity to training sample size by constructing five training samples ranging from 300 months (1927:1–1951:12) to 60 months (1947:1–1951:12). Longer training samples provide more informative (i.e., more precise) priors. For a given training sample, we estimate each of the 40 models using uninformative priors. For each model, the posterior densities obtained from the training sample are used as a guide for specifying the informed priors for the 1952:1–2003:12 "estimation" sample.[26]

The model comparison results are very robust to alternative prior specifications based on training samples. Table 3 summarizes the results of sensitivity analysis to alternative priors. We consider six prior specifications: the uninformed base prior and five "informed" priors based on training samples of different length. For each prior, models are assigned a ranking (i.e., 1 through 40) and a posterior probability. Like the Bayes factors reported in Table 2, these model comparisons are based on estimated marginal likelihoods. Regardless of prior specification, the APT-constrained three-factor MSV model is strongly favored by the data. Model MSV3f is ranked first for every prior, and assigned a posterior probability of 1.00 for four of the six priors. It is clear from Table 3 that one-factor models and homoskedastic (i.e., CF) models are decisively rejected by the data. MSV models are strongly favored over SVF and SVE specifications, and APT-constrained models are generally favored over unconstrained models. All of these conclusions are relatively insensitive to prior specification.

Our second approach analyzes the effect of prior perturbations on model comparison results. Our base priors are very diffuse, so we focus on the effects of making priors more precise. Since the number of potential prior perturbations is infinite, we confine ourselves to analyzing alternative priors that are economically interesting. The most interesting cases are alternative priors that are biased against our empirical findings. We are interested in how precise these alternative priors would have to be in order to alter our conclusions. We reestimate and compare models with more precise (i.e., informed) priors that are intentionally biased against our empirical findings. We consider three experiments along these lines. The results are summarized below.[27]

[26]The means and standard deviations of the training sample posterior densities are used to calibrate the location and precision hyperparameters for the corresponding prior densities. Given possible structural differences between the training and estimation samples (e.g., there was no Great Depression in the 1952–2003 sample), we do not want the informed priors to be overly precise. Accordingly, we select precision hyperparameters such that the informed priors are less precise (i.e., standard deviations roughly doubled) than would be suggested by strict interpretation of the training sample. The resulting informed priors are still much more precise than their uninformed counterparts.

[27]More detailed analysis is available from the authors.

TABLE 3

Sensitivity of Model Comparisons to Prior Specification

Table 3 presents model rankings and posterior probabilities for 40 models with constant expected returns or risk premia. The data are size decile portfolios for the period 1952:1 to 2003:12. Priors are uninformative or informed by a training sample. The model nomenclature indicates the stochastic volatility specification (MSV, SVF, SVE, or CF), the number of factors, and whether the model is constrained or unconstrained (*) by APT pricing restrictions.

	Base Priors		Training Sample Priors									
			47–51		42–51		37–51		32–51		27–51	
Model	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.
MSV1f	33		33		33		35		34		33	
MSV2f	7		5		7		8		10		13	
MSV3f	1	1.00	1	1.00	1	1.00	1	0.62	1	0.95	1	1.00
MSV4f	2		3		4		3	0.10	2	0.05	2	
MSV5f	4		6		3		4		5		5	
SVF1f	37		37		37		38		38		38	
SVF2f	21		24		29		23		23		24	
SVF3f	8		17		20		11		8		19	
SVF4f	9		15		18		15		18		20	
SVF5f	10		16		15		17		16		16	
SVE1f	35		35		35		34		35		35	
SVE2f	22		21		19		18		21		15	
SVE3f	14		9		8		10		7		9	
SVE4f	18		10		9		9		13		8	
SVE5f	16		11		10		12		11		12	
CF1f	38		39		39		39		40		39	
CF2f	31		32		30		31		32		31	
CF3f	24		27		24		27		29		29	
CF4f	25		23		22		26		25		27	
CF5f	26		26		25		25		26		25	
MSV1f*	34		34		34		33		33		34	
MSV2f*	12		8		14		7		9		7	
MSV3f*	5		2		2		2	0.28	6		3	
MSV4f*	3		4		5		5		3		4	
MSV5f*	6		7		6		6		4		6	
SVF1f*	39		38		38		37		37		37	
SVF2f*	23		25		28		24		24		23	
SVF3f*	15		18		21		19		20		22	
SVF4f*	11		19		17		16		19		21	
SVF5f*	13		20		16		21		17		17	
SVE1f*	36		36		36		36		36		36	
SVE2f*	27		22		23		22		22		18	
SVE3f*	20		14		11		20		14		11	
SVE4f*	19		12		12		13		12		10	
SVE5f*	17		13		13		14		15		14	
CF1f*	40		40		40		40		39		40	
CF2f*	32		31		31		32		31		32	
CF3f*	30		28		26		28		30		30	
CF4f*	29		29		27		29		27		28	
CF5f*	28		30		32		30		28		26	

In the first experiment, we attempt to challenge the conclusion that APT-constrained models are favored over unconstrained models. The experiment focuses on the priors for μ_i in the unconstrained model. For each size portfolio, we set the hyperparameter μ_i^0 equal to the sample mean of the portfolio's returns, and set M_i^0 to the standard error of the sample mean. These priors would be reasonable for a researcher who believes that average portfolio returns are unrelated to factor loadings (i.e., $\alpha \neq B\gamma$). Note that we have not tightened the priors on γ in the corresponding constrained models. Tightening the prior on μ_i for the unconstrained model while retaining the base prior for the constrained model strongly biases the model comparison against the APT pricing restriction. Yet, even with these biased priors, APT-constrained models are still favored over their unconstrained counterparts in most cases. In particular, the constrained three-factor MSV model remains decisively favored over the unconstrained model with a Bayes factor of 3.93.

The second experiment challenges our conclusion that the data favor models with three factors. This experiment focuses on the joint prior for γ and B. Recall that our base prior for each of the free elements in B is very diffuse and centered on zero. The base prior on γ, the factor risk premia, is also very diffuse and centered on zero. Now, consider a researcher who is biased against models with more than two factors. To accommodate this view, we entertain alternative priors that are increasingly precise, but still centered on zero, for the loadings on f_3, f_4, and f_5. Furthermore, our alternative priors on the factor risk premia (γ_3, γ_4, and γ_5) are much tighter: centered on zero with a standard deviation of only 10 bp per month. The alternative priors for loadings and risk premia on f_1 and f_2 are identical to the base prior. We reestimate constrained MSV models with three, four, and five factors, and compare these models to the previously estimated two-factor MSV model. Even with relatively tight priors, the constrained three-factor MSV model is decisively favored over the otherwise identical two-factor model. Only when the priors on B_{i3}, B_{i4}, and B_{i5} become dogmatic (i.e., prior standard deviations of only 0.003) does the data favor the two-factor model over the three-factor model. Torturing the data in this manner clearly is not reasonable. Overall, our conclusion that the data favor the three-factor model is very robust to these prior perturbations.

The third experiment challenges the conclusion that homoskedastic models are strongly rejected by the data. For this experiment, we perturb the priors on ϕ_j and σ_j in (6) to reflect the belief that the V_t and/or D_t matrices in (5) are time invariant. If the volatility of a particular factor or idiosyncratic shock (e.g., f_j) is homoskedastic, then $\phi_j = 0$ and $\sigma_j = 0$. Accordingly, we consider alternative priors on ϕ_j and σ_j that are more precise and closer to zero. We do not perturb the prior on κ_j, the steady-state log-variance. When $\phi_j \rightarrow 0$ and $\sigma_j \rightarrow 0$, κ_j must be free to fit the level of the log-variance. We reestimate 24 models: three volatility specifications (MSV, SVF, and SVE), two-factor through five-factor, constrained and unconstrained. These priors should bias model comparison results in favor of the CF models estimated under the base priors. Of all of the models compared, the constrained three-factor MSV model is decisively favored by the data for all but the most biased prior. When priors on ϕ_j are made unreasonably tight around zero, the data favor the constrained three-factor SVF model over the constrained three-factor MSV model. Our conclusion that the data decisively favor MSV models over SVF, SVE, and CF models appears to be very robust to prior perturbations.

Taken as a whole, this extensive battery of sensitivity checks indicates that our model comparison results are very robust to changes in the priors.

VI. Empirical Results: Models with Constant Expected Returns

Section V reported that the data strongly favor the APT-constrained three-factor MSV model over all other competing models with constant expected returns. Accordingly, this section is devoted to further analysis of that model. Subsection A discusses the estimated factor loadings and factor risk premia. In subsection B, we discuss cross-sectional variation in portfolio expected returns for the APT-constrained and unconstrained variants of the model. We also analyze the time series of latent factor shocks. Subsection C discusses the nature of the stochastic volatility processes for factor shocks and idiosyncratic shocks. In subsection D, we revisit the comparisons between models with constant expected returns discussed in Section V and summarize the main empirical differences between model specifications.

A. Factor Loadings and Factor Risk Premia

Table 4 reports summary statistics describing the posterior densities of each parameter in the γ (constant factor risk premia) vector and the B (factor loading) matrix. The mean of the posterior density, which we refer to as the posterior estimate, is analogous to the point estimate in frequentist methods. We evaluate the precision of posterior estimates in two ways. The standard deviation of the MCMC draws from the posterior density much like the standard error of the estimate. The 5% and 95% quantile values are the endpoints of a 90% Bayesian confidence interval.

TABLE 4

Factor Loadings and Factor Risk Premia

Table 4 reports summary statistics describing the Bayesian posterior densities of parameters for the three-factor MSV model with constant factor risk premia and APT pricing restrictions (MSV3f). The data are monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). Priors on the parameters are uninformative. The model was estimated using Markov chain Monte Carlo (MCMC) techniques. The posterior mean, standard deviation, and 5th and 95th percentile critical values for each parameter are based on 10,000 Gibbs sampler iterations from a suitably constructed Markov chain.

	Factor Risk Premia (γ_j) (% per month)				Factor Average Return ($\gamma_j + \overline{f_j}$) (% per month)			
Factor	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
f_1	0.961	0.190	0.645	1.270	0.742	0.015	0.716	0.766
f_2	0.178	0.080	0.047	0.310	0.125	0.037	0.065	0.187
f_3	0.315	0.112	0.133	0.501	0.503	0.073	0.383	0.626

	Loadings on f_1 (B_{i1})				Loadings on f_2 (B_{i2})				Loadings on f_3 (B_{i3})			
Port.	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Cap1	1.061	0.035	1.004	1.119	−1.077	0.107	−1.254	−0.902	1			
Cap2	1.052	0.024	1.013	1.092	−0.765	0.075	−0.889	−0.642	0.652	0.032	0.600	0.705
Cap3	1.035	0.018	1.006	1.064	−0.576	0.056	−0.670	−0.486	0.415	0.026	0.371	0.458
Cap4	1.019	0.014	0.997	1.042	−0.404	0.041	−0.473	−0.337	0.232	0.023	0.194	0.271
Cap5	1.012	0.010	0.995	1.028	−0.217	0.032	−0.271	−0.165	0.095	0.020	0.062	0.128
Cap6	1				0				0			
Cap7	0.961	0.010	0.945	0.977	0.224	0.026	0.182	0.267	−0.069	0.021	−0.104	−0.035
Cap8	0.907	0.012	0.887	0.927	0.459	0.028	0.414	0.506	−0.078	0.020	−0.112	−0.046
Cap9	0.833	0.016	0.807	0.859	0.718	0.033	0.664	0.774	−0.060	0.023	−0.098	−0.022
Cap10	0.635	0.025	0.594	0.676	1				0			

Since factors are latent, we impose a hierarchical structural constraint on the loading matrix B. Since the manner in which B is constrained affects parameter estimates and their interpretation, it deserves further explanation. Rather than imposing an arbitrary constraint, we are guided by the results of a preliminary principal components analysis. This analysis indicated that the *Cap6* portfolio was the most highly correlated with the first static factor. Accordingly, we assign

Cap6 a factor loading of one on f_1 and factor loadings of zero on the remaining factors. Following this approach, we assign unit factor loadings on f_2 and f_3 to *Cap10* and *Cap1*, respectively.

We loosely interpret f_1 as a "stock market factor." The relation between B_{i1} and size is relatively flat for all but the largest decile portfolios. For the smallest size portfolios (*Cap1–Cap3*), posterior estimates of the factor loadings (B_{i1}) are slightly greater than one. For portfolios *Cap4* and *Cap5*, posterior estimates of B_{i1} are not significantly different from one (i.e., one falls within the Bayesian confidence interval). The factor loadings decrease monotonically with size for the larger stock portfolios (*Cap7–Cap10*). It is noteworthy that the relation between B_{i1} and size bears little resemblance to the inverse monotonic relation between size and average excess returns reported in Table 1.

Exposure to f_1 is rewarded with a significant risk premium. The posterior estimate of γ_1 is 0.961 with a Bayesian confidence interval of $[0.645, 1.270]$. This indicates that the stock market factor has a significant constant risk premium of about 96 bp per month per unit of exposure. Table 4 also reports the posterior density of $\gamma_1 + \overline{f}_1$ (where $\overline{f}_1$ is the time-series average of f_1 shocks), which we interpret as the ex post average factor return. The posterior estimate of $\gamma_1 + \overline{f}_1$ is 0.742 with a Bayesian confidence interval of $[0.716, 0.766]$. The tighter confidence interval indicates that $\gamma_1 + \overline{f}_1$ is estimated with higher precision than γ_1 alone. Taken together, the posterior estimates of B_{i1} and γ_1 suggest that exposure to f_1 can explain the average level of expected returns for size decile portfolios. However, they cannot fully explain cross-sectional variation in expected returns of size decile portfolios.

We loosely interpret f_2 as a "size factor" since the factor loadings decrease monotonically with size. Posterior estimates of B_{i2} are negative for *Cap1–Cap5*, and positive for *Cap7–Cap9*. We find that the size factor earns a small, but significant, constant risk premium. The posterior estimate of γ_2 reported in Table 4 is 0.178 or about 18 bp per month with a Bayesian confidence interval of $[0.047, 0.310]$. These estimates of B_{i2} and γ_2 suggest slightly lower expected returns for small size portfolios and slightly higher expected returns for large size portfolios.

We assign a unit factor loading on f_3 to *Cap1*, the smallest market capitalization decile portfolio. Posterior estimates of B_{i3} are positive and decrease monotonically with size from *Cap2–Cap5*. Posterior estimates of B_{i3} for *Cap7–Cap9* are close to zero. Factor loadings for *Cap6* and *Cap10* are constrained to be zero. We loosely interpret f_3 as a "small stock factor" that is distinct from f_2 (the size factor). Exposure to f_3 also earns a significant constant risk premium. The posterior estimate of γ_3 is 0.315 (about 32 bp per month) with a Bayesian confidence interval of $[0.133, 0.501]$. Higher factor loadings on f_3 contribute substantially to explaining the higher volatility and expected returns of small stock portfolios.

B. Analysis of Ex Post Average Portfolio Returns

In the models we estimate, ex post average excess returns can be decomposed into a constant expected return component ($B_i\gamma$ for the constrained model, μ_i for the unconstrained model), an average factor shock component ($B_i\overline{f}$), and

https://doi.org/10.1017/S0022109000003422 Published online by Cambridge University Press

an average idiosyncratic shock component ($\bar{\epsilon}_i$). Unlike frequentist methods, the Bayesian estimation methodology does not constrain average factor shocks ($\bar{f}$) to be zero. Rather, the latent factor shocks are sampled conditional on the data, the latent stochastic volatilities, and the other parameters of the model.

How well does model MSV3f fit the cross section of average returns? Panel A of Table 5 reports summary statistics describing the posterior densities of $B_i\gamma$, $B_i\bar{f}$, and $B_i(\gamma+\bar{f})$ for the APT-constrained three-factor MSV model. Posterior estimates of $B_i(\gamma+\bar{f})$, the average return on portfolio i implied by the constrained model, range from 59.5 bp per month for *Cap10* to 115.7 bp per month for *Cap1*. Posterior estimates of $B_i(\gamma+\bar{f})$ match the mean excess returns reported in Table 1 very closely. We conclude that the APT-constrained three-factor MSV model performs adequately in explaining the cross section of average excess returns for size decile portfolios.

TABLE 5

Decomposition of Ex Post Average Portfolio Returns

Table 5 reports summary statistics describing the Bayesian posterior densities of the risk premia and average factor shock components for both the APT-constrained and unconstrained variants of the three-factor MSV model with constant risk premia. Returns are expressed in percent per month. The data are monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). Priors on the parameters are uninformative. The model was estimated using Markov chain Monte Carlo (MCMC) techniques. The posterior mean, standard deviation, and 5th and 95th percentile critical values for each parameter are based on 10,000 Gibbs sampler iterations from a suitably constructed Markov chain.

Panel A. Constrained Three-Factor MSV Model

	Posterior Density of $B_i\gamma$				Posterior Density of $B_i\bar{f}$				Posterior Density of $B_i(\gamma+\bar{f})$			
Port.	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Cap1	1.145	0.235	0.757	1.531	0.013	0.233	−0.371	0.394	1.157	0.042	1.087	1.228
Cap2	1.081	0.215	0.726	1.436	−0.068	0.214	−0.422	0.285	1.013	0.023	0.975	1.051
Cap3	1.023	0.204	0.685	1.357	−0.119	0.203	−0.454	0.219	0.904	0.018	0.875	0.934
Cap4	0.981	0.196	0.654	1.303	−0.159	0.196	−0.482	0.168	0.823	0.016	0.796	0.850
Cap5	0.964	0.192	0.644	1.278	−0.192	0.192	−0.506	0.127	0.771	0.016	0.745	0.797
Cap6	0.961	0.190	0.645	1.270	−0.219	0.189	−0.528	0.096	0.742	0.015	0.716	0.766
Cap7	0.942	0.183	0.636	1.238	−0.236	0.183	−0.533	0.069	0.706	0.015	0.682	0.730
Cap8	0.929	0.175	0.637	1.212	−0.238	0.175	−0.522	0.053	0.691	0.013	0.668	0.713
Cap9	0.909	0.167	0.631	1.177	−0.232	0.166	−0.501	0.045	0.677	0.016	0.650	0.703
Cap10	0.788	0.144	0.549	1.024	−0.192	0.141	−0.424	0.043	0.595	0.030	0.546	0.646

Panel B. Unconstrained Three-Factor MSV Model

	Posterior Density of μ_i				Posterior Density of $B_i\bar{f}$				Posterior Density of $\mu_i + B_i\bar{f}$			
Port.	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Cap1	1.250	0.242	0.852	1.641	0.014	0.237	−0.366	0.411	1.265	0.058	1.168	1.359
Cap2	1.074	0.220	0.709	1.432	−0.060	0.218	−0.413	0.305	1.014	0.033	0.960	1.067
Cap3	0.942	0.209	0.594	1.285	−0.108	0.207	−0.447	0.239	0.834	0.032	0.782	0.885
Cap4	0.970	0.201	0.638	1.304	−0.146	0.199	−0.477	0.185	0.824	0.030	0.774	0.873
Cap5	0.941	0.196	0.617	1.270	−0.178	0.194	−0.501	0.142	0.762	0.028	0.716	0.808
Cap6	0.952	0.192	0.640	1.273	−0.204	0.190	−0.522	0.106	0.748	0.025	0.707	0.790
Cap7	0.935	0.184	0.637	1.242	−0.219	0.183	−0.523	0.075	0.716	0.025	0.674	0.757
Cap8	0.928	0.176	0.645	1.224	−0.221	0.174	−0.515	0.058	0.707	0.023	0.668	0.745
Cap9	0.896	0.166	0.629	1.176	−0.215	0.164	−0.493	0.051	0.681	0.019	0.648	0.713
Cap10	0.682	0.148	0.442	0.930	−0.177	0.139	−0.413	0.051	0.505	0.055	0.415	0.594

Recall that the Bayes factors reported in Table 2 indicate that the data decisively favor APT-constrained models over otherwise identical unconstrained models. For completeness, Panel B of Table 5 reports summary statistics describing the posterior densities of μ_i, $B_i\bar{f}$, and $\mu_i + B_i\bar{f}$ for the unconstrained three-factor MSV model (i.e., MSV3f*). Comparing Panels A and B of Table 5 provides

https://doi.org/10.1017/S0022109000003422 Published online by Cambridge University Press

some intuition to complement the model comparison results. The patterns observed in Panels A and B are generally similar. The posterior estimates of μ_i for the unconstrained model exhibit slightly more cross-sectional variation than their constrained counterparts, $B_i\gamma$, but less cross-sectional variation than the sample means reported in Table 1. This is probably because the μ_i vector is free to fit the cross section of average realized excess returns without having to simultaneously explain the covariance matrix of returns (i.e., the APT pricing restriction). That the Bayes factor favors the constrained model indicates that the improved fit provided by the μ_i vector over the $B_i\gamma$ vector is not worth the cost of seven fewer degrees of freedom.

Figure 1 plots the time series of cumulative factor returns (i.e., $\gamma_j + f_j$) for f_1, f_2, and f_3. Figure 2 plots the time series of cumulative factor shocks (i.e., f_j alone) and is essentially equivalent to Figure 1 with the cumulative factor risk premia removed. Each point represents the mean of 10,000 MCMC sample draws. Several features of these plots are remarkable. First, the cumulative factor shocks for f_1 are negative for much of the sample period. Cumulative shocks for f_1 are relatively flat (i.e., close to zero) prior to 1968, drop sharply during 1969–1973, and trend slightly downward for the last 20 years of the sample. This pattern might indicate the existence of regimes with different risk premiums. Recall from Table 4 that posterior estimates of $\gamma_1 + \overline{f}_1$ are about 20 bp per month lower than posterior estimates of γ_1 alone. This is consistent with the plot of cumulative factor shocks for f_1 in Figure 2. The behavior of the two size-related factors (f_2 and f_3) is also interesting. Recall that the smaller size decile portfolios have negative factor loadings on f_2 and that the larger size decile portfolios have positive factor loadings on f_2. The size factor (f_2) experiences a substantial decline from the mid-1960s to the early 1980s. This is consistent with the size effect first documented by Banz (1981) and Reinganum (1981). The size factor rebounded in the late 1980s and has been relatively flat since then. This corresponds with the apparent end of the size effect documented by Schwert (2002). The small stock factor (f_3) experienced substantial cumulative gains in the late 1960s and again in the 1990s.

C. Factor and Idiosyncratic Stochastic Volatility Processes

Table 6 reports summary statistics describing the posterior densities of the stochastic volatility parameters for the APT-constrained three-factor MSV model with constant factor risk premia. For each of the factor and idiosyncratic volatility processes, we report on each of the three parameters in equation (6). κ_j, the intercept of the log-variance equation, is difficult to interpret. Instead, we report the density of $\exp(\kappa_j/2)$ (in percent per month), which can be interpreted as a steady-state standard deviation. ϕ_j is the persistence parameter and σ_j is the volatility of the log-variance. The most volatile factor is f_1, for which the posterior estimate of $\exp(\kappa_j/2)$ is 4.78% per month. Given the factor loadings on f_1, it is apparent that the stock market factor accounts for most of the variance in portfolio excess returns. Posterior estimates of $\exp(\kappa_j/2)$ are 1.84% and 2.09% per month, respectively, for f_2 and f_3. Idiosyncratic volatilities are small relative to factor volatilities. The posterior estimates of $\exp(\kappa_j/2)$ are less than 1% per month for

FIGURE 1

Cumulative Factor Returns

Time-series plots of cumulative returns ($\gamma_j + f_j$) for factors 1, 2, and 3 from the APT-constrained three-factor MSV model with constant factor risk premia. Each point is the mean of 10,000 MCMC draws. Note the log scale for the y-axis.

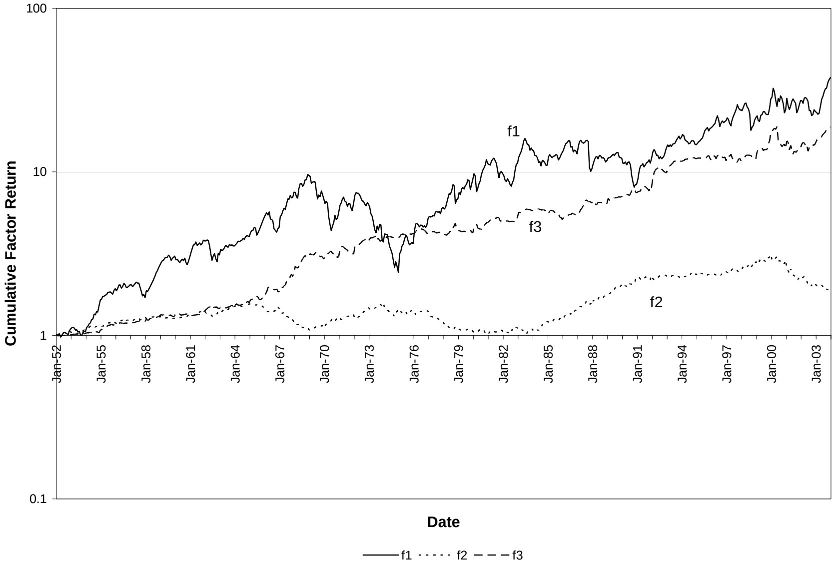
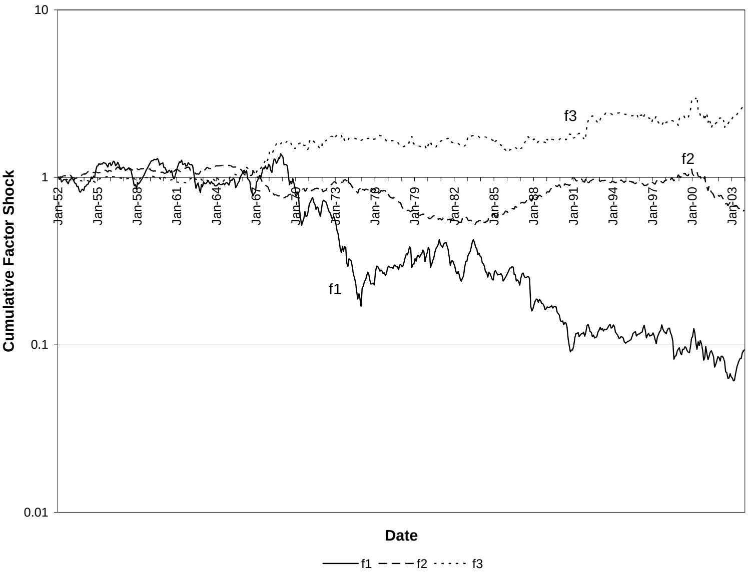


FIGURE 2

Cumulative Factor Shocks

Time-series plots of cumulative shocks (f_j only) for factors 1, 2, and 3 from the APT-constrained three-factor MSV model with constant factor risk premia. Each point is the mean of 10,000 MCMC draws. Note the log scale for the y-axis.



eight of the 10 size decile portfolios. We conclude that idiosyncratic volatilities contribute relatively little to total return volatility in a three-factor linear factor model. Posterior estimates of the persistence parameters, ϕ_j, are greater than 0.9 for the latent factor volatilities, and greater than 0.79 for the idiosyncratic shock volatilities. The small stock factor is the most likely to experience large volatility shocks. The posterior estimate of σ_j for f_3 is 0.499 with a confidence interval of [0.244, 0.789]. Similarly, the posterior estimate of σ_j for the smallest size decile portfolio (*Cap1*) is much larger than the posterior estimates of σ_j for the other size decile portfolios.

TABLE 6

Stochastic Volatility Processes

Table 6 reports summary statistics describing the Bayesian posterior densities of parameters for the stochastic volatility processes (equation (6)) for the APT-constrained three-factor MSV model with constant factor risk premia (MSV3f). The data are monthly excess returns for 10 NYSE-AMEX-NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). Priors on the parameters are uninformative. The model was estimated using Markov chain Monte Carlo (MCMC) techniques. The posterior mean, standard deviation, and 5th and 95th percentile critical values for each parameter are based on 10,000 Gibbs sampler iterations from a suitably constructed Markov chain.

	$\exp(\kappa_j/2)$ (% per month)				ϕ_j				σ_j			
Factor/ Port.	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
f_1	4.778	0.461	4.069	5.554	0.916	0.035	0.852	0.964	0.319	0.068	0.217	0.437
f_2	1.837	0.245	1.472	2.264	0.959	0.022	0.918	0.987	0.206	0.057	0.126	0.309
f_3	2.086	0.346	1.602	2.627	0.902	0.057	0.794	0.977	0.499	0.167	0.244	0.789
Cap1	1.336	0.154	1.091	1.599	0.875	0.048	0.788	0.943	0.541	0.114	0.362	0.735
Cap2	0.793	0.052	0.714	0.884	0.829	0.117	0.603	0.970	0.149	0.077	0.059	0.300
Cap3	0.794	0.063	0.698	0.901	0.937	0.039	0.865	0.982	0.177	0.060	0.096	0.288
Cap4	0.749	0.066	0.651	0.857	0.933	0.032	0.874	0.976	0.209	0.052	0.135	0.303
Cap5	0.689	0.028	0.642	0.735	0.794	0.114	0.576	0.936	0.119	0.052	0.053	0.218
Cap6	0.619	0.041	0.555	0.684	0.889	0.063	0.769	0.964	0.224	0.079	0.115	0.371
Cap7	0.641	0.102	0.493	0.813	0.977	0.015	0.950	0.993	0.133	0.036	0.085	0.198
Cap8	0.570	0.058	0.480	0.665	0.953	0.026	0.905	0.986	0.168	0.048	0.101	0.253
Cap9	0.475	0.039	0.419	0.534	0.868	0.109	0.654	0.983	0.115	0.047	0.054	0.204
Cap10	1.360	0.185	1.084	1.673	0.957	0.023	0.913	0.986	0.231	0.061	0.145	0.344

Figure 3 plots the time series of stochastic volatilities for f_1, f_2, and f_3. All three of these plots display clear evidence of volatility clustering and persistence. This is consistent with the posterior estimates of ϕ_j and σ_j reported in Table 6. In particular, the stochastic volatility plot for f_3 shows that the small stock factor is more prone to large volatility shocks (i.e., higher σ_j) and is less persistent (i.e., lower ϕ_j) than f_1 or f_2. With the exception of *Cap1*, plots of stochastic volatilities for the idiosyncratic shocks are relatively flat and unremarkable.[28] This is consistent with high persistence and the relatively low posterior estimates of σ_j for *Cap2* through *Cap10*. High volatility for f_1 is associated with several episodes in the 1970s, the period surrounding the Federal Reserve regime change in 1979, the October 1987 stock market crash, and the Internet/technology "bubble." The highest volatilities for f_2 and f_3 both coincide with the collapse of the Internet/technology bubble in 2000.

D. Model Comparisons Revisited

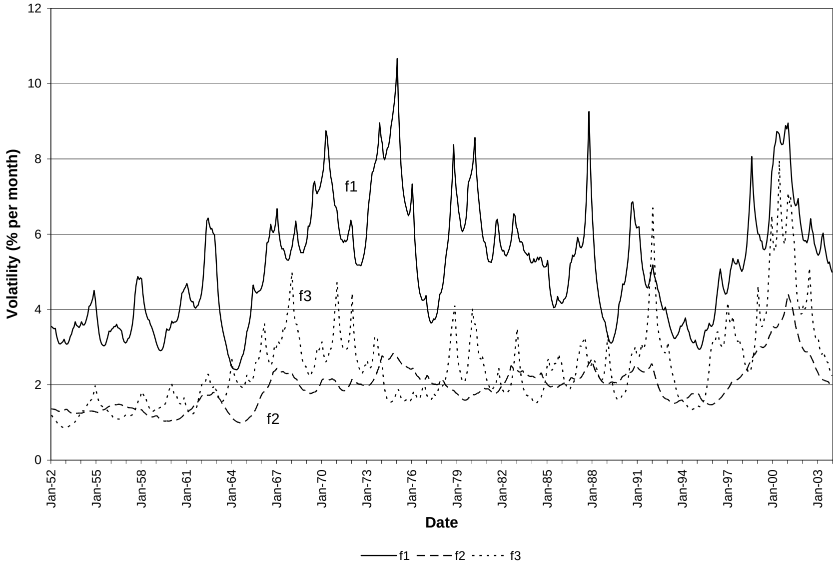
We have spent much of this section describing empirical results for the APT-constrained three-factor MSV model since that specification is strongly favored

[28]We omit these plots to conserve space.

FIGURE 3

Factor Stochastic Volatilities

Time-series plots of stochastic volatilities (in percent per month) for factors 1, 2, and 3 from the APT-constrained three-factor MSV model with constant factor risk premia. Each point is the mean of 10,000 MCMC draws.



by the Bayes factors reported in Table 2. Although detailed analysis of empirical results for 40 different specifications is not practical, it is possible to summarize, in general terms, the major differences between classes of models. In doing so, we hope to provide some intuition for the Bayesian model comparison results.

For a given stochastic volatility specification, additional factors improve the fit of the covariance matrix by explaining common variation in portfolio excess returns. As one would expect, idiosyncratic volatilities decline as the number of factors increases. We find that cross-sectional variation in the loading on f_1 (B_{i1}) across stock portfolios changes little when f_2 is added. We also find that the precision of the posterior estimates of B_{i1} decreases somewhat as the number of factors increases for MSV and SVF models. However, increasing the number of factors does not appear to significantly change estimates of B, μ, γ, or patterns in cumulative factor shocks.

Behavior of factor shocks is generally similar across models. MSV and SVF models behave somewhat differently than SVE and CF models because their factor volatilities are stochastic rather than static. For example, the cumulative declines in f_1 during 1973–1974 are larger for MSV and SVF models (see Figures 1 and 2) since the volatility of f_1 is very high during that period (see Figure 3). Differences in average factor shocks across models are associated with compensating differences in posterior estimates of μ or γ, the constant expected returns or factor risk premia. For example, the posterior estimate of γ_1 (i.e., the constant risk premia for factor f_1) in the APT-constrained three-factor CF model is 74.7 bp per month. The comparable posterior estimate of γ_1 for the APT-constrained three-factor MSV model (reported in Table 4) is 96.1 bp per month. The difference is associated with the magnitudes of the average factor shocks for these two models.

The Bayes factors reported in Table 2 indicate that, all else equal, models constrained by APT pricing restrictions are favored over unconstrained models in every case. However, we find few substantive differences in posterior estimates of B, θ, or in factor shocks between otherwise identical models. It appears that any gains from estimating the N-vector μ, rather than the K-vector γ, are insufficient to overcome the implicit penalty for additional parameters imposed in the computation of the marginal likelihood.

VII. Empirical Results: Models with Time-Varying Expected Returns

In this section, we consider models with time-varying expected returns. We are interested in two questions: i) do models with time-varying expected returns provide a better fit of the data than models with constant expected returns, and ii) is return predictability related to exposures to latent factors?

Panel A of Table 7 addresses the first question. It reports additional Bayes factors comparing models with time-varying expected returns to model MSV3f (the reference model for Table 2) for the period 1952–2003. Comparisons are based on marginal likelihoods computed under uninformative base priors. Given the decisive evidence in favor of models with MSV, we consider only MSV specifications in this section. Since the reference model (MSV3f) is the same, Bayes factors comparing any pair of models from Tables 2 or 7 are easy to compute. In Table 7, an asterisk (*) denotes an unconstrained model, and a dagger (†) denotes a model with time-varying returns or risk premia. Among models with time-varying expected returns, the model with the highest marginal likelihood is the three-factor MSV model constrained by APT pricing restrictions (MSV3f†). Furthermore, a Bayes factor of 8.82 indicates that the data decisively favor the APT-constrained three-factor MSV model with time-varying factor risk premia (MSV3f†) over the otherwise identical model with constant factor risk premia (MSV3f). Similar results hold for one- and two-factor MSV models, but not for four- and five-factor MSV models. However, unconstrained MSV models with time-varying expected returns are not favored over otherwise identical models with constant expected returns.[29] We conclude that allowing for time-varying risk premia improves the fit of APT-constrained models with up to three factors, but does not improve the fit for more parameter-rich models (i.e., constrained models with more than three factors or unconstrained models). It should also be noted that the improvement in marginal likelihoods from allowing for time-varying expected returns is much less than the improvement in marginal likelihoods from allowing for stochastic volatility or adding a second latent factor.

[29]To make these comparisons, one must examine the differences between the Bayes factors reported in Tables 2 and 7. For example, the Bayes factor comparing an unconstrained model with time-varying expected returns (e.g., MSV3f†*) to an otherwise identical unconstrained model with constant expected returns (e.g., MSV3f*) is the difference between the Bayes factor reported in Table 7 (−58.05) and the Bayes factor reported in Table 2 (−26.45). The resulting Bayes factor, −31.60, indicates that the data decisively favor the model with constant expected returns.

TABLE 7

Model Comparisons: Models with Time-Varying Expected Returns

Table 7 presents Bayes factors ($\log_{10}(BF_{ij})$) comparing models with time-varying expected returns for 1952:1 to 2003:12. Priors on parameters are uninformative. Each cell compares the model identified with the row ($\mathcal{M}_i$) to the model identified with the column ($\mathcal{M}_j$). The model nomenclature indicates the stochastic volatility specification (MSV), the number of factors, whether expected returns/risk premia are constant or time varying (†), and whether the model is constrained by APT pricing restrictions or unconstrained (*). We interpret Bayes factors using the scale in Kass and Raftery (1995).

Panel A. Bayes Factors Comparing Models with Time-Varying Expected Returns to Models with Constant Expected Returns

$\mathcal{M}_i$	$\mathcal{M}_j$ MSV3f	$\mathcal{M}_i$	$\mathcal{M}_j$ MSV3f
MSV1f†	−659.43	MSV1f†*	−785.08
MSV2f†	−27.15	MSV2f†*	−130.98
MSV3f†	8.82	MSV3f†*	−58.05
MSV4f†	−10.94	MSV4f†*	−81.87
MSV5f†	−22.89	MSV5f†*	−76.69

Panel B. Bayes Factors Comparing Models Constrained by APT Pricing Restrictions to Unconstrained Models

	$\mathcal{M}_j$				
$\mathcal{M}_i$	MSV1f†*	MSV2f†*	MSV3f†*	MSV4f†*	MSV5f†*
MSV1f†	125.65				
MSV2f†		103.82			
MSV3f†			66.87		
MSV4f†				70.93	
MSV5f†					53.80

The second question we address is whether return predictability is related to exposures to latent factors. If return predictability is due to time-varying risk premia, as suggested by proponents of market efficiency and rational expectations, then it should be proportional to exposures (i.e., factor loadings) to latent factors. We investigate by comparing APT-constrained models with time-varying risk premia to otherwise identical unconstrained models with time-varying expected returns. Bayes factors making these comparisons are reported in Panel B of Table 7. In each case, the data decisively favor the APT-constrained model over the otherwise identical (and less parsimonious) unconstrained model. These results are similar in spirit to those reported in Ferson and Korajczyk (1995). We conclude that return predictability is related to exposures to latent factors.

The model comparisons discussed in this section are very robust to changes in prior specification. Table 8 summarizes the results of sensitivity analysis to alternative priors. As in Table 3, we consider six alternative prior specifications: the uninformed base prior and five informed priors based on training samples of different length. For each prior, Table 8 reports the rankings (1 through 20) and posterior probabilities for the 20 MSV models with and without time-varying expected returns/risk premia. Regardless of prior, the constrained three-factor model with time-varying risk premia (MSV3f†) is ranked first and assigned a posterior probability of 1.00. It is interesting to note that the corresponding model with constant risk premia (i.e., MSV3f, the top-ranked model in Tables 2 and 3), is ranked second for four of the six priors in Table 8.

We also compare models under alternative priors biased against time-varying risk premia. Under the base prior, the densities of the elements of the G matrix are i.i.d. normal and centered on zero. Under the "biased" prior, we tighten the distributions around zero for all elements of the G matrix except those in the first

https://doi.org/10.1017/S0022109000003422 Published online by Cambridge University Press

TABLE 8

Sensitivity of Model Comparisons to Prior Specification

Table 8 presents model rankings and posterior probabilities for 20 models with both constant and time-varying expected returns/risk premia. The data are size decile portfolios for the period 1952:1 to 2003:12. Priors are uninformative or informed by a training sample. The model nomenclature indicates the stochastic volatility specification (MSV), the number of factors, and whether the model is constrained or unconstrained (*) by APT pricing restrictions, and whether expected returns/risk premia are time varying (†).

	Base Priors		Training Sample Priors									
			47–51		42–51		37–51		32–51		27–51	
Model	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.
MSV1f	18		17		17		20		18		18	
MSV2f	11		8		13		14		15		14	
MSV3f	2		2		2		3		4		2	
MSV4f	3		6		7		5		5		4	
MSV5f	7		10		6		6		7		9	
MSV1f*	19		19		19		19		17		19	
MSV2f*	13		14		15		12		14		13	
MSV3f*	8		4		5		4		8		5	
MSV4f*	5		7		8		8		3		6	
MSV5f*	10		13		12		9		6		11	
MSV1f†	17		18		18		18		20		17	
MSV2f†	9		9		11		15		13		15	
MSV3f†	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00	1	1.00
MSV4f†	4		3		3		2		2		3	
MSV5f†	6		5		4		7		9		7	
MSV1f†*	20		20		20		17		19		20	
MSV2f†*	16		16		16		16		16		16	
MSV3f†*	12		11		9		10		10		10	
MSV4f†*	15		12		10		11		11		8	
MSV5f†*	14		15		14		13		12		12	

column (i.e., the constants). We find that the priors must be extremely tight in order for constant risk premia models to dominate models with time-varying risk premia. Details are available from the authors.

All of the results reports thus far are based on the 1952–2003 estimation sample. As an additional robustness check, we partition the full 1927–2003 sample into training and estimation samples at different breakpoints, and repeat the analysis described above. Table 9 reports model rankings and posterior probabilities for seven estimation samples: 1927–2003 (estimated under the base prior; no training sample), 1932–2003, 1937–2003, 1942–2003, 1946–2003, 1952–2003 (the sample analyzed in Table 8), and 1963–2003. The training sample in each case begins in 1927. Unlike the analysis reported in Tables 3 and 8, this analysis checks the sensitivity of the model comparison results to simultaneous changes in the prior and in the estimation sample. The conclusion that APT-constrained models are favored over unconstrained models is very robust. Three-factor models are strongly favored in most periods. However, five-factor models are favored in the 1942–2003 sample and four-factor models are favored in the 1963–2003 sample. Models with constant risk premia are favored over models with time-varying risk premia in two estimation samples: 1937–2003 and 1963–2003. Since the model comparisons reported in Table 8 for the 1952–2003 estimation sample are relatively insensitive to changes in priors, it is likely that the changes in model comparisons reported in Table 9 are due to changes in the estimation sample rather than changes in the priors.

Tables 10 and 11 report summary statistics describing the posterior densities (estimated under the base priors for the 1952–2003 sample) of the factor loading

TABLE 9

Sensitivity of Model Comparisons to Sample Selection and Prior Specification

Table 9 reports model rankings and posterior probabilities for 20 models with both constant and time-varying expected returns/risk premia. The data are size decile portfolios for the period 1927:1 to 2003:12. Each column reports results for a different partition of the data into a training sample and estimation sample. The model nomenclature indicates the stochastic volatility specification (MSV), the number of factors, and whether the model is constrained or unconstrained (*) by APT pricing restrictions, and whether expected returns/risk premia are time varying (†).

Training	N/A		27–31		27–36		27–41		27–45		27–51		27–62	
Estimation	27–03		32–03		37–03		42–03		46–03		52–03		63–03	
Model	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.	Rank	Prob.
MSV1f	18		18		18		19		18		18		17	
MSV2f	10		13		9		11		14		14		15	
MSV3f	4		4		1	1.00	5		2	0.01	2		5	
MSV4f	5		5		3		6		5		4		1	1.00
MSV5f	6		7		10		2	0.04	6		9		6	
MSV1f*	17		19		19		18		19		19		19	
MSV2f*	16		15		16		16		15		13		14	
MSV3f*	9		10		6		8		9		5		7	
MSV4f*	11		9		11		10		8		6		4	
MSV5f*	12		8		12		12		7		11		9	
MSV1f†	19		17		17		17		17		17		18	
MSV2f†	13		3		7		14		13		15		13	
MSV3f†	1	0.73	1	1.00	2		3		1	0.99	1	1.00	3	
MSV4f†	2	0.27	2		4		4		3		3		2	
MSV5f†	3		6		5		1	0.96	4		7		8	
MSV1f†*	20		20		20		20		20		20		20	
MSV2f†*	15		12		8		15		16		16		16	
MSV3f†*	14		11		15		7		10		10		11	
MSV4f†*	7		14		14		9		12		8		10	
MSV5f†*	8		16		13		13		11		12		12	

matrix (B) and the stochastic volatility parameters ($\exp(\kappa_j/2)$, ϕ_j, and σ_j) for the APT-constrained three-factor model with time-varying risk premia. The results are remarkably similar to those reported in Tables 4 and 6 for the otherwise identical constant risk premia model. This indicates that allowing factor risk premia to be time varying has very little effect on posterior estimates of the model's parameters. Likewise, plots of cumulative factor returns and stochastic volatilities for the time-varying risk premia model are almost indistinguishable from Figures 1 and 3.[30] The unique feature of the time-varying risk premia model is the coefficient matrix G. Table 12 reports summary statistics describing the posterior densities of the G matrix. The risk premia for f_1 (the stock market factor) is positively related to the default risk yield spread and the January dummy, and inversely related to the short-term interest rate. This is somewhat consistent with the literature on predictability of stock index returns. γ_2 is inversely related to the January dummy. And γ_3 is positively related to the default risk yield spread and the January dummy, and inversely related to the short-term interest rate and the dividend yield.

VIII. Conclusions

We analyze a new class of linear factor models in which the factors are latent and the covariance matrix of returns follows an MSV process. Our interest in this class of models is twofold. First, we are interested in determining which features of the model provide the best fit of the data. We examine models with different

[30]We omit these redundant plots to conserve space.

TABLE 10

Factor Loadings for a Model with Time-Varying Factor Risk Premia

Table 10 reports summary statistics describing the Bayesian posterior densities of factor loadings for the APT-constrained three-factor MSV model with time-varying factor risk premia (MSV3f†). The data are monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). Priors on the parameters are uninformative. The model was estimated using Markov chain Monte Carlo (MCMC) techniques. The posterior mean, standard deviation, and 5th and 95th percentile critical values for each parameter are based on 10,000 Gibbs sampler iterations from a suitably constructed Markov chain.

	Loadings on f_1 (B_{i1})				Loadings on f_2 (B_{i2})				Loadings on f_3 (B_{i3})			
Port.	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Cap1	1.050				−1				1			
Cap2	1.038	0.011	1.020	1.055	−0.718	0.032	−0.770	−0.665	0.637	0.028	0.592	0.684
Cap3	1.027	0.009	1.013	1.041	−0.545	0.028	−0.590	−0.499	0.404	0.024	0.366	0.443
Cap4	1.016	0.009	1.002	1.030	−0.389	0.026	−0.431	−0.347	0.223	0.021	0.188	0.257
Cap5	1.012	0.008	1.000	1.025	−0.212	0.025	−0.253	−0.171	0.085	0.018	0.056	0.113
Cap6	1				0				0			
Cap7	0.968	0.007	0.956	0.980	0.222	0.023	0.184	0.261	−0.066	0.018	−0.096	−0.036
Cap8	0.915	0.008	0.903	0.928	0.459	0.024	0.421	0.499	−0.071	0.017	−0.099	−0.043
Cap9	0.843	0.008	0.830	0.856	0.731	0.027	0.688	0.777	−0.044	0.020	−0.076	−0.011
Cap10	0.650				1				0			

TABLE 11

Stochastic Volatility Processes for a Model with Time-Varying Factor Risk Premia

Table 11 reports summary statistics describing the Bayesian posterior densities of parameters for the stochastic volatility processes (equation (6)) for the APT-constrained three-factor MSV model with time-varying risk premia (MSV3f†). The data are monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). Priors on parameters are uninformative. The model was estimated using Markov chain Monte Carlo (MCMC) techniques. The posterior mean, standard deviation, and 5th and 95th percentile critical values for each parameter are based on 10,000 Gibbs sampler iterations from a suitably constructed Markov chain.

	$\exp(\kappa_j/2)$ (% per month)				ϕ_j				σ_j			
Factor/ Port.	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
f_1	4.637	0.469	3.922	5.401	0.935	0.030	0.881	0.974	0.259	0.060	0.168	0.364
f_2	1.702	0.268	1.301	2.152	0.968	0.018	0.935	0.990	0.187	0.051	0.115	0.278
f_3	1.831	0.298	1.380	2.332	0.927	0.038	0.855	0.977	0.439	0.115	0.269	0.649
Cap1	1.303	0.159	1.057	1.571	0.890	0.045	0.807	0.953	0.507	0.108	0.343	0.695
Cap2	0.814	0.049	0.736	0.895	0.825	0.114	0.605	0.967	0.154	0.082	0.063	0.315
Cap3	0.794	0.064	0.695	0.898	0.932	0.043	0.847	0.981	0.183	0.063	0.100	0.300
Cap4	0.748	0.062	0.650	0.853	0.932	0.033	0.870	0.975	0.207	0.052	0.130	0.300
Cap5	0.689	0.028	0.645	0.736	0.797	0.108	0.591	0.933	0.117	0.050	0.053	0.212
Cap6	0.620	0.043	0.555	0.689	0.893	0.064	0.768	0.966	0.220	0.077	0.111	0.359
Cap7	0.643	0.100	0.498	0.814	0.977	0.015	0.948	0.993	0.136	0.038	0.085	0.208
Cap8	0.576	0.068	0.485	0.678	0.954	0.025	0.908	0.986	0.171	0.047	0.104	0.256
Cap9	0.467	0.042	0.412	0.525	0.869	0.107	0.655	0.984	0.114	0.048	0.054	0.207
Cap10	1.379	0.194	1.092	1.703	0.960	0.023	0.918	0.987	0.220	0.060	0.140	0.330

numbers of factors, models with different specifications for the time-varying covariance matrix of returns, and models with predictable returns. Second, we are interested in whether pricing restrictions implied by the equilibrium APT are supported by the data. We compare unconstrained models to models constrained by APT pricing restrictions. We also examine the implications of the APT regarding return predictability. If predictable returns are due to time-varying factor risk premia, then asset pricing theory suggests that predictability should be related to exposures to systematic risk factors. We compare models with unconstrained return predictability to models in which predictability is proportional to loadings on latent factors.

TABLE 12

Time-Varying Factor Risk Premia

In the APT-constrained model with time-varying expected returns, factor risk premia are linear in predetermined instrumental variables, $\gamma_t = Gz_{t-1}$. Table 12 reports summary statistics describing the Bayesian posterior densities of coefficients from the *G* matrix for the APT-constrained three-factor MSV model with time-varying factor risk premia (MSV3f†). The data are monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization (*Cap*) decile portfolios for the period 1952:1 to 2003:12 (624 monthly observations). Priors on parameters are uninformative. The model was estimated using Markov chain Monte Carlo (MCMC) techniques. The posterior mean, standard deviation, and 5th and 95th percentile critical values for each parameter are based on 10,000 Gibbs sampler iterations from a suitably constructed Markov chain.

	γ_1				γ_2				γ_3			
z	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%	Mean	Std. Dev.	5%	95%
Constant	0.001	0.008	−0.013	0.014	−0.001	0.003	−0.006	0.004	0.012	0.005	0.004	0.019
Term	0.100	0.248	−0.306	0.507	0.017	0.103	−0.153	0.186	−0.113	0.155	−0.367	0.141
Qual	2.248	0.689	1.106	3.374	0.099	0.269	−0.344	0.541	0.969	0.353	0.385	1.549
Jan	0.028	0.007	0.017	0.039	−0.022	0.003	−0.026	−0.017	0.031	0.004	0.025	0.038
Bill	−4.827	1.183	−6.754	−2.886	−0.121	0.481	−0.920	0.666	−1.486	0.648	−2.547	−0.416
Div Yld	0.127	0.195	−0.191	0.452	0.109	0.080	−0.021	0.239	−0.364	0.102	−0.530	−0.197

We examine 50 years of monthly excess returns for 10 NYSE/AMEX/NASDAQ market capitalization decile portfolios. Recent advances in Bayesian MCMC techniques make estimation/comparison of models with latent factors and latent stochastic volatilities feasible. Using Bayes factors, we compare models on a number of dimensions. We find that linear factor models with MSV fit the data far better than models with homoskedastic factor and/or idiosyncratic shocks. Models with three latent factors best explain the common variation in size portfolio excess returns. Although exposure to the first latent factor explains the average level of portfolio expected returns, exposures to the second and third latent factors contribute substantially to explaining the cross section of portfolio expected returns. In every case we examine, the data strongly favor models constrained by APT pricing restrictions over otherwise identical unconstrained models. We also find support for APT pricing restrictions in models with time-varying expected returns. As suggested by APT, the data strongly favor models in which predictability is due to time-varying factor risk premia over models in which predictability is unrelated to factor loadings. Overall, we find no evidence against the APT's central prediction that expected returns should be related to loadings on latent factors that explain common variation in asset returns. All conclusions are very robust to the choice of priors.

Appendix A. MCMC Algorithm

In this appendix, we discuss the estimation of the constrained MSV model. Since estimation of the unconstrained model requires only a minor modification to our procedure, we will omit its description for the sake of brevity.[31] Similarly, since the SVE, SVF, and CF models are special cases of the MSV structure, their respective MCMC estimation algorithms do not present any additional complication and are not described in this appendix. Complete details for the omitted algorithms are available from the authors.

Note that it is the stochastic volatilities themselves (h_t), and not the parameters of the SV processes (θ), that appear in the likelihood function (11). Marginalizing over f_t, the conditional sampling density of y_t can be written

$$y_t | h_t, \gamma, B \quad \sim \quad \mathcal{N}_N(B\gamma, \Omega_t). \tag{14}$$

[31] The parameter vector for the unconstrained model is $\psi = (\mu, \beta, \theta_1, \ldots, \theta_{N+K})$.

With this marginalization, the posterior distribution of β is not of a known form. A Metropolis-Hastings (M-H) algorithm is, thus, necessary to carry out the sampling.[32] Chib, Nardari, and Shephard (2002) show that integrating out the factors and generating β by M-H is essential to produce draws that efficiently converge to the target distribution.

Conditioning on β, the sampling of γ and the latent factors is straightforward as the posterior updates are given by standard Bayesian results for multivariate regression analysis (see, for example, Zellner (1971)).

Next, given β, γ, and f, each idiosyncratic and, respectively, factor shock can be represented by

$$\begin{aligned} y_{jt} - B_j\gamma - B_j f_t &= \varepsilon_{jt}\exp(h_{jt}/2), \quad j \leq N \\ f_{jt} &= \varepsilon_{jt}\exp(h_{jt}/2), \quad N+1 \leq j \leq N+k, \end{aligned}$$

where $\varepsilon_{jt} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. We exploit a clever change of variable suggested by Kim, Shephard, and Chib (1998). Let

$$z_{jt} = \begin{cases} \ln(y_{jt} - B_j\gamma - B_j f_t)^2 = h_{jt} + \ln(\varepsilon_{jt}^2), & j \leq N \\ \ln(f_{jt})^2 = h_{jt} + \ln(\varepsilon_{jt}^2), & N+1 \leq j \leq N+k \end{cases}.$$

Note that z_{jt} is the sum of h_{jt} and a log χ^2 random variable with one degree of freedom. Kim, Shephard, and Chib (1998) show that the density of a log χ^2 random variable can be approximated quite precisely with a seven-component mixture of Gaussian distributions. This change of variable permits us to represent the MSV model as $N + K$ independent conditionally Gaussian state space models:

$$z_{jt}|s_{jt}, h_{jt} \sim \mathcal{N}(h_{jt} + m_{s_{jt}}, \upsilon^2_{s_{jt}}) \quad \text{and} \tag{15}$$

$$h_{jt} = \kappa_j + \phi_j(h_{jt-1} - \kappa_j) + \sigma_j\eta_{jt}, \quad j \leq N+K, \tag{16}$$

where s_{jt} is a discrete component indicator with mass function $\Pr(s_{jt}) = q_i, i \leq 7, t \leq T$, and $m_{s_{jt}}$, $\upsilon^2_{s_{jt}}$, and q_i are the parameters of each component tabulated in Kim, Shephard, and Chib (1998). Given the factors, the factor risk premia, and the free elements of B, one can sample θ_j and the log-variances separately for each portfolio and factor series using MCMC methods developed for univariate SV models.[33] Let $\{h_{j.}\}$ denote the set of T-vectors $h_{j.} = (h_{j1}, \ldots, h_{jT})$. For the APT-constrained model with constant factor risk premia, the steps of the Gibbs sampler are summarized below.

MCMC Algorithm for Constrained MSV Model with Constant Risk Premia

1. Initialize $\{h_{j.}\}$ and γ.
2. Sample $\beta, \gamma, \{f_t\}|y, \{h_{j.}\}$ by drawing:
   (a) β from $\beta|y, \gamma, \{h_{j.}\}$
   (b) γ from $\gamma|y, \beta, \{h_{j.}\}$
   (c) f_t from $f_t|y_t, B, \gamma, \{h_{j.}\}, t \leq T$.
3. Compute $\{z_{jt}\}$ for $t \leq T$ and $j \leq N+K$.
4. Sample $s_{j.}$, θ_j, and $\{h_{j.}\}$ by repeating the following steps for $j \leq N+K$:
   (a) Draw $s_{j.}$ from $s_{j.}|z_{j.}, h_{j.}$
   (b) Draw θ_j from $\theta_j|z_{j.}, h_{j.}$
   (c) Draw $h_{j.}$ from $h_{j.}|z_{j.}, s_{j.}, \theta_j$.
5. Go to step 2 and repeat.

In our applications, we cycle through steps 2, 3, and 4 for 15,000 iterations. We discard the initial 5,000 "burn-in" draws and retain the remaining draws for inferential purposes.

[32] Chib and Greenberg (1995) present a thorough illustration of the M-H algorithm.

[33] See Kim, Shephard, and Chib (1998) and Chib, Nardari, and Shephard (2006) for details on univariate SV sampling.

For models with time-varying expected returns, the sampling of vectors γ (for the constrained model) and μ (for the unconstrained model) is replaced by the sampling of coefficient matrices G and M. Recall that $\gamma_t = Gz_{t-1}$ in the constrained model and $\mu_t = Mz_{t-1}$ in the unconstrained model, where z_{t-1} is the vector of predetermined instrumental variables. Conditioning on B and f, the row by row sampling of G or M is straightforward since the posterior updates are standard Bayesian results for multivariate regression analysis (see, for example, Zellner (1971)).

Appendix B. Marginal Likelihood Calculation

We follow the approach suggested by Chib, Nardari, and Shephard (2006) for the computation of marginal likelihood estimates. Using the Bayes rule, the marginal likelihood of the data given model $\mathcal{M}_i$ is

$$m(y|\mathcal{M}_i) = \frac{p(y|\mathcal{M}_i, \psi_i^*)\pi(\psi_i^*|\mathcal{M}_i)}{\pi(\psi_i^*|y, \mathcal{M}_i)}, \tag{17}$$

where $p(y|\mathcal{M}_i, \psi_i^*)$ is the likelihood function under $\mathcal{M}_i$, and $\pi(\psi_i^*|\mathcal{M}_i)$ and $\pi(\psi_i^*|y, \mathcal{M}_i)$ are the corresponding prior and posterior densities of the parameters. Each density is evaluated at the parameter vector ψ_i^*.

Using the basic marginal likelihood identity (see Chib (1995)) in (17), the log Bayes factor comparing $\mathcal{M}_i$ to $\mathcal{M}_j$ can be written as

$$\begin{aligned} &\log p(y|\mathcal{M}_i) - \log p(y|\mathcal{M}_j) \\ &\quad = \{\log p(y|\mathcal{M}_i, \psi_i^*) + \log \pi(\psi_i^*|\mathcal{M}_i) - \log \pi(\psi_i^*|y, \mathcal{M}_i)\} \\ &\qquad - \left\{\log p(y|\mathcal{M}_j, \psi_j^*) + \log \pi(\psi_j^*|\mathcal{M}_j) - \log \pi(\psi_j^*|y, \mathcal{M}_j)\right\}. \end{aligned} \tag{18}$$

Since (17) is an identity, the choice of vectors ψ_i^* and ψ_j^* is arbitrary. For computational accuracy, we choose to evaluate these expressions at the posterior means from the MCMC output. For each model being compared, there are two key quantities that must be computed: the posterior ordinate $\pi(\psi^*|y, \mathcal{M})$ and the likelihood ordinate $p(y|\mathcal{M}, \psi^*)$. Because of the high dimensionality of the vector ψ and the presence of the latent factors and volatilities, calculation of $\pi(\psi^*|y, \mathcal{M})$ and the likelihood ordinate $p(y|\mathcal{M}, \psi^*)$ is not analytically tractable. Fortunately, Chib (1995) and Chib and Jeliazkov (2001) provide simulation-based strategies to efficiently estimate the posterior ordinate. Furthermore, building on the work of Pitt and Shephard (1999) and Doucet, de Freitas, and Gordon (2001), Chib, Nardari, and Shephard (2002) develop filtering methods to estimate the likelihood ordinate. Chib, Nardari, and Shephard (2002), (2006) provide ample evidence of the reliability of these approaches for computing marginal likelihood estimates in the context of univariate and multivariate stochastic volatility models.

References

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

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

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

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

Banz, R. W. “The Relationship between Return and Market Value of Common Stocks.” *Journal of Financial Economics*, 9 (1981), 3–18.

Berger, J., and L. R. Pericchi. “The Intrinsic Bayes Factor for Model Selection and Prediction.” *Journal of the American Statistical Association*, 91 (1996), 109–122.

Bossaerts, P., and R. C. Green. "A General Equilibrium Model of Changing Risk Premia: Theory and Tests." *Review of Financial Studies*, 2 (1989), 467–493.

Brown, S. J., and T. Otsuki. "Risk Premia in Pacific-Basin Capital Markets." *Pacific-Basin Finance Journal*, 1 (1993), 235–261.

Brown, S. J., and M. I. Weinstein. "A New Approach to Asset Pricing Models: The Bilinear Paradigm." *Journal of Finance*, 38 (1983), 711–743.

Burmeister, E., and M. B. McElroy. "Joint Estimation of Factor Sensitivities and Risk Premia for the Arbitrage Pricing Theory." *Journal of Finance*, 43 (1988), 721–733.

Campbell, J. Y. "Stock Returns and the Term Structure." *Journal of Financial Economics*, 18 (1987), 373–399.

Chib, S. "Marginal Likelihood From the Gibbs Output." *Journal of the American Statistical Association*, 90 (1995), 1313–1321.

Chib, S., and E. Greenberg. "Understanding the Metropolis-Hastings Algorithm." *American Statistician*, 49 (1995), 327–335.

Chib, S., and I. Jeliazkov. "Marginal Likelihood from the Metropolis-Hastings Output." *Journal of the American Statistical Association*, 96 (2001), 270–281.

Chib, S.; F. Nardari; and N. Shephard. "Markov Chain Monte Carlo Methods for Stochastic Volatility Models." *Journal of Econometrics*, 108 (2002), 281–316.

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

Connolly, R. A. "A Posterior Odds Analysis of the Weekend Effect." *Journal of Econometrics*, 49 (1991), 51–104.

Connor, G. "A Unified Beta Pricing Theory." *Journal of Economic Theory*, 34 (1984), 13–31.

Connor, G., and R. A. Korajczyk. "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." *Journal of Financial Economics*, 15 (1986), 373–394.

Connor, G., and R. A. Korajczyk. "Risk and Return in an Equilibrium APT: Application of a New Test Methodology." *Journal of Financial Economics*, 21 (1988), 255–289.

Connor, G., and R. A. Korajczyk. "An Intertemporal Equilibrium Beta Pricing Model." *Review of Financial Studies*, 2 (1989), 373–392.

Connor, G., and R. A. Korajczyk. "The Arbitrage Pricing Theory and Multifactor Models of Asset Returns." In *Finance,* R. A. Jarrow, V. Maksimovic, W. T. Ziemba, eds., Vol. 9, *Handbooks in Operations Research and Management Science*. Amsterdam: North-Holland (1995).

Doucet, A.; N. de Freitas; and N. Gordon. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag (2001).

Engle, R. F.; V. K. Ng; and M. Rothschild. "Asset Pricing with a Factor-ARCH Covariance Structure: Empirical Estimates for Treasury Bills." *Journal of Econometrics*, 45 (1990), 213–237.

Evans, M. D. D. "Expected Returns, Time-Varying Risk, and Risk Premia." *Journal of Finance*, 49 (1994), 655–679.

Fama, E. F., and K. R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 3–25.

Fama, E. F., and K. R. French. "Business Conditions and Expected Returns on Stocks and Bonds." *Journal of Financial Economics*, 25 (1989), 23–49.

Fama, E. F., and G. W. Schwert. "Asset Returns and Inflation." *Journal of Financial Economics*, 5 (1977), 115–146.

Ferson, W. E. "Changes in Expected Security Returns, Risk, and the Level of Interest Rates." *Journal of Finance*, 44 (1989), 1191–1217.

Ferson, W. E., and C. R. Harvey. "The Variation of Economic Risk Premiums." *Journal of Political Economy*, 99 (1991), 385–415.

Ferson, W. E., and R. A. Korajczyk. "Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?" *Journal of Business*, 68 (1995), 309–349.

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

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

Ghysels, E.; A. C. Harvey; and E. Renault. "Stochastic Volatility." In *Handbook of Statistics,* G. S. Maddala, C. R. Rao, eds., Vol. 14. Amsterdam: North-Holland (1996).

Gibbons, M. R., and W. E. Ferson. "Testing Asset Pricing Models with Changing Expectations and an Unobservable Market Portfolio." *Journal of Financial Economics*, 14 (1985), 217–236.

Gibbons, M. R.; S. A. Ross; and J. Shanken. "A Test of the Efficiency of a Given Portfolio." *Econometrica*, 57 (1989), 1121–1152.

Harvey, C. R. "Time-Varying Conditional Covariances in Tests of Asset Pricing Models." *Journal of Financial Economics*, 24 (1989), 289–317.

Harvey, C. R., and G. Zhou. "Bayesian Inference in Asset Pricing Tests." *Journal of Financial Economics*, 26 (1990), 221–254.
Jacquier, E.; N. G. Polson; and P. E. Rossi. "Bayesian Analysis of Stochastic Volatility Models (with discussion)." *Journal of the American Statistical Association*, 12 (1994), 371–417.
Jones, C. S. "Extracting Factors from Heteroskedastic Asset Returns." *Journal of Financial Economics*, 62 (2001), 293–325.
Kass, R. E., and A. E. Raftery. "Bayes Factors." *Journal of the American Statistical Association*, 90 (1995), 773–795.
Keim, D. B. "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence." *Journal of Financial Economics*, 12 (1983), 13–32.
Kim, S.; N. Shephard; and S. Chib. "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." *Review of Economic Studies*, 65 (1998), 361–393.
King, M.; E. Sentana; and S. Wadhwani. "Volatility and Links between National Stock Markets." *Econometrica*, 62 (1994), 901–933.
Kirby, C. "The Restrictions on Predictability Implied by Rational Asset Pricing Models." *Review of Financial Studies*, 11 (1998), 343–382.
Klein, R. W., and S. J. Brown. "Model Selection When There is 'Minimal' Prior Information." *Econometrica*, 52 (1984), 1291–1312.
Lehmann, B. N., and D. M. Modest. "The Empirical Foundations of the Arbitrage Pricing Theory." *Journal of Financial Economics*, 21 (1988), 213–254.
McCulloch, R., and P. E. Rossi. "A Bayesian Approach to Testing the Arbitrage Pricing Theory." *Journal of Econometrics*, 49 (1991), 141–168.
McElroy, M. B., and E. Burmeister. "Arbitrage Pricing Theory as a Restricted Nonlinear Multivariate Regression Model: ITNLSUR Estimates." *Journal of Business and Economic Statistics*, 6 (1988), 29–42.
Moskowitz, T. J. "An Analysis of Covariance Risk and Pricing Anomalies." *Review of Financial Studies*, 16 (2003), 417–457.
Ng, V. K.; R. F. Engle; and M. Rothschild. "A Multi-Dynamic-Factor Model for Stock Returns." *Journal of Econometrics*, 52 (1992), 245–266.
O'Hagan, A. *Bayesian Inference, Kendall's Advanced Theory of Statistics: Vol. 2B*. London: Edward Arnold (1994).
Pitt, M. K., and N. Shephard. "Filtering via Simulation: Auxiliary Particle Filters." *Journal of the American Statistical Association*, 94 (1999), 590–599.
Reinganum, M. R. "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yields and Market Values." *Journal of Financial Economics*, 9 (1981), 19–46.
Roll, R., and S. A. Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance*, 35 (1980), 1073–1103.
Ross, S. A. "The Arbitrage Pricing Theory of Capital Asset Pricing." *Journal of Economic Theory*, 13 (1976), 341–360.
Ross, S. A. "Return, Risk and Arbitrage." In *Risk and Return in Finance,* I. Friend, J. L. Bicksler, eds. Cambridge, MA: Ballinger (1977).
Schwert, G. W. "Anomalies and Market Efficiency." NBER Working Paper 9277 (2002).
Shanken, J. "A Bayesian Approach to Testing Portfolio Efficiency." *Journal of Financial Economics*, 19 (1987), 195–215.
Shanken, J. "Intertemporal Asset Pricing: An Empirical Investigation." *Journal of Econometrics*, 45 (1990), 99–120.
Sharpe, W. F. "A Simplified Model For Portfolio Analysis." *Management Science*, 99 (1963), 277–293.
Shiller, R. J. "Stock Prices and Social Dynamics." *Brookings Papers on Economic Activity*, 2 (1984), 457–498.
Taylor, S. J. "Modeling Stochastic Volatility: A Review and Comparative Study." *Mathematical Finance*, 4 (1994), 183–204.
Zellner, A. *An Introduction to Bayesian Inference in Econometrics*. New York: J. Wiley and Sons, Inc. (1971).
Zhou, G. "Analytical GMM Tests: Asset Pricing with Time-Varying Risk Premiums." *Review of Financial Studies*, 7 (1994), 687–709.