

Why Does Stock Market Volatility Change Over Time?  
A Time-Varying Variance Decomposition for Stock Returns\*

Federico Nardari<sup>†</sup>

John T. Scruggs<sup>‡</sup>

Department of Finance

Department of Banking and Finance

W. P. Carey School of Business

Terry College of Business

Arizona State University

University of Georgia

April 25, 2005

\*Scruggs acknowledges financial support from a Terry-Sanford Summer Research Award.

<sup>†</sup>Correspondence: Federico Nardari, Main Campus, PO Box 873906, Tempe, AZ 85287-3906, tel: (480) 965-7961, fax: (480) 965-8539, e-mail: Federico.Nardari@asu.edu.

<sup>‡</sup>Correspondence: John T. Scruggs, Department of Banking and Finance, Terry College of Business, University of Georgia, Brooks Hall, Athens, GA, 30602-6253, tel: (706) 542-3649, fax: (706) 542-9434, e-mail: jscruggs@terry.uga.edu.

# Why Does Stock Market Volatility Change Over Time?

## A Time-Varying Variance Decomposition for Stock Returns

### Abstract

We extend the variance decomposition model of Campbell (1991) to allow for time-varying stock market volatility. Specifically, we introduce a model in which the covariance matrix of the predictive vector autoregression (VAR) follows a multivariate stochastic volatility (MSV) process. This new VAR-MSV model permits the decomposition of unexpected real stock return variance into three time-varying components: variance of “news about future dividends,” variance of “news about future returns,” and a covariance term. We develop Bayesian Markov chain Monte Carlo (MCMC) econometric techniques for estimating the VAR-MSV model. These methods are well-suited for estimating models with latent stochastic volatilities, and are not subject to the small-sample biases and unit root problems that plague frequentist estimation of predictive regressions. We report strong evidence that real stock returns are predictable when the dividend-price ratio and a stochastically detrended short-term interest rate are employed as forecasting variables. The time-varying variance of “news about future returns” is the primary determinant of stock market volatility (both levels and changes). The variance of “news about future dividends” increased dramatically during the 1973–1974 recession and peaked during the 1980 recession before descending in the 1980s. However, its contribution to stock market volatility was offset by positive correlation between “news about future dividends” and “news about future returns” from 1974–1984.

Key Words: Variance decomposition; Return predictability; Vector autoregression; Multivariate stochastic volatility; Markov chain Monte Carlo; Gibbs sampling.

JEL Classification: G12 (Asset Pricing); C11 (Bayesian Analysis); C15 (Statistical Simulation Methods; Monte Carlo Methods); C32 (Multiple Equation Time Series Models).

Campbell (1991) asks the fundamental question: What moves the stock market? Intuition derived from a dynamic dividend growth model suggests that stock price movements are caused by either changes in rational expectations of future dividends (i.e., “news about future dividends”) or changes in rational expectations of future returns (i.e., “news about future returns”). Campbell (1991) captures this intuition in a clever model where expected returns are time-varying. Campbell’s model decomposes the variance of unexpected real stock returns into components associated with uncertainty about future dividends and uncertainty about future real returns. The model is composed of: (1) an asset pricing framework based on a log-linear approximation of a dynamic dividend growth model, and (2) a vector autoregression (VAR) for forecasting stock returns. Campbell finds that “the variability and persistence of expected stock returns account for a considerable degree of volatility in unexpected stock returns.” He finds that the variance of news about future dividends accounts for only one-third to one-half of the variance of unexpected stock returns.

In this paper we ask a different, but closely related, question: Why does stock market volatility change over time? It is well-established that stock market volatility is time-varying. This fact has important implications for asset pricing models, portfolio allocation decisions, risk management, and derivative security pricing. Previous research on the question has examined both microstructure explanations (e.g., trading volume, margin requirements, bid-ask spreads, etc.) and macroeconomic explanations (e.g., levels and volatilities of macroeconomic variables).<sup>1</sup> A well-known example is Schwert (1989), which reports that stock market volatility is higher during recessions, but is puzzled to find only weak evidence that it is associated with the time-varying volatility of macroeconomic variables (e.g., inflation, monetary growth, industrial production).

Motivated by Campbell’s approach, this paper is the first to explore whether time-varying stock market volatility is driven by uncertainty about future dividends or uncertainty about future real returns.<sup>2</sup> We introduce an extension of the Campbell (1991) model that allows the covariance matrix of the VAR to vary over time. Specifically, we assume that the covariance matrix of the VAR follows a multivariate stochastic volatility (MSV) process. The VAR-MSV model enables the decomposition of unexpected real stock return variance into three time-varying components: variance of “news about future returns,” variance of “news about future dividends,” and a covariance term. This time-varying variance decomposition allows us to investigate how the variance components vary

---

<sup>1</sup>Bollerslev, Chou, and Kroner (1992) reviews some of this research.

<sup>2</sup>Using intuition similar to Campbell’s, Schwert (1989) suggests (p. 1116) that “If macroeconomic data provide information about the volatility of either future expected cash flows or future discount rates, they can help explain why stock return volatility changes over time.”

over time, and how they contribute (in levels and changes) to the variance of unexpected returns.

We find that stock market volatility changes over time primarily because of the time-varying volatility of “news about future returns.” A time-series plot of the variance of “news about future returns” closely tracks (both in levels and in changes) the variance of unexpected real returns for our post-war sample period (1952:1–2002:12). The share of return variance attributable to “news about future returns” ranges from 0.71 to 1.20, but is typically less than one (the time-series average is 0.84). Many, but not all, episodes of high variance of “news about future returns” are associated with recessions. The other two components of stock market volatility play secondary roles. The variance of “news about future dividends” is typically much smaller than the variance of “news about future returns.” However, it did increase dramatically during the recessions of 1973–1975, 1980, and 1981–1982. The share of return variance attributable to “news about future dividends” ranges from 0.10 to 0.61 (time-series average is 0.24) in the post-war period, but is always less than the share attributable to “news about future returns.” When the variance of “news about future dividends” is at its highest (i.e., 1974–1983), its contribution to return variance is tempered by a covariance term that implies positive correlation between “news about future dividends” and “news about future returns.” When this correlation is positive (negative), return variance is less (more) than the sum of the variances of “news about future dividends” and “news about future returns.” The share of return variance attributable to this covariance term ranges from 0.21 (when the correlation is negative) to  $-0.79$  (during the 1980 recession when the correlation is positive), but is typically in the neighborhood of zero.

We find that monthly real stock returns are predictable using the dividend-price ratio and the stochastically detrended short-term interest rate. Since the predictive regressors (especially the dividend-price ratio) are highly autocorrelated, it follows that expected real stock returns are highly persistent. When expectations regarding future returns are revised, they are revised for many periods in the future. Since “news about future returns” is the discounted sum of such revisions in expectations, a small innovation in the expected return can cause a relatively large stock price movement. Likewise, a small increase in the volatility of expected return innovations can contribute to a relatively large increase in stock market volatility. Such changes in the variance of “news about future returns” are the primary reason that stock market volatility changes over time.

We employ Bayesian Markov chain Monte Carlo (MCMC) methods to analyze the model. These

methods are well-suited for making inferences regarding return predictability, and this paper makes an important contribution in that area. Frequentist econometric methods (e.g., OLS) are problematic for three main reasons. First, OLS estimators are biased in small samples when lagged endogenous regressors are employed as predictive variables (see Stambaugh (1999)). Bayesian MCMC methods, in contrast, provide exact small-sample posterior densities for the model’s parameters and other functions of interest. Second, the correct asymptotic distribution theory is unclear if a predictive variable is nearly integrated or exhibits a unit root (see Lewellen (2004), Campbell and Yogo (2005) and Torous, Valkanov, and Yan (2005)). Inferences based on Bayesian posterior densities are valid whether or not the predictive variable has a unit root (see Sims (1988) and Sims and Uhlig (1991)). Third, the OLS estimator bias corrections derived in Stambaugh (1999) and Lewellen (2004) apply to univariate predictive regressions. Stambaugh (1999) finds that extending these approaches to the VAR framework with multiple predictive regressors is not straightforward. In contrast, Stambaugh shows that the Bayesian approach can be readily extended to VAR predictive regressions. We implement just such an extension.<sup>3</sup> The Bayesian approach has other attractive features. Models with latent variables (stochastic volatilities in this case) are amenable to estimation using MCMC methods. Using the Bayesian approach, the parameter space is augmented with the stochastic volatilities and the Gibbs sampler algorithm effectively “integrates out” these nuisance parameters. The model parameters and latent variables are estimated simultaneously, so posterior densities implicitly incorporate estimation error (i.e., parameter uncertainty). Our Bayesian MCMC estimation methodology for estimating SV models builds on techniques developed in Chib, Nardari and Shephard (2002, 2005).

The VAR approach to forecasting asset returns is used in many strands of the asset pricing literature. Following Campbell (1991), many researchers employ a VAR framework coupled with a log-linear approximation of the present value relation to decompose unexpected asset returns. Papers that employ this approach include Campbell (1993), Campbell and Ammer (1993), Campbell (1996), Ammer and Mei (1996), Lamont (1998), Lamont and Polk (2001), Hollifield, Koop, and Li (2003), and Campbell and Vuolteenaho (2004). Hodrick (1992) and Stambaugh (1999) examine the statistical properties of predictive regressions (including VARs) using lagged endogenous regressors. Kandel and Stambaugh (1996), Barberis (2000), and Shanken and Tamayo (2004) use a VAR framework to examine the sensitivity of asset allocations to evidence about return predictability.

---

<sup>3</sup>In related work, Hollifield, Koop, and Li (2003) employ Bayesian MCMC methods to estimate VAR predictive regressions where the VAR errors are homoskedastic.

The VAR-MSV model introduced in this paper has potential applications in all of these strands of the literature.

The remainder of the paper is organized as follows. Section I derives the VAR-MSV model. Section II discusses our Bayesian MCMC estimation methodology. The more technical aspects are included in the appendices. The data are described in Section III. We discuss our empirical results in Section IV. Section V concludes and suggests directions for future research.

## I The Model

### A Decomposing Unexpected Stock Returns

We begin by decomposing unexpected real stock returns into components related to changes in rational expectations of future real dividends and future real stock returns. We employ a model derived by Campbell (1991). The model is based on a clever log-linear approximation of a dynamic accounting identity.<sup>4</sup>

Let  $h_{t+1}$  denote the log real return on a stock held from times  $t$  to  $t + 1$ , and let  $d_{t+1}$  denote the log real dividend paid during period  $t + 1$ . Campbell (1991) shows that unexpected real stock returns can be decomposed into components related to changes in expectations regarding future dividends and changes in expectations regarding future real returns:

$$h_{t+1} - E_t h_{t+1} = (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j h_{t+1+j}. \quad (1)$$

The parameter  $\rho$  is the average ratio of the stock price to the sum of the stock price and the dividend. It should be slightly less than one. In our empirical work, we set  $\rho = 0.997$  for the 1952:1–2002:12 sample of monthly data. It is clear from this model that an unexpected positive real return shock (i.e., good news) must be caused by either an upward revision in expected dividend growth or a downward revision in expected future real returns.

Let  $\eta_{t+1} \equiv h_{t+1} - E_t h_{t+1}$  define the unexpected real stock return in period  $t + 1$ . For notational

---

<sup>4</sup>Campbell derives the equation by taking a first-order Taylor series approximation of the present value relation for a dividend-paying stock. The approximate equation is solved forward, imposing the terminal condition that the log dividend-price ratio does not follow an explosive process. The log-linear approximation technique was originally proposed in Campbell and Shiller (1988a, b).

convenience, we express (1) more simply as:

$$\eta_{t+1} = \eta_{d,t+1} - \eta_{h,t+1}, \quad (2)$$

where  $\eta_{d,t+1}$  represents “news about future dividends” and  $\eta_{h,t+1}$  represents “news about future real returns.”

## B VAR Approach to Estimating Unobservable Components

Following Campbell (1991), we employ a VAR model to forecast future real stock returns. Let  $\mathbf{z}_{t+1}$  denote a  $k \times 1$  state vector of variables known at the end of period  $t + 1$ . The real stock return  $h_{t+1}$  is the first element in  $\mathbf{z}_{t+1}$ . The remaining elements of  $\mathbf{z}_{t+1}$  are forecasting variables. We assume that the state vector  $\mathbf{z}_{t+1}$  follows a first-order VAR process:

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{w}_{t+1} \quad (3)$$

where  $\mathbf{z}_t$  is the lagged state vector,  $\mathbf{A}$  is a conforming coefficient matrix, and  $\mathbf{w}_{t+1}$  is the error vector. In the homoskedastic VAR model employed by Campbell (1991), the error vector is assumed to be distributed multivariate normal,  $\mathbf{w}_{t+1} \sim \mathcal{N}_k(\mathbf{0}, \Sigma)$ .

Multi-period forecasts of the state vector are obtained by matrix multiplication,  $E_t \mathbf{z}_{t+j} = \mathbf{A}^j \mathbf{z}_t$ . Let  $\mathbf{e1}$  denote a  $k \times 1$  vector with a 1 in the first row and zeroes in the others.  $\mathbf{e1}$  is designed to pick the first element from a state or error vector. For example,  $\mathbf{e1}'\mathbf{z}_{t+1} = h_{t+1}$  picks the real stock return from the state vector, and  $\mathbf{e1}'\mathbf{w}_{t+1} = \eta_{t+1} = h_{t+1} - E_t h_{t+1}$  picks the unexpected stock return from the error vector.

Campbell (1991) combines the first-order VAR forecasting model with the previously described asset pricing framework to decompose unexpected real stock returns into “news about future returns” and “news about future dividends.” It follows from (1) that the “news about future returns”

(i.e., the discounted sum of revisions in expected real returns) can be written as

$$\begin{aligned}
\eta_{h,t+1} &= (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j h_{t+1+j} \\
&= \mathbf{e}\mathbf{1}' \sum_{j=1}^{\infty} \rho^j \mathbf{A}^j \mathbf{w}_{t+1} \\
&= \mathbf{e}\mathbf{1}' \rho \mathbf{A} (\mathbf{I} - \rho \mathbf{A})^{-1} \mathbf{w}_{t+1} \\
&= \boldsymbol{\lambda}' \mathbf{w}_{t+1}
\end{aligned} \tag{4}$$

where  $\boldsymbol{\lambda}' = \mathbf{e}\mathbf{1}' \rho \mathbf{A} (\mathbf{I} - \rho \mathbf{A})^{-1}$  is a nonlinear function of the VAR coefficients. Since  $\eta_{t+1} = \mathbf{e}\mathbf{1}' \mathbf{w}_{t+1}$ , it follows from the error decomposition in (2) that

$$\eta_{d,t+1} = (\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}') \mathbf{w}_{t+1}. \tag{5}$$

The vectors  $\boldsymbol{\lambda}'$  and  $(\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}')$  map the VAR innovations into “news about future returns” and “news about future dividends,” respectively.

It should be emphasized that Campbell (1991) doesn’t explicitly model dividend growth or “news about future dividends.” The term  $\eta_{d,t+1}$ , as defined in (5), is residual in nature. Although Campbell refers to this term as “news about future dividends,” it might reasonably include a “noise” component. Campbell doesn’t discuss the possible contribution of irrational mispricing or noise trading (i.e., trading for reasons unrelated to fundamentals) to unexpected stock returns. It’s not clear whether predictable variation in real stock returns tracks rational variation in the market risk premium or is due to predictable corrections of irrational mispricing (e.g., mean reversion). The model in (1) is based on a dynamic accounting identity, not an economic model of rational investor behavior. Distinguishing between rational and irrational stories about return predictability is an important pursuit, but it is not the focus of this paper.<sup>5</sup> In his conclusion, Campbell emphasizes several caveats that apply equally to this paper. In particular, he cautions that the return decomposition “cannot be given an unambiguous structural interpretation.” Since the variables in the VAR are determined simultaneously, it is probably overly simplistic to conclude that “news about future returns” and “news about future dividends” determine unexpected stock returns.

---

<sup>5</sup>Shanken and Tamayo (2004) investigates the relative roles of played by risk and mispricing in explaining stock return predictability.



## C Multivariate Stochastic Volatility

Campbell (1991) assumes that the error vector in (3) is homoskedastic (i.e.,  $\mathbf{w}_{t+1} \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma})$ ). A distinctive feature of our model is the multivariate stochastic volatility (MSV) specification for the covariance matrix  $\boldsymbol{\Sigma}_{t+1}$ . In extending the VAR model to the heteroskedastic case, we would like a specification for  $\boldsymbol{\Sigma}_{t+1}$  that is sufficiently rich to allow for non-zero time-varying correlations between VAR errors. However, our methodology for estimating stochastic volatility processes, discussed Section II below, assumes that the errors are orthogonal. Thus, we need to transform the covariance matrix  $\boldsymbol{\Sigma}_{t+1}$ . This is not an obstacle. Because  $\boldsymbol{\Sigma}_{t+1}$  is a positive definite symmetric matrix, there exists a unique triangular factorization  $\boldsymbol{\Sigma}_{t+1} = \mathbf{B}\mathbf{V}_{t+1}\mathbf{B}'$  where  $\mathbf{B}$  is a  $k \times k$  lower triangular matrix with unit elements on the main diagonal,

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ b_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & 1 \end{pmatrix}, \quad (6)$$

and  $\mathbf{V}_{t+1}$  is a  $k \times k$  diagonal matrix with positive elements on the main diagonal.<sup>6</sup> With this transformation, the VAR system in (3) can be rewritten

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{V}_{t+1}^{1/2}\boldsymbol{\epsilon}_{t+1} \quad (7)$$

where  $\boldsymbol{\epsilon}_{t+1} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$  and  $\mathbf{V}_{t+1}$  is a diagonal matrix of stochastic variances. With this specification, the VAR errors are permitted to have non-zero, time-varying contemporaneous correlations.

Let  $v_{j,t+1}$  denote the log of the  $j$ th element of  $\mathbf{V}_{t+1}$ , such that

$$\mathbf{V}_{t+1} = \text{diag}\{\exp(v_{1,t+1}), \dots, \exp(v_{k,t+1})\}. \quad (8)$$

We assume that each of the log-variances in (8) follows an independent autoregressive (AR) process,

$$v_{j,t+1} = \mu_j + \phi_j(v_{j,t} - \mu_j) + \sigma_j\zeta_{j,t+1}, \quad (9)$$

where  $\zeta_{j,t+1} \sim \mathcal{N}(0, 1)$  and  $E[\boldsymbol{\zeta}_{t+1}\boldsymbol{\zeta}'_{t+1}] = \mathbf{I}_k$ . Stochastic volatility (SV) models such as this offer an

---

<sup>6</sup>Section 4.4 of Hamilton (1994) discusses the triangular factorization.

attractive alternative to models from the GARCH family. The chief difference between a GARCH model and (9) is the variance innovation term  $\sigma_j \zeta_{jt+1}$  in the SV model. The variance innovation in a GARCH model is typically the lagged squared innovation from the return process. SV models permit innovations in both the return and variance processes. Taylor (1994) and Ghysels, Harvey, and Renault (1996) review SV models such as (9). Articles that develop econometric techniques for estimating SV models include Taylor (1986), Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998), and Chib, Nardari and Shephard (2002, 2005).

## D Variance Decomposition

From (1), the variance of unexpected real returns can be decomposed into three terms,

$$\begin{aligned} \text{Var}(\eta_{t+1}) &= \text{Var}(\eta_{d,t+1} - \eta_{h,t+1}) \\ &= \text{Var}(\eta_{d,t+1}) - 2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1}) + \text{Var}(\eta_{h,t+1}). \end{aligned} \quad (10)$$

Uncertainty about returns in period  $t + 1$  can be expressed as

$$\text{Var}(\eta_{h,t+1}) = \boldsymbol{\lambda}' \boldsymbol{\Sigma}_{t+1} \boldsymbol{\lambda}. \quad (11)$$

Likewise, uncertainty about dividends in period  $t + 1$  can be expressed as

$$\text{Var}(\eta_{d,t+1}) = (\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}') \boldsymbol{\Sigma}_{t+1} (\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}'). \quad (12)$$

And, the covariance is

$$\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1}) = (\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}') \boldsymbol{\Sigma}_{t+1} \boldsymbol{\lambda}. \quad (13)$$

Campbell (1991) reports ratios of the unconditional versions of these three terms to the unconditional variance of real returns. We plot corresponding time-varying ratios involving these three terms. The denominator of each ratio is the variance of the unexpected real return,  $\text{Var}(\eta_{t+1}) = \boldsymbol{\Sigma}_{11,t+1} = \mathbf{e}\mathbf{1}' \boldsymbol{\Sigma}_{t+1} \mathbf{e}\mathbf{1}$ . By construction, these three ratios sum to one at each point in time.

The difference between the shares of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{d,t+1})$  and  $\text{Var}(\eta_{h,t+1})$  is an

intertemporal constant. Given the assumption that  $\Sigma_{t+1} = \mathbf{B}\mathbf{V}_{t+1}\mathbf{B}'$ , this is a mechanical result:

$$\begin{aligned} \frac{\text{Var}(\eta_{d,t+1}) - \text{Var}(\eta_{h,t+1})}{\text{Var}(\eta_{t+1})} &= \frac{(\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}')\Sigma_{t+1}(\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}')' - \boldsymbol{\lambda}'\Sigma_{t+1}\boldsymbol{\lambda}}{\mathbf{e}\mathbf{1}'\Sigma_{t+1}\mathbf{e}\mathbf{1}} \\ &= 1 + 2\lambda_1 + 2\lambda_2\mathbf{B}_{21} + 2\lambda_3\mathbf{B}_{31}. \end{aligned} \quad (14)$$

$\text{Var}(\eta_{t+1})$  and  $\text{Var}(\eta_{h,t+1})$  are estimated directly. Recall from equations (2) and (5) that that  $\eta_d$  is residual in nature. It follows that  $\text{Var}(\eta_{d,t+1})$  and  $\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$  are also somewhat residual in nature, and are estimated indirectly conditional on the variance decomposition in (10).

## II Bayesian Estimation

OLS estimation of VAR predictive regressions is problematic for several reasons. Employing lagged endogenous regressors, especially highly autocorrelated ones, can induce finite-sample biases in predictive regressions.<sup>7</sup> This is because return innovations are correlated with innovations in the predictive regressors. Stambaugh (1999) derives the finite-sample properties of the OLS estimator in this case and proposes a correction for the bias. However, the validity of the correction depends critically on the assumed stationarity of the predictive variable (see Lewellen (2004) and Torous, Valkanov, and Yan (2005)). For many well-known predictive variables, this is a heroic assumption. If the predictive variable's order of integration is uncertain, then the asymptotic distribution of the OLS estimator is of non-standard form and traditional inferential tools (e.g.,  $t$ -tests and  $p$ -values) are invalid (see Campbell and Yogo (2005) and Torous, Valkanov, and Yan (2005)). Furthermore, the bias corrections suggested in Stambaugh (1999) and Lewellen (2004) are for univariate predictive regressions. As discussed in Stambaugh (1999), the extension to models with multiple predictive variables is not straightforward.

We employ Bayesian Markov chain Monte Carlo (MCMC) methods to estimate the model. Unlike frequentist methods, Bayesian methods treat the parameters as random variables given a likelihood function and fixed data. The Bayesian approach is natural in this setting because the regressors are stochastic rather than deterministic (i.e., fixed in repeated samples). The Bayesian approach to estimating predictive regressions has many attractive features. Bayesian methods deliver exact finite-sample posterior densities for both parameters and features of interest. Inferences

---

<sup>7</sup>See Mankiw and Shapiro (1986), Nelson and Kim (1993), and Stambaugh (1999) for detailed discussions.

made from posterior densities are valid even if the predictive regressor is nearly-integrated or exhibits a unit root (see Sims (1988) and Sims and Uhlig (1991)). And, as shown in Stambaugh (1999), the Bayesian approach can be readily extended to models with multiple predictive regressors. We implement such an extension for the VAR forecasting model.

Bayesian estimation requires three elements: the data, a likelihood function dictated by the model, and prior densities for the model's parameters. For illustration of the general principle, let  $D$  denote the data,  $p(D|\psi)$  denote the likelihood function, and  $\pi(\psi)$  denote the prior density for the parameter set  $\psi$ . In Bayesian analysis, the object of interest is the joint posterior density of the parameters given the data,  $\pi(\psi|D)$ . Following Bayes rule, the joint posterior density for the parameters is proportional to the product of the likelihood function and the prior density on the parameters:

$$\pi(\psi|D) \propto p(D|\psi)\pi(\psi).$$

Bayesian inference is accomplished by analyzing the joint posterior density of the model's parameters, or other functions of interest.

## A Likelihood Functions

We consider two models: the homoskedastic VAR model in equation (3), and the VAR-MSV model in (7). For notational convenience, let  $\mathbf{z} = \{\mathbf{z}_1, \dots, \mathbf{z}_T\}$  denote the dependent variables of the VAR systems in (3) and (7).

### A.1 Homoskedastic VAR Model

Let  $\psi_H = \{\mathbf{A}, \mathbf{\Sigma}\}$  denote the parameters of the homoskedastic VAR model in (3). The likelihood function (i.e., sampling density) for the homoskedastic VAR model is

$$\begin{aligned} p(\mathbf{z}|\psi_H) &= \prod_{t=1}^T p(\mathbf{z}_t|\psi_H, \mathbf{z}_{t-1}) \\ &= \prod_{t=1}^T \mathcal{N}_k(\mathbf{z}_t|\mathbf{A}\mathbf{z}_{t-1}, \mathbf{\Sigma}) \end{aligned} \tag{15}$$

where  $\mathcal{N}_k$  a  $k$ -variate Normal density. Appendix A describes the probability distribution functions and notation used in this paper.

## A.2 VAR-MSV Model

The parameters of the VAR-MSV model in (7) are the VAR matrix  $\mathbf{A}$  and the lower triangular matrix  $\mathbf{B}$ . The latent variances in the diagonal matrix  $\mathbf{V}_{t+1}$  are generated by the  $k$  stochastic volatility processes described in (9). Let  $\theta_j = \{\mu_j, \phi_j, \sigma_j\}$  denote the parameters of the  $j$ th stochastic volatility process. And, let  $\psi_{\text{MSV}} = \{\mathbf{A}, \mathbf{B}, \theta_1, \dots, \theta_k\}$  denote the full parameter set for the VAR-MSV model.

The likelihood function for the VAR-MSV model in (7) is

$$\begin{aligned} p(\mathbf{z}|\psi_{\text{MSV}}) &= \prod_{t=1}^T \int p(\mathbf{z}_t|\psi_{\text{MSV}}, \mathbf{z}_{t-1}, \mathbf{v}_t) p(\mathbf{v}_t|\psi_{\text{MSV}}) d\mathbf{v}_t \\ &= \prod_{t=1}^T \int \mathcal{N}_k(\mathbf{z}_t|\mathbf{A}\mathbf{z}_{t-1}, \mathbf{B}\mathbf{V}_t\mathbf{B}') p(\mathbf{v}_t|\theta, \mathcal{F}_{t-1}) d\mathbf{v}_t \end{aligned} \quad (16)$$

where  $\mathcal{F}_{t-1} = \{\mathbf{v}_1, \dots, \mathbf{v}_{t-1}\}$  denotes the history of the latent log-variance processes through  $t - 1$ . Note that the parameters  $\theta = \{\theta_1, \dots, \theta_k\}$  do not enter (16) directly. Rather, the likelihood function is evaluated by integrating over the latent log-variances  $\mathbf{v}_t = [v_{1,t+1} \cdots v_{k,t+1}]'$ . Estimation of this model using frequentist methods (e.g., maximum likelihood) is impractical because the integral in (16) is analytically intractable. However, the model is amenable to estimation using Bayesian MCMC methods. The Gibbs sampler, discussed below in Section II.C, effectively integrates over the latent log-variances to obtain the joint posterior density of the model's parameters and functions of interest (e.g., variance components).

## B Priors

In Bayesian econometrics, prior densities summarize the researcher's prior beliefs about the model's parameters. Following Hollifield, Koop, and Li (2003), we consider four priors:

**Prior 1** An uninformative base prior.

**Prior 2** Base prior with covariance stationarity imposed.

**Prior 3** Base prior with covariance stationarity imposed and stochastic initial condition.

**Prior 4** Prior 3 with an additional "features of interest" prior.

## B.1 Base Prior

For the homoskedastic VAR model, we assume an independent Normal-Wishart prior. We will refer to this as the “base prior” for the homoskedastic VAR model. Specifically, we assume that the priors are independent and can be factored

$$\pi(\mathbf{A}, \boldsymbol{\Sigma}) = \pi(\mathbf{A})\pi(\boldsymbol{\Sigma}), \quad (17)$$

where

$$\pi(\mathbf{a}) = \mathcal{N}_{k^2}(\mathbf{a}|\mathbf{a}_0, \mathbf{A}_0^{-1}) \quad (18)$$

$$\pi(\boldsymbol{\Sigma}^{-1}) = \mathcal{W}(\boldsymbol{\Sigma}^{-1}|\mathbf{S}_0, s_0). \quad (19)$$

Note that  $\mathbf{a} = \text{vec}(\mathbf{A}')$  where the  $\text{vec}(\cdot)$  operator stacks the columns of its argument. We assume that  $\boldsymbol{\Sigma}^{-1}$  is distributed Wishart. This is equivalent to assuming that  $\boldsymbol{\Sigma}$  is distributed inverse-Wishart. Appendix A describes the probability distributions and notation used in the paper.

The researcher’s prior beliefs are reflected in the choice of hyperparameters  $\mathbf{a}_0$ ,  $\mathbf{A}_0$ ,  $s_0$  and  $\mathbf{S}_0$ . The Normal-Wishart prior converges to the diffuse or flat prior employed by Stambaugh (1999) as  $\mathbf{a}_0$ ,  $\mathbf{A}_0$ ,  $s_0$  and  $\mathbf{S}_0$  all converge to zero. Our choice of hyperparameters for the base prior corresponds to relatively uninformative prior beliefs regarding the parameters. We discuss our choice of hyperparameters further with the empirical results.

For the VAR-MSV model, we assume the priors are independent and can be factored

$$\pi(\mathbf{A}, \mathbf{B}, \theta_1, \dots, \theta_k) = \pi(\mathbf{A})\pi(\mathbf{B})\pi(\theta_1) \cdots \pi(\theta_k). \quad (20)$$

We assume  $\pi(\mathbf{A})$  is the same as the homoskedastic VAR model (see (18)). Let  $\mathbf{b}$  denote the  $(k(k-1)/2) \times 1$  vector containing the free elements of  $\mathbf{B}$ . The base prior for  $\mathbf{b}$  is also Normal:

$$\pi(\mathbf{b}) = \mathcal{N}_{k(k-1)/2}(\mathbf{b}|\mathbf{b}_0, \mathbf{B}_0^{-1}). \quad (21)$$

The priors  $\pi(\theta_j) = \pi(\mu_j)\pi(\phi_j)\pi(\sigma_j)$  are chosen to be relatively uninformative regarding the

dynamics of the SV processes described in (9). For  $\mu_j$ , the unconditional log variance, we assume:

$$\pi(\mu_j) = \mathcal{N}(\mu_j | m_{0j}, M_{0j}). \quad (22)$$

$\phi_j$  reflects the persistence of the SV process. We constrain  $\phi_j$  to the region of stationarity (i.e.,  $|\phi_j| < 1$ ) by invoking the change of variable  $\phi_j = 2\phi_j^* - 1$  and assuming that  $\phi_j^*$  is distributed Beta:

$$\pi(\phi_j^*) = \mathcal{B}(\phi_j^* | \phi_{0j}^{(1)}, \phi_{0j}^{(2)}). \quad (23)$$

We assume that  $\sigma_j$ , the volatility of the SV process, is distributed Inverse-Gamma:

$$\pi(\sigma_j) = \mathcal{IG}(\sigma_j | \alpha_{0j}, \beta_{0j}). \quad (24)$$

The choice of hyperparameters  $\{m_{0j}, M_{0j}, \phi_{0j}^{(1)}, \phi_{0j}^{(2)}, \alpha_{0j}, \beta_{0j}\}$  are discussed with the empirical results.

## B.2 Base Priors with Covariance Stationarity

When lagged endogenous regressors are employed in predictive regressions (including VARs), inference is very sensitive to high autocorrelation, and especially a unit root, in a predictive variable. If the researcher has prior beliefs regarding the stationarity of the predictive variable (covariance stationarity in the case of a VAR), then inference can be enhanced by incorporating those prior beliefs. The matrix  $\mathbf{A}$  is covariance stationary if its eigenvalues lie inside the unit circle (i.e.,  $\|\mathbf{A}\| < 1$ ).<sup>8</sup> The base prior does not restrict  $\mathbf{A}$  to the region of covariance stationarity. As demonstrated in Hollifield, Koop, and Li (2003), sampling  $\mathbf{A}$  from outside the region of covariance stationarity has dramatic (perhaps pathological) effects on posterior densities. Let  $S \equiv \{\mathbf{A} \in \mathfrak{R}^{k \times k} \text{ such that } \|\mathbf{A}\| < 1\}$  denote the region of covariance stationarity, and let  $I(\mathbf{A} \in S)$  be the corresponding indicator variable. Following Stambaugh (1999) and Hollifield, Koop, and Li (2003), we entertain prior beliefs regarding the covariance stationarity of the VAR. Under this prior belief,  $\mathbf{A}$  is constrained to the region of covariance stationarity and (18) is rewritten as:

$$\pi(\text{vec}(\mathbf{A})) = \mathcal{N}_{k^2}(\text{vec}(\mathbf{A}) | \mathbf{a}_0, \mathbf{A}_0^{-1}) I(\mathbf{A} \in S). \quad (25)$$

---

<sup>8</sup>See Hamilton (1994) p. 259.

The covariance stationarity prior on  $\mathbf{A}$  applies to both the homoskedastic VAR and VAR-MSV models. Priors on all other parameters are identical to the base prior.

### B.3 Base Prior with Covariance Stationarity and Stochastic Initial Condition

In the standard regression framework, the initial condition  $\mathbf{z}_0$  is assumed to be deterministic. In that case, it would be appropriate to work with the priors and likelihood functions discussed above. For the application we consider, the assumption of deterministic regressors is clearly invalid. Following Stambaugh (1999) and Hollifield, Koop, and Li (2003), we consider estimation of the VAR models under the alternative assumption that the initial condition  $\mathbf{z}_0$  is stochastic rather than deterministic.

First, consider the homoskedastic VAR model. Since the initial condition is stochastic,  $\mathbf{z}_0$  can be treated as a parameter with its own prior density. Our prior on  $\mathbf{z}_0$  is a function of the other parameters of the model. Given  $\psi_H$  and assuming covariance stationarity, we assume that  $\mathbf{z}_0$  is drawn from its unconditional or steady-state density,

$$p(\mathbf{z}_0|\psi_H) = \mathcal{N}_k(\mathbf{z}_0|\mathbf{0}, \mathbf{\Omega}_0), \quad (26)$$

where

$$\text{vec}(\mathbf{\Omega}_0) = [\mathbf{I} - (\mathbf{A} \otimes \mathbf{A})]^{-1} \text{vec}(\mathbf{\Sigma}). \quad (27)$$

Equations (26) and (27) imply that the vector  $\mathbf{z}_0$  is drawn from a multivariate Normal density with mean zero (since we work with demeaned data) and unconditional covariance matrix  $\mathbf{\Omega}_0$ .

Stambaugh (1999) distinguishes between “conditional” and “exact” likelihood functions. In Stambaugh’s terminology, the likelihood function in (15) is “conditional” on observing the initial condition  $\mathbf{z}_0$ . When the initial condition is stochastic rather than deterministic, Stambaugh (1999) employs an “exact” likelihood function which corresponds to the product of the “conditional” likelihood in (15) and the density in (26). Hollifield, Koop, and Li (2003) treat the density in (26) as a prior. So, depending on one’s perspective, the density in (26) could be considered a prior or a component of the likelihood function. Since the posterior density is proportional to the product of the likelihood function and the priors, the distinction is immaterial. The inference will be the same.

Next, consider the VAR-MSV model. Given  $\psi_{MSV}$ , we assume that  $\mathbf{z}_0$  is drawn from its uncon-



ditional density,

$$p(\mathbf{z}_0|\psi_{\text{MSV}}) = \mathcal{N}_k(\mathbf{z}_0|\mathbf{0}, \mathbf{\Omega}_0) \quad (28)$$

where

$$\text{vec}(\mathbf{\Omega}_0) = [\mathbf{I} - (\mathbf{A} \otimes \mathbf{A})]^{-1} \text{vec}(\mathbf{B}\mathbf{V}_0\mathbf{B}') \quad (29)$$

$$\mathbf{V}_0 = \text{diag} \left\{ \exp \left( \mu_1 + \frac{1}{2} \frac{\sigma_1^2}{(1 - \phi_1^2)} \right), \dots, \exp \left( \mu_k + \frac{1}{2} \frac{\sigma_k^2}{(1 - \phi_k^2)} \right) \right\}. \quad (30)$$

The exponential term  $\exp(\mu_j + \sigma_j^2/2(1 - \phi_j^2))$  in (30) is the unconditional or steady-state variance of the  $j$ th independent SV processes described in (9). It follows that  $\mathbf{B}\mathbf{V}_0\mathbf{B}'$  is the unconditional covariance matrix for the VAR system error vector.

#### B.4 Features of Interest Prior

Hollifield, Koop, and Li (2003) suggests that parameters for a reduced-form VAR such as (3) are difficult to interpret. This makes eliciting prior beliefs regarding the model’s parameters equally difficult. However, researchers may have prior beliefs regarding “features of interest.” In the present model, we are interested in the terms of the variance decomposition. These features of interest are highly non-linear functions of the model’s parameters. We discuss technical aspects of the features of interest prior in Appendix B.

### C Prior-Posterior Analysis

Combining the likelihood function and the priors via Bayes rule, one obtains the joint posterior density of the model’s parameters given the data. Given the form of the likelihood functions in (15) and, especially, in (16), and given any of the priors described above, the joint posterior cannot be estimated (i.e., sampled) directly. Fortunately, the Gibbs sampling method bypasses the computation of the likelihood function and computation of the joint posterior density. Rather, the Gibbs sampler algorithm generates draws from the conditional distribution of each block of parameters (i.e., the distribution of each block given the data, the prior and the other blocks of parameters). The draws from these conditional densities eventually converge to draws from the joint posterior density. Inference is based on summary statistics (e.g., mean, standard deviation, etc.) describing the distribution of the sample draws of the model’s parameters, and of functions

thereof.

Bayesian estimation of the homoskedastic VAR model is discussed in detail in Hollifield, Koop, and Li (2003). Our Bayesian treatment of the homoskedastic VAR model is essentially the same, and we refer the reader to that paper for details.

For the VAR-MSV model, note that it is the latent log-variances ( $\mathbf{v}_t$ ), and not the parameters of the SV processes ( $\theta$ ), that appear in the likelihood function (16). Bayesian MCMC methods treat latent variables as parameters and, thus, produce a sample from their posterior distribution as well. The other parameters are then drawn conditioning not only on the data, but also on the simulated values of the latent variables. This approach is known as data augmentation. Conditioning on the latent variables, the sampling density of  $\mathbf{z}_t$  can be written

$$\mathbf{z}_t | \mathbf{A}, \mathbf{B}, \mathbf{z}_{t-1}, \mathbf{v}_t \sim \mathcal{N}_k(\mathbf{A}\mathbf{z}_{t-1}, \mathbf{B}\mathbf{V}_t\mathbf{B}') \quad (31)$$

Let  $y_{jt} = \ln(\mathbf{B}_j^{-1}(\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1}))^2$  and  $\{v_j\} = (v_1, \dots, v_k)$ , where  $v_j = (v_{j1}, \dots, v_{jT})$  denotes the time series of the  $j$ th latent SV process. In similar fashion, let  $y_j = (y_{j1}, \dots, y_{jT})$ . Draws from the (augmented) posterior density are obtained by cycling through the following steps.

1. Initialize  $\mathbf{B}$ ,  $\mathbf{z}_0$  and  $v_j$  for  $j = 1, \dots, k$ .
2. Sample  $\mathbf{A} | \mathbf{z}, \{v_j\}, \mathbf{B}, \mathbf{z}_0$ .
3. Sample  $\mathbf{B} | \mathbf{z}, \{v_j\}, \mathbf{A}, \mathbf{z}_0$ .
4. Compute  $y_j$  and sample  $\theta_j$  and  $\{v_j\}$  by repeating the following steps for  $j = 1, \dots, k$ :
  - (a) Draw  $\theta_j$  from  $\theta_j | y_j$
  - (b) Draw  $v_j$  from  $v_j | y_j, \theta_j$ .
5. Sample  $\mathbf{z}_0 | \mathbf{B}, \mathbf{z}, \{v_j\}, \mathbf{A}$
6. Go to step 2 and repeat.

Under Prior 1 and Prior 2, step 5 and any conditioning on  $\mathbf{z}_0$  in the other steps are omitted. Details on each step of the simulation are provided in Appendix B.

### III Data

We closely follow Campbell (1991) in constructing the data set. The vector  $\mathbf{z}_t = \{h_t (D/P)_t rb_t\}'$  includes three elements: a real stock index return, a dividend-price ratio, and a stochastically detrended short-term interest rate. The dividend-price ratio (or dividend yield) and relative T-bill rate are chosen for their well-documented ability to forecast stock returns. Fama and Schwert (1977) first document the forecasting power of short-term interest rates. Shiller (1984), Fama and French (1988), and Campbell and Shiller (1988a) were among the first papers to report the significant forecasting power of the dividend-price ratio. These results are not without controversy. Mankiw and Shapiro (1986), Nelson and Kim (1993), and Stambaugh (1999) find that OLS estimates of the dividend-price ratio's predictive power are biased in small samples (i.e., forecasting power is overstated) since the dividend-price ratio (a lagged endogenous regressor) is highly autocorrelated. Stambaugh (1999) and Lewellen (2004) find that incorporating prior information about the stationarity of the predictive regressor strengthens evidence of predictability. Ang and Bekaert (2004) report that the ability of the dividend-price ratio to predict returns is "best visible at short horizons with the short rate as an additional regressor."<sup>9</sup> Given the state of the literature, we believe that our choice of predictive regressors for the VAR is reasonable, and that estimation of the model using Bayesian MCMC methods may provide useful insights on the predictability debate.

We study the period 1952:1–2002:12 (612 monthly observations). We choose this period because of evidence that the data generating process was fundamentally different prior to 1952. As noted by Schwert (1989) and Kim, Nelson, and Startz (1991), stock returns in the pre-war period (particularly the depression years) were substantially more volatile than in any period since 1952. Furthermore, interest rates were artificially smooth prior to the 1951 Fed-Treasury accord.

We construct a real total stock return index by deflating the CRSP monthly value-weighted index of NYSE stocks (with dividends) using the Consumer Price Index (CPI) obtained from CRSP. The log total return is defined  $h_t = \ln(P_t + d_t) - \ln(P_{t-1})$  where  $P_t$  is the real index level at  $t$  and  $d_t$  is the real dividend paid between  $t - 1$  and  $t$ . Real returns are expressed in percent per month.

We compute the dividend-price ratio in the manner of Fama and French (1988). We construct monthly dividend and price index series using the return series (with and without dividends) for

---

<sup>9</sup>Other recent papers in the debate include Goyal and Welch (2003) and Campbell and Yogo (2005).

the CRSP monthly value-weighted index of NYSE stocks. The dividend-price ratio is the sum of the dividends paid over the previous twelve months divided by the current level of the stock price index,

$$\left(\frac{D}{P}\right)_t = \frac{1}{P_t} \sum_{j=0}^{11} d_{t-j}.$$

The dividend-price ratio is expressed in percent per annum.

We construct the stochastically detrended short-term interest rate in the manner of Campbell (1991) and Hodrick (1992). The relative bill rate is the one-month Treasury bill yield,  $y_{1,t+1}$ , less its twelve-month moving average,

$$rb_t = y_{1,t} - \frac{1}{12} \sum_{i=1}^{12} y_{1,t-i}.$$

The Treasury bill data are from Ibbotson & Associates. The relative bill rate is multiplied by 1200 to express it in percent per annum.

The time-series of  $(D/P)_t$  and  $rb_t$  are plotted in Figure 1.

Figure 1 goes about here.

## IV Empirical Results

### A Priors

We assume a prior for the  $\mathbf{A}$  matrix that contains almost no information. Specifically, we set the hyperparameters  $\mathbf{a}_0 = 0$  and  $\mathbf{A}_0^{-1} = 10^6 \cdot \mathbf{I}_{k^2}$ . This implies that VAR coefficients are normally distributed with mean zero and standard deviation 1000 (i.e., nearly flat). For the homoskedastic VAR model, the hyperparameters for the prior on the  $\mathbf{\Sigma}$  matrix are  $\mathbf{S}_0 = \mathbf{I}_k$  and  $s_0 = 8$ . The latter choice is the least informative within the values for which the prior mean of  $\mathbf{\Sigma}$  exists. Our choices for  $\mathbf{A}$  and  $\mathbf{\Sigma}$  are identical to Hollifield, Koop, and Li (2003).

For the  $\mathbf{B}$  matrix in the VAR-MSV model, we set  $\mathbf{b}_0 = 0$  and  $\mathbf{B}_0^{-1} = 10 \cdot \mathbf{I}_{k(k-1)/2}$ . For  $\mu_j$ , the average log-variance for error shocks, we set  $m_{0j} = -2$  and  $M_{0j} = 25$  for all  $j$ . At the prior mean, this implies an average volatility of about 8.3% per month with a standard deviation of over

200%. Again, this prior is very uninformative. For  $\phi_j$ , we set  $\phi_{0j}^1 = 20$  and  $\phi_{0j}^2 = 1.5$  so that the distribution has a mean of 0.86 with standard deviation of 0.11, reflecting the expected persistence in the stochastic volatility dynamics. For  $\sigma_j$ , we set  $\alpha_{0j} = 2.39$  and  $\beta_{0j} = 0.347$  implying a mean of 0.25 and a standard deviation of 0.4 for the volatility of the log-variance.

## B Homoskedastic VAR Models

The Campbell (1991) model allows us to address the question: What moves the stock market? Table I reports summary statistics describing the posterior densities for parameters of the VAR coefficient matrix  $\mathbf{A}$  and the covariance matrix  $\Sigma$  for the homoskedastic VAR model. We report results obtained under the four priors discussed in Section II. Since the posterior densities for  $\mathbf{A}$  and  $\Sigma$  are very robust to our different prior specifications, we discuss them together. Unless otherwise noted, our discussion will focus on the posterior densities obtained under Prior 4 (i.e., the features of interest prior).

We find that the dividend-price ratio and the relative bill rate do forecast future real stock returns. Posterior estimates of  $A_{12}$  and  $A_{13}$  are 0.312 and  $-0.599$ , respectively. The inference is statistically reliable. For the dividend-price ratio, the 90% Bayesian confidence interval for  $A_{12}$  is  $[0.091, 0.548]$ , and only 0.0077 of the posterior density is less than zero.<sup>10</sup> For the relative bill rate, the 90% Bayesian confidence interval for  $A_{13}$  is  $[-0.837, -0.361]$ , and only 0.0006 of the posterior density is greater than zero. High dividend-price ratios forecast higher than average future real stock returns, and high relative short-term interest rates forecast lower than average future real stock returns. The parameters in the second and third rows of  $\mathbf{A}$  indicate that the dividend-price ratio and the relative bill rate would be described well by univariate first-order autoregressions. As expected, the dividend-price ratio is very highly autocorrelated. The posterior estimate of  $A_{22}$  is 0.988 with a 90% Bayesian confidence interval of  $[0.978, 0.997]$ . This is where one might reasonably expect the covariance stationarity prior to play an important role. However, posterior estimates of  $A_{22}$  are nearly unaffected by the imposition of covariance stationarity. The proportion of the posterior density of  $A_{22}$  greater than one ranges from 0.0154 under Prior 1 to 0.0108 under Prior 4. We conclude that the dividend-price ratio is stationary for the 1952–2002 sample period. It is important to note that our inferences regarding predictability are not sensitive to the existence of

---

<sup>10</sup>We use the term posterior estimate to denote the mean of the posterior density. The 90% Bayesian confidence interval describes the region between the 5% and 95% quantiles of the posterior density. Alternatively, we could use the standard deviation of the posterior density to describe the precision of the posterior estimate.

a unit root in the predictive regressors. As discussed in Sims (1988) and Sims and Uhlig (1991), Bayesian posterior densities, and hence inferences, are valid even when regressors are non-stationary. Campbell and Yogo (2005) and Torous, Valkanov, and Yan (2005) show that traditional frequentist inference tools (e.g.,  $t$ -tests and  $p$ -values) may be invalid when regressors are nearly integrated.

Table I goes about here.

The posterior estimates of  $\Sigma$  and  $R^2$  reported in Table I are also very robust to prior specification. The unconditional variance of unexpected stock returns,  $\Sigma_{11}$ , is 16.680. As expected, the covariance between unexpected real stock returns and innovations in the dividend-price ratio is strongly negative. The posterior estimate of  $\Sigma_{21}$  is  $-0.618$ , which corresponds to a contemporaneous correlation of  $\rho_{21} = -0.895$ . Innovations in the relative bill rate are weakly correlated with unexpected real returns ( $\Sigma_{31} = -0.260$ ,  $\rho_{31} = -0.082$ ) and innovations in the dividend-price ratio ( $\Sigma_{32} = 0.010$ ,  $\rho_{32} = 0.077$ ). Note that the parameters of  $\Sigma$  are estimated simultaneously with the parameters of  $\mathbf{A}$ . Posterior estimates of  $R^2$  are 0.030, 0.978 and 0.558 for  $h_{t+1}$ ,  $(D/P)_{t+1}$  and  $rb_{t+1}$ , respectively.<sup>11</sup> Campbell (1991) reported corresponding  $R^2$  statistics of 0.024, 0.937 and 0.450 for 1952:1–1988:12.

Posterior densities for the variance decompositions are reported in Table II. We report the shares of unexpected real stock return variance attributable to the variance components defined in (10). It is here that the importance of the covariance stationarity prior becomes evident. Under Prior 1 (i.e., the base prior), the posterior mean and standard deviation are “blown up” by a few MCMC draws from outside the region of covariance stationarity. In contrast, the median, 5% and 95% quantiles of the posterior densities appear reasonable and are very similar to those obtained under the other priors. The results for Prior 1 are consistent with the “pathological” results reported in Hollifield, Koop, and Li (2003).

Table II goes about here.

We find that most of the variance of unexpected real returns in the post-war period is attributable to “news about real returns.” Under Prior 2, posterior estimates of the shares due to  $\text{Var}(\eta_d)$ ,  $\text{Var}(\eta_h)$  and  $-2\text{Cov}(\eta_d, \eta_h)$  are 0.267, 0.786 and  $-0.054$ , respectively. Under Prior 4 (i.e., the features of interest prior), posterior estimates of the shares change very little (0.261, 0.739 and

---

<sup>11</sup>For each equation in the VAR, we sample  $R^2 = 1 - \text{SSE}/\text{SST}$  for each iteration of the Gibbs sampler.

0.000), but the estimates are more precise (i.e., lower posterior standard deviations). These results are reasonably consistent with Campbell (1991), which reports shares of 0.127, 0.772 and 0.101, respectively, for the 1952:1–1988:12 period. Even when covariance stationarity is imposed (i.e., under Priors 2, 3 and 4), variance decompositions remain sensitive to draws of  $\mathbf{A}$  near the edge of the region of stationarity. The resulting skewness in the posterior densities of the shares of  $\text{Var}(\eta)$  is evident in the differences between the posterior means and medians reported in Table II. Under Prior 4, the posterior medians are 0.236, 0.711 and 0.070, respectively.

Bayesian posterior densities for two other features of interest are also reported in Table II. Sample draws for  $\text{Corr}(\eta_d, \eta_h)$  are constructed from the sample draws of the model’s other parameters. The posterior mean {median} for  $\text{Corr}(\eta_d, \eta_h)$  is  $-0.068$   $\{-0.090\}$ , but the Bayesian confidence interval  $[-0.520, 0.468]$  is relatively wide (i.e., imprecise) and includes zero. Thus, we can offer no conclusive inference regarding the sign of  $\text{Corr}(\eta_d, \eta_h)$ . Campbell (1991) reports the point estimate of  $\text{Corr}(\eta_d, \eta_h) = -0.161$  for 1952:1–1988:12, but is also inconclusive regarding the sign (std. error = 0.256). Following Campbell (1991), we also define a VAR persistence measure. Let

$$P_h \equiv \frac{\sigma(\eta_{h,t+1})}{\sigma(u_{t+1})} = \frac{(\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda})^{1/2}}{(\mathbf{e}\mathbf{1}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'\mathbf{e}\mathbf{1})^{1/2}} \quad (32)$$

where  $\eta_{h,t+1}$  is the previously defined “news about future returns,”  $u_{t+1} = \mathbf{e}\mathbf{1}'\mathbf{A}\mathbf{w}_{t+1}$  is the innovation to the expected return, and the  $\sigma(\cdot)$  operator gives the standard deviation of its argument. The VAR persistence measure summarizes the economic effect of persistent expected returns: a 1% increase in the expected return causes a  $P_h\%$  decrease in the real stock price index. We report the posterior densities of  $P_h$  for the homoskedastic VAR model in Table II. Under Prior 4, the posterior mean {median} is 7.252 {6.844} and the Bayesian 90% confidence interval is [4.076, 11.720]. For comparison, Campbell (1991) reports a point estimate of 5.794 (std. error = 1.469) for 1952:1–1988:12. Note that Campbell (1991) uses the delta method to compute standard errors for these features of interest. The Bayesian MCMC approach produces exact small-sample posterior densities for these features of interest.

Although the predictive regressors explain only a modest fraction of the variance in real stock returns (i.e.,  $R^2 = 0.030$ ), we conclude that “news about future returns” explains the lion’s share of the variance in unexpected real returns. How do we reconcile these two seemingly disparate conclusions? Expected real stock returns are very persistent. This characteristic is related to the persistence of the predictive regressors, particularly the dividend-price ratio. Viewed through the

lens of the asset pricing framework, a small innovation in the expected real return can cause a large unexpected stock return (i.e.,  $P_h = 7.252$ ). If expected returns were constant (i.e., unforecastable), we would expect all of the variance in unexpected real stock returns to be attributable to “news about future dividends.” The data does not support this alternative hypothesis. Our results are consistent with previous empirical evidence that aggregate dividend growth is approximately constant (i.e., unforecastable).<sup>12</sup>

## C VAR-MSV Models

The VAR-MSV model introduced in this paper allows us to address an additional question: Why does stock market volatility change over time? Given the asset pricing framework and VAR-MSV model for forecasting real stock returns, the variance of unexpected real returns has three time-varying components: variance of “news about future returns,” variance of “news about future dividends,” and a covariance term.

We first examine the predictability of real stock returns. Table III reports summary statistics describing the posterior densities for parameters of the VAR-MSV model. Posterior estimates of the  $\mathbf{A}$  matrix are very robust to prior specification, and are not appreciably different than those reported in Table I for the homoskedastic VAR model. We focus primarily on the posterior densities under Prior 4. The first row of  $\mathbf{A}$  indicates strong evidence of real stock return predictability. Future real returns are positively related to the dividend-price ratio ( $A_{12} = 0.295$ ) and inversely related to the relative bill rate ( $A_{13} = -0.526$ ). For  $A_{12}$ , the proportion of the posterior density less than zero ranges from 0.0378 under Prior 1 to 0.0049 under Prior 4. For  $A_{13}$ , the proportion of the posterior density greater than zero is less than 0.0003 for all four priors. The dividend-price ratio is highly persistent. The posterior mean of  $A_{22}$  is 0.991 with a 90% Bayesian confidence interval of [0.983, 0.998]. The covariance stationarity prior appears to play a more important role for the VAR-MSV model than for the homoskedastic VAR model. For  $A_{22}$ , the proportion of the posterior density greater than one ranges from 0.0904 under Prior 1 to 0.0123 under Prior 4. The relative bill rate is less persistent. The posterior mean of  $A_{33}$  is 0.756 with a 90% Bayesian confidence interval of [0.738, 0.828]. Most of the persistence in expected returns is associated with the forecasting power and persistence of the dividend-price ratio. The vectors  $\boldsymbol{\lambda}'$  and  $(\mathbf{e}\mathbf{1}' + \boldsymbol{\lambda}')$  (not reported in Table III)

---

<sup>12</sup>Cochrane (2001) provides a lucid review and interpretation of the empirical evidence on long-run stock return predictability and excess volatility. Cochrane concludes (p. 405) that “Return forecastability follows from the fact that dividends are not forecastable, and that the dividend/price ratio is highly but not completely persistent.”



map the VAR innovations into “news about future returns.” and “news about future dividends,” respectively. As defined in (4),  $\boldsymbol{\lambda}' = \mathbf{e}\mathbf{1}'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}$  is a highly nonlinear function of the VAR coefficient matrix  $\mathbf{A}$ . The posterior densities of  $\boldsymbol{\lambda}'$  are easily simulated from the Gibbs sampler output. Under Prior 4, the posterior means and standard deviations of  $\boldsymbol{\lambda}'$  are [0.017 23.97 0.588] and [0.021 6.27 0.840], respectively. Only the second element of  $\boldsymbol{\lambda}'$  is reliably non-zero. This indicates that innovations in the dividend-price ratio are positively correlated with innovations in  $\eta_{h,t+1}$ , and is consistent with mean reversion in stock prices.

Table III goes about here.

Table III also reports summary statistics describing the posterior densities for the  $\mathbf{B}$  matrix and for parameters of the SV processes. Posterior estimates of  $B_{21}$  reflect the strong inverse relation between unexpected real returns and innovations in the dividend-price ratio. This strong inverse relation, along with the persistence of the dividend-price ratio, contributes to the small-sample bias in univariate predictive regressions discussed in Stambaugh (1999). Bayesian estimates of the predictive equation (i.e., the first row of the  $\mathbf{A}$ ) are free of this small-sample bias. For real stock returns, posterior estimates of  $\mu_j$ , the intercept of the log-variance specification, corresponds to a standard deviation of about 3.8% per month.<sup>13</sup> Posterior estimates of  $\phi_j$  indicate that all three SV processes are quite persistent. The posterior mean of  $\phi_1$  is 0.917. From the posterior draws of  $\phi_1$  we can construct the posterior distribution of the half-life for volatility shocks. We find that a shock to the log-variance of unexpected real stock returns has a half-life of about 10 months. This level of volatility persistence for the U.S. stock market is consistent with estimates reported in the GARCH literature. For example, Scruggs (1998) estimates a univariate EGARCH(1,1)-M model for monthly excess stock index returns and reports persistence of about 0.90 for 1950:3–1994:12.

Stock market volatility varied widely in the 1952–2002 sample period. The top panel of Figure 2 plots the posterior mean of the variance of unexpected real stock returns,  $\text{Var}(\eta_{t+1})$ .<sup>14</sup> Monthly variance peaks in October 1974, and again in October 1987. Sustained periods of high volatility are evident from 1978–1982 and later from 1998–2002. The vertical gray bars in Figure 2 indicate recessions as defined by the National Bureau of Economic Research (NBER). Many, but not all, periods of high stock market volatility are associated with recessions. This is consistent with

<sup>13</sup>We report on  $\exp(\mu_j/2)$  rather than  $\mu_j$  to make interpretation easier.

<sup>14</sup>Variances and covariances are expressed in percent squared per month. Each point on the graph is based on 50,000 MCMC draws from the VAR-MSV model under Prior 4.

Schwert (1989), which finds that volatility is significantly higher during recessions.

Figure 2 goes about here.

Real stock return variance changes over time primarily because of the time-varying variance of “news about future returns.” The middle panel of Figure 2 plots the posterior medians (under Prior 4) of the three components of real return variance:  $\text{Var}(\eta_{d,t+1})$ ,  $\text{Var}(\eta_{h,t+1})$  and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$ . Variation in  $\text{Var}(\eta_{h,t+1})$  tracks the variation in  $\text{Var}(\eta_{t+1})$  (see the top panel) very closely for much of the sample period, and is responsible for most of the changes in real stock return variance. The time-varying variance of “news about future dividends,”  $\text{Var}(\eta_{d,t+1})$ , and the covariance term,  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$ , play secondary roles for much of the sample period. However, there are episodes during which  $\text{Var}(\eta_{d,t+1})$  and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$  play more prominent roles. The variance of “news about future dividends” increases dramatically during the 1973–1975 recession and remains high well into the 1980s. During this period, the sum  $\text{Var}(\eta_{d,t+1}) + \text{Var}(\eta_{h,t+1})$  exceeds  $\text{Var}(\eta_{t+1})$ . The covariance term  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$  turns negative (as it must according to equation (10)) and increases dramatically in magnitude during this period. If  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1}) < 0$ , then  $\text{Corr}(\eta_{d,t+1}, \eta_{h,t+1}) > 0$ . A positive correlation between  $\eta_{d,t+1}$  and  $\eta_{h,t+1}$  implies that upward revisions in expectations regarding future dividends (i.e., good news) are accompanied, on average, by upward revisions in expectations regarding future returns (i.e., bad news). From (2), we know that such revisions in expectations should have opposite effects on unexpected real stock returns, and should thus dampen stock market volatility.

Campbell and Vuolteenaho (2004) recently proposed a two-beta variant of the ICAPM in which the risk factors are “news about future dividends” and “news about future returns.” Since the model is based on the unexpected return decomposition in (1), the two betas sum to the conventional market beta. Their model suggests that exposure to “news about future dividends” should have a higher price (hence the name “bad beta”) than exposure to “news about future returns” (i.e., “good beta”). They find that value stocks and small stocks have higher cash flow betas (i.e., “bad betas”) than growth stocks and large stocks, and that this explains their higher average returns. The plots in the middle panel of Figure 2 indicate that the volatilities of these two risk factors change in interesting ways over time.

The bottom panel of Figure 2 plots posterior medians (under Prior 4) of the variance components as shares of  $\text{Var}(\eta_{t+1})$ . The story is much the same as that in the preceding paragraphs.

Although the share of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{h,t+1})$  varies over the sample period, it is always the largest component of unexpected real return variance. This ratio is perhaps the most meaningful of the three because  $\text{Var}(\eta_{h,t+1})$  and  $\text{Var}(\eta_{t+1})$  are estimated directly. The share of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{d,t+1})$  increases dramatically starting in the 1973-1975 recession and peaks during the 1980 recession. It then slowly descends back to previous levels during the 1980s. However, as noted previously in equation (14), the difference between the shares of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{d,t+1})$  and  $\text{Var}(\eta_{h,t+1})$  is an intertemporal constant. This is evident in Figure 2. The posterior mean of this difference is  $-0.603$ . Increases in the shares attributable to  $\text{Var}(\eta_{d,t+1})$  and  $\text{Var}(\eta_{h,t+1})$  are offset exactly by increases in the share attributable to  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$ . Since the shares must sum to one, this is another mechanical result. The shares of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{d,t+1})$ ,  $\text{Var}(\eta_{h,t+1})$  and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$  range from  $[0.10, 0.71, 0.21]$  to  $[0.61, 1.20, -0.79]$  during the sample period.<sup>15</sup> The time-series means of these shares are 0.24, 0.84, and  $-0.04$ , respectively.

As was noted for the homoskedastic case, the posterior densities of the variance component shares are skewed. Thus, posterior standard deviations don't accurately describe the dispersion of the posterior densities. We employ graphical methods to convey information regarding the posterior densities of the time-varying shares of unexpected real stock return variance. The graphs in Figure 3 plot the median, 5% and 95% quantiles of the posterior densities (under Prior 4) of the shares attributable to  $\text{Var}(\eta_{d,t+1})$ ,  $\text{Var}(\eta_{h,t+1})$ , and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$ .<sup>16</sup> For each graph, the median corresponds to the share plotted in the bottom graph of Figure 2. The posterior densities of the shares are clearly skewed, and the degree of skewness changes over time. Of particular note is the posterior density of the share attributable to  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$  plotted in the bottom graph. The posterior median varies considerably over time, and even changes sign several times. However, the 90% Bayesian confidence interval straddles zero for every date in the sample period.

The graphs in Figure 4 depict the posterior densities of three additional time-varying features of interest:  $\text{Corr}(\eta_{d,t+1}, \eta_{h,t+1})$ ,  $\sigma(u_{t+1})$ , and  $P_{h,t+1}$ . Each graph plots the median, 5% and 95% quantiles of the posterior density. The posterior median of  $\text{Corr}(\eta_{d,t+1}, \eta_{h,t+1})$  ranges from less than  $-0.4$  to over  $0.4$ , and changes sign several times. This pattern, which is a reflection of the previously discussed results regarding the covariance term, is very intriguing. Theorists and empiricists alike

<sup>15</sup>These "triples" are posterior medians rather than posterior means. The posterior medians don't necessarily sum to exactly one.

<sup>16</sup>Quantiles describe samples based on 50,000 MCMC draws.

are becoming more interested in the correlation between “news about future returns” and “news about future dividends.” Menzly, Santos, and Veronesi (2004) derive a general equilibrium model that predicts a positive correlation. Lettau and Ludvigson (2005) investigate a consumption-based present-value relation. They report that dividend growth is forecastable at horizons from one to six years, and that dividend forecasts are positively correlated with excess return forecasts. Although our results regarding  $\text{Corr}(\eta_{d,t+1}, \eta_{h,t+1})$  are intriguing, we hesitate to make too much of them because the posterior density plotted in Figure 4 is relatively wide (i.e., imprecise) and the Bayesian 90% confidence interval brackets zero for every date.

The volatility of expected real return innovations,  $\sigma(u_{t+1})$ , also varies considerably over time. The posterior median of  $\sigma(u_{t+1})$ , plotted in the middle graph of Figure 4, ranges from as little as 20 basis points per month to over one percent per month in the post-war sample. The time-series average is 37 basis points per month. The volatility of expected real return innovations tends to increase during recessions and reaches its peak during the 1980 recession. This peak is driven by the unusually high volatility of the relative bill rate at that time.

We also define a time-varying counterpart for the VAR persistence measure defined in (32). Let

$$P_{h,t+1} \equiv \frac{\sigma(\eta_{h,t+1})}{\sigma(u_{t+1})} = \frac{(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{t+1}\boldsymbol{\lambda})^{1/2}}{(\mathbf{e}\mathbf{1}'\mathbf{A}\boldsymbol{\Sigma}_{t+1}\mathbf{A}'\mathbf{e}\mathbf{1})^{1/2}}. \quad (33)$$

Note that the denominator of  $P_{h,t+1}$ , as defined in (33), is the same  $\sigma(u_{t+1})$  term plotted in the middle graph of Figure 4. The bottom graph in Figure 4 plots the time series of the posterior density of  $P_{h,t+1}$ . The posterior median of  $P_{h,t+1}$  fluctuates widely, usually in the range from 5 to 20. The time-series average is 10.17. If the  $\mathbf{A}$  matrix is constant, why does the “persistence” measure vary over time? The explanation is simple. The level of  $P_{h,t+1}$  is largely determined by the persistence and time-varying volatilities of the predictive regressors. For example, the posterior median of  $P_{h,t+1}$  reaches its lowest level during the 1980 recession, when the volatility of expected return innovations,  $\sigma(u_{t+1})$ , reaches its highest level. As discussed above, the high levels of  $\sigma(u_{t+1})$  during this period are due to the unusually high volatility of the relative bill rate. Since shocks to the relative bill rate are much less persistent ( $A_{33} = 0.784$ , half-life of about 2.5 months) than shocks to the dividend-price ratio ( $A_{22} = 0.991$ , half-life of over 50 months), expected return innovations associated with shocks to the relative bill rate will cause smaller capital gains/losses. Hence,  $P_{h,t+1}$  is low. It’s interesting to note that the posterior median of  $P_{h,t+1}$  reaches its historical high near the end of the sample in 2002. This is due largely to the confluence of very high levels of  $\Sigma_{22,t+1}$

with very low levels of  $\Sigma_{33,t+1}$ . Expected return innovations in this environment are likely to be highly persistent and cause relatively large stock price changes. Hence,  $P_{h,t+1}$  is high.

In summary, we conclude that time-series patterns in stock market volatility are primarily driven by the time-varying variance of “news about future returns.” The variance of “news about future returns,” in turn, is explained by the persistence and time-varying volatility of expected return innovations. The variance of “news about future dividends” and the covariance term typically play a secondary role. The variance of “news about future dividends” increases substantially from the mid-1970s to the mid-1980s (a period which includes three recessions), but its contribution to stock market volatility during that period is offset by positive correlation between “news about future returns” and “news about future dividends.”

## V Conclusions

This paper considers the question: Why does stock market volatility change over time? Using the asset pricing framework of Campbell (1991), we decompose unexpected real stock market returns into two components: “news about future dividends” and “news about future returns.” Where Campbell (1991) uses a homoskedastic VAR to predict future returns, we introduce a VAR in which the conditional covariance and correlation matrices are time-varying. Specifically, we assume that the VAR covariance matrix follows a multivariate stochastic volatility (MSV) process. The VAR-MSV model allows us to decompose the variance of unexpected real stock returns into three time-varying components: variance of “news about future dividends,” variance of “news about future returns,” and a covariance term.

We find that time-varying variance of “news about future returns” is the primary reason that stock market volatility changes over time. Many, but not all, episodes of high uncertainty regarding future returns are associated with recessions. Volatility of “news about future dividends” normally plays a secondary role, but did increase dramatically during the recessions of 1973-1975, 1980, and 1981-1982. However, the contribution of “news about future dividends” to unexpected return variance during this period is offset by temporarily high positive correlation between “news about future dividends” and “news about future returns.” The shares of unexpected real return variance attributable to “news about future returns,” “news about future dividends” and the covariance term, respectively, range from [0.10, 0.71, 0.21] to [0.61, 1.20, -0.79] during the sample period.

The time-series means of these shares are 0.24, 0.84, and  $-0.04$ , respectively.

This study also contributes, both methodologically and empirically, to the debate on stock return predictability. We find that monthly real stock returns are predictable using the dividend-price ratio and the stochastically detrended short-term interest rate. Since the predictive regressors (especially the dividend-price ratio) are highly autocorrelated, it follows that expected real stock returns are highly persistent. Therefore, small innovations in expected returns can cause relatively large capital gains or losses. Estimates of the VAR-MSV model indicate that the volatility of expected return innovations ranges from 20 basis points per month to one percent per month during the post-war period. Taken together, results from the VAR-MSV model lead us to conclude that observed time-series patterns in stock market volatility can be explained by the persistence and time-varying volatility of expected real return innovations.

We employ innovative Bayesian MCMC methods to estimate the model. There are several advantages to employing Bayesian methods for this particular analysis. Bayesian methods are ideal for making inferences regarding return predictability. Since they deliver exact small-sample posterior densities for both the model's parameters and functions of interest, they are not subject to the small-sample biases that plague OLS estimators in predictive regressions with lagged endogenous regressors. Furthermore, inferences remain valid when regressors are highly autocorrelated, regardless of whether a unit root exists. We extend the Bayesian approach suggested in Stambaugh (1999) to VAR forecasting models with multiple predictive variables and multivariate stochastic volatility. Bayesian MCMC methods are also especially well-suited for estimating models with latent variables. In the VAR-MSV model, the stochastic volatilities are latent. MCMC methods essentially simulate and then integrate out these latent variables. We use the posterior densities of the SV processes to construct posterior densities for the time-varying components of unexpected real return variance.

The Campbell (1991) method for decomposing unexpected returns into “news about future returns” and “news about future dividends” has now become a standard modeling tool in empirical asset pricing. Allowing the volatilities of these unexpected return components to vary over time has important implications for asset pricing models, and is an important area for future research.<sup>17</sup> The VAR-MSV model and Bayesian MCMC methods introduced in this paper have potential applications in many strands of the literature.

---

<sup>17</sup>One example of work in this area is Chen (2003), which introduces a VAR-GARCH generalization of the of Campbell (1993, 1996) model.

## A Probability Distributions

The table below briefly describes the probability distribution functions (i.e., PDFs or densities) and notation used in the paper. We use both  $\pi(\cdot)$  and  $p(\cdot)$  to denote a probability density. However, we typically reserve the use of  $\pi(\cdot)$  for prior or posterior densities of the parameters. Further details on these probability distributions can be found in most elementary texts on Bayesian statistics (e.g., Zellner (1971), Poirier (1995), or Gelman, Carlin, Stern, and Rubin (1995)).

Distribution	Notation	Density	Comments
Normal	$x \sim \mathcal{N}(\mu, \sigma^2)$	$p(x \mu, \sigma^2) = \mathcal{N}(x \mu, \sigma^2)$	$\mu$ and $\sigma^2$ are scalars.
Multivariate Normal	$\mathbf{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$p(\mathbf{x} \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}_d(\mathbf{x} \boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\boldsymbol{\mu}$ is a $d \times 1$ vector. $\boldsymbol{\Sigma}$ is a $d \times d$ symmetric, positive definite covariance matrix.
Inverse-Gamma	$x \sim \mathcal{IG}(\alpha, \beta)$	$p(x \alpha, \beta) = \mathcal{IG}(x \alpha, \beta)$	$x > 0$ . Shape parameter $\alpha > 0$ . Scale parameter $\beta > 0$ .
Wishart	$\mathbf{X} \sim \mathcal{W}(\mathbf{S}, \nu)$	$p(\mathbf{X} \mathbf{S}, \nu) = \mathcal{W}(\mathbf{X} \mathbf{S}, \nu)$	$\mathbf{S}$ is a $d \times d$ symmetric, positive definite scale matrix. $\nu$ is degrees of freedom.
Beta	$x \sim \mathcal{B}(\alpha, \beta)$	$p(x \alpha, \beta) = \mathcal{B}(x \alpha, \beta)$	$x \in [0, 1]$ , $\alpha > 0$ , $\beta > 0$ .

## B Prior-Posterior Analysis for VAR-MSV Model

As one varies the priors on the model parameters and/or on the functions of interest, the resulting posterior distributions clearly change. As a consequence, although the steps of the Gibbs sampler remain the same as those described in section II.C, the specific implementation of each step differs depending on the prior. In what follows we will first detail the MCMC scheme under Prior 1 (the base uninformative prior). We will then report the differentiating elements induced in each step by the alternative priors. For notational convenience, let  $\mathbf{x} = \{\mathbf{z}_0, \dots, \mathbf{z}_{T-1}\}$  denote the  $T \times k$  matrix containing the lagged dependent variables of the VAR system.

### B.1 Sampling of A under Prior 1

The VAR model can be rewritten as

$$\mathbf{z}_t = \mathbf{X}_t \mathbf{a} \tag{B.1}$$

where  $\mathbf{X}_t = \mathbf{I}_k \otimes \mathbf{z}_{t-1}$ . Using standard Bayesian results in multivariate regression with heteroskedastic errors, the posterior distribution is then

$$\pi(\mathbf{a}|\mathbf{z}, \mathbf{x}, \mathbf{B}, \{v_j\}) \sim \mathcal{N}_{k^2}(\mathbf{a}_1, \mathbf{A}_1^{-1}) \quad (\text{B.2})$$

where  $\mathbf{A}_1 = (\mathbf{A}_0 + \sum_{t=1}^T \mathbf{X}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{X}_t)$  and  $\mathbf{a}_1 = \mathbf{A}_1^{-1}(\mathbf{A}_0 \mathbf{a}_0 + \sum_{t=1}^T \mathbf{X}'_t \boldsymbol{\Sigma}_t^{-1} \mathbf{z}_t)$

## B.2 Sampling of $\mathbf{B}$ under Prior 1

Conditioning on  $\{v_j\}$  and  $\mathbf{A}$  we can write the VAR as

$$\tilde{\mathbf{z}}_t = \mathbf{B} \mathbf{u}_t$$

or

$$\mathbf{B}^{-1} \tilde{\mathbf{z}}_t = \mathbf{u}_t$$

where  $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{A} \mathbf{z}_{t-1}$  and  $\mathbf{u}_t \sim \mathcal{N}_k(0, \mathbf{V}_t)$ . Given the diagonal structure of  $\mathbf{V}_t$ , the system can then be treated as  $k$  independent regressions where the conditional posterior density for the free elements of  $\mathbf{B}$  are obtained again from standard Bayesian updating formulas.

## B.3 Sampling of Volatilities and Volatility Parameters under Prior 1

Conditional on  $\mathbf{A}$  and  $\mathbf{B}$ , the independent error terms in (7) can be isolated as follows:

$$\mathbf{B}^{-1}(\mathbf{z}_t - \mathbf{A} \mathbf{z}_{t-1}) = \mathbf{V}_t^{1/2} \boldsymbol{\epsilon}_t \quad (\text{B.3})$$

where  $\boldsymbol{\epsilon}_t \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$ . We exploit a clever change of variable suggested by Kim, Shephard, and Chib (1998). Let

$$y_{jt} = \ln(\mathbf{B}_j^{-1}(\mathbf{z}_t - \mathbf{A} \mathbf{z}_{t-1}))^2 = v_{jt} + \ln(\epsilon_{jt}^2), \quad j = 1, \dots, k \quad (\text{B.4})$$

Note that  $y_{jt}$  is the sum of  $v_{jt}$  and a log chi-square random variable with one degree of freedom. Kim, Shephard, and Chib (1998) show that the density of a log chi-square random variable can be approximated quite precisely with a seven-component mixture of Normal distributions. This change of variable permits us to represent the  $\mathbf{v}_t$  vector as  $k$  independent, conditionally Normal



state-space models

$$y_{jt}|s_{jt}, v_{jt} \sim \mathcal{N}\left(v_{jt} + m_{s_{jt}}, v_{s_{jt}}^2\right) \quad (\text{B.5})$$

$$v_{jt} = \mu_j + \phi_j(v_{jt-1} - \mu_j) + \sigma_j \zeta_{jt}, \quad j = 1, \dots, k \quad (\text{B.6})$$

where  $s_{jt}$  is a discrete component indicator with mass function  $\Pr(s_{jt}) = q_i, i = 1, \dots, 7$ , and  $m_{s_{jt}}, v_{s_{jt}}^2$  and  $q_i$  are the parameters of each component as tabulated in Kim, Shephard, and Chib (1998). Conditional on  $\mathbf{A}$  and  $\mathbf{B}$ , one can sample  $\theta_j$  and the latent log-variances separately for each independent process in  $\mathbf{V}_t$  using methods developed for univariate SV models. See Kim, Shephard, and Chib (1998) and Chib, Nardari and Shephard (2002, 2005) for details on sampling parameters and latent log-variances for univariate SV processes.

#### B.4 Posterior Sampling under Prior 2

Imposing covariance stationarity on the VAR only restricts the sampling of  $\mathbf{A}$ . This is achieved by simply rejecting draws from (B.2) that violate the covariance stationarity condition and, in such cases, keeping the previous valid Gibbs draw. In all our applications, the number of  $\mathbf{A}$  draws rejected under Prior 2 is never larger than 10 to 12 percent of the sample.

#### B.5 Posterior Sampling under Prior 3

With a slight change of notation, let  $q(\mathbf{A}, \mathbf{B}, \boldsymbol{\theta}) = p(\mathbf{z}_0|\psi_{\text{MSV}})$  denote the prior on  $\mathbf{z}_0$ . Since the other parameters enter this prior,  $q(\cdot)$  will be part of their posterior density. Therefore, the posteriors cannot be sampled as above. We need to rely on the Metropolis - Hastings (MH) algorithm (see Chib and Greenberg (1995) for a thorough illustration) to sample these more involved densities. Briefly, the MH algorithm requires draws from a candidate (or, proposal) density and accepts them with a probability proportional to the proposal density and the actual posterior density, both evaluated at the current draw and, respectively, the candidate draw. We now provide some additional details for the sampling of the  $\mathbf{A}$  matrix. Similar strategies are applied for  $\mathbf{B}$  and  $\boldsymbol{\theta}$ .

### Sampling of $\mathbf{A}$ under Prior 3

The posterior density is proportional to  $\pi(\mathbf{a}|\mathbf{z}, \mathbf{B}, \{v_j\})q(\mathbf{A}, \mathbf{B}, \boldsymbol{\theta})$ . Let  $\mathbf{A}^{(g)}$  denote the draw at the iteration  $g$  of the Gibbs sampler, and let  $\mathbf{A}^{(*)}$  denote the candidate draw at Gibbs iteration  $g + 1$ . The latter is generated from (B.2) and it is accepted with probability

$$\alpha_A = \min \left( 1, \frac{q(\mathbf{A}^{(*)}, \mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)})}{q(\mathbf{A}^{(g)}, \mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)})} \right). \quad (\text{B.7})$$

If the candidate draw is rejected, the value of  $\mathbf{A}^{(g+1)}$  is set to  $\mathbf{A}^{(g)}$ .

### Sampling of $\mathbf{z}_0$ under Prior 3

The conditional posterior density of  $\mathbf{z}_0$  is

$$\pi(\mathbf{z}_0|\mathbf{z}, \mathbf{x}, \mathbf{A}, \mathbf{B}, \{v_j\}, \boldsymbol{\theta}) \propto p(\mathbf{z}_0|\psi_{\text{MSV}})p(\mathbf{z}|\mathbf{x}, \mathbf{A}, \mathbf{B}, \{v_j\}) \quad (\text{B.8})$$

where the latter density is proportional to  $\exp[-\frac{1}{2} \sum_{t=1}^T (\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1})' \boldsymbol{\Sigma}_t^{-1} (\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1})]$ .

This density is also of a non-standard form and, thus, requires the MH algorithm. In this case the candidate draw,  $\mathbf{z}_0^{(g+1)}$ , is generated according to the process  $\mathbf{z}_0^{(g+1)} = \mathbf{z}_0^{(g)} + \mathbf{c}$  where  $\mathbf{c}$  is a vector simulated from a normal distribution with mean zero and covariance matrix equal to  $\boldsymbol{\Sigma}_0^{(g)}(\mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)})$  that is, the unconditional covariance matrix of the errors as a function of the current draws for  $\mathbf{B}$  and  $\boldsymbol{\theta}$ . Since the proposal draw is equal to the current value plus noise, this implementation of the MH algorithm is called a random walk chain. The probability of acceptance is given by

$$\alpha_{z_0} = \min \left( 1, \frac{p(\mathbf{z}|\mathbf{x}^{(g+1)}, \mathbf{A}^{(g)}, \mathbf{B}^{(g)}, \{v_j\}^{(g)}) p(\mathbf{z}_0^{(g+1)}|\psi_{\text{MSV}}^{(g)})}{p(\mathbf{z}|\mathbf{x}^{(g)}, \mathbf{A}^{(g)}, \mathbf{B}^{(g)}, \{v_j\}^{(g)}) p(\mathbf{z}_0^{(g)}|\psi_{\text{MSV}}^{(g)})} \right) \quad (\text{B.9})$$

where  $\mathbf{x}^{(\cdot)}$  is  $\mathbf{x}$  with the first row set equal to  $\mathbf{z}_0^{(\cdot)}$ . We find that the proposal density described above works quite well for the problem at hand, as it generates draws whose acceptance rates are steadily between 50 and 60 percent.

## B.6 Posterior Sampling under Prior 4

The features of interest prior implicitly imposes additional structure on the priors for  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\boldsymbol{\theta}$ . This is because the functions of interest (in this case, the variance decomposition ratios) are functions of the model's parameters. We choose to impose a prior on the unconditional variance decomposition terms. Since the three ratios sum to one by construction, only two priors are necessary. As we have a better sense for the variance terms than for the covariance term, we impose independent normal priors on  $\text{Var}(\eta_{d,t+1})/\text{Var}(\eta_{t+1})$  and on  $\text{Var}(\eta_{h,t+1})/\text{Var}(\eta_{t+1})$ . Let  $p(\Psi(\mathbf{A}, \mathbf{B}, \boldsymbol{\theta}))$  denote the joint prior, where  $\Psi(\cdot)$  is the functional form linking the parameters to the decomposition terms. We first assume a prior mean equal to 1/3 for the ratios implying that, a priori, innovations in dividends and innovations in expected returns (and, consequently, the covariance term as well) have the same impact on the variance of returns. To allow the data to move the prior view, we set the prior variance to 2. We then experiment with prior means of 1/2 and 1/2 and the same prior variance. Although clearly subjective, we perceive these choices as economically sensible. Hollifield, Koop, and Li (2003) propose a similar approach, although their priors are significantly tighter (i.e., more informed) than ours.

Under the features of interest prior, the posterior distributions are modified by the inclusion of one additional term for each parameter block. As a consequence, we can still use the Metropolis-Hastings algorithm to generate posterior draws, as described for Prior 3. We can use the same candidate densities, but the acceptance probabilities are different in this case because of the additional prior term.

It should be noted that the specific functional form for the priors is not dictated by computational convenience. Since the posterior sampling is done through the Metropolis-Hastings algorithm, any functional form could be used. We next provide details for the sampling of the  $\mathbf{A}$  matrix. Similar strategies are applied for  $\mathbf{B}$  and  $\boldsymbol{\theta}$ . No changes are necessary for the posterior sampling of  $\mathbf{z}_0$  used under Prior 3.

### Sampling of $\mathbf{A}$ under Prior 4

The posterior density is proportional to  $\pi(\boldsymbol{\alpha}|\mathbf{z}, \mathbf{B}, \{v_j\})q(\mathbf{A}, \mathbf{B}, \boldsymbol{\theta})p(\Psi(\mathbf{A}, \mathbf{B}, \boldsymbol{\theta}))$ . Let  $\mathbf{A}^{(g)}$  denote the value at iteration  $g$  of the Gibbs sampler, and let  $\mathbf{A}^{(*)}$  denote the candidate draw at Gibbs iteration  $g + 1$ . The latter is generated from (B.2) and it is accepted with probability  $\alpha_A$  given by

$$\alpha_A = \min \left( 1, \frac{q(\mathbf{A}^{(*)}, \mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)}) p(\Psi(\mathbf{A}^{(*)}, \mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)}))}{q(\mathbf{A}^{(g)}, \mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)}) p(\Psi(\mathbf{A}^{(g)}, \mathbf{B}^{(g)}, \boldsymbol{\theta}^{(g)}))} \right) \quad (\text{B.10})$$

To compute  $p(\Psi(\cdot))$ , the unconditional variance decomposition ratios are first computed from the relevant parameter values and their prior densities are evaluated at the calculated points.

## References

- Ammer, John, and Jianping Mei, 1996, Measuring international economic linkages with stock market data, *Journal of Finance* 51, 1743–1763.
- Ang, Andrew, and Geert Bekaert, 2004, Stock return predictability: Is it there?, Working Paper, Columbia University.
- Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.
- Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner, 1992, ARCH modeling in finance, *Journal of Econometrics* 52, 5–59.
- Campbell, John Y., 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.
- , 1993, Intertemporal asset pricing without consumption data, *American Economic Review* 183, 487–512.
- , 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.
- , and John Ammer, 1993, What moves the stock and bond markets? A variance decomposition for long-term asset returns, *Journal of Finance* 48, 3–37.
- Campbell, John Y., and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.
- , 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.
- Campbell, John Y., and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.
- Campbell, John Y., and Motohiro Yogo, 2005, Efficient tests of stock return predictability, Working Paper, Harvard University.
- Chen, Joseph, 2003, Intertemporal CAPM and the cross-section of stock returns, Working Paper, University of Southern California.

- Chib, Siddhartha, Federico Nardari, and Neil Shephard, 2002, Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics* 108, 281–316.
- , 2005, Analysis of high dimensional multivariate stochastic volatility models, forthcoming in *Journal of Econometrics*.
- Cochrane, John H., 2001, *Asset Pricing* (Princeton University Press).
- Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.
- Fama, Eugene F., and G. William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 1995, *Bayesian Data Analysis* (Chapman and Hall: London).
- Ghysels, Eric, Andrew C. Harvey, and Eric Renault, 1996, Stochastic volatility, in G. S. Maddala, and C. R. Rao, ed.: *Handbook of Statistics*, vol. 14 . pp. 119–191 (North-Holland: Amsterdam).
- Goyal, Amit, and Ivo Welch, 2003, Predicting the equity premium with dividend ratios, *Management Science* 49, 639–654.
- Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press: Princeton, New Jersey).
- Harvey, Andrew, Esther Ruiz, and Neil Shephard, 1994, Multivariate stochastic variance models, *Review of Economic Studies* 61, 247–264.
- Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.
- Hollifield, Burton, Gary Koop, and Kai Li, 2003, A Bayesian analysis of a variance decomposition for stock returns, *Journal of Empirical Finance* 10, 583–601.
- Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, 1994, Bayesian analysis of stochastic volatility models (with discussion), *Journal of the American Statistical Association* 12, 371–417.
- Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385–424.

- Kim, Myung Jig, Charles R. Nelson, and Richard Startz, 1991, Mean reversion in stock prices? A reappraisal of the empirical evidence, *Review of Economic Studies* 58, 515–528.
- Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, 1998, Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361–393.
- Lamont, Owen A., 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.
- , and Christopher Polk, 2001, The diversification discount: Cash flows versus returns, *Journal of Finance* 56, 1693–1721.
- Lettau, Martin, and Sydney C. Ludvigson, 2005, Expected returns and expected dividend growth, Forthcoming in *Journal of Financial Economics*.
- Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.
- Mankiw, N. Gregory, and Matthew D. Shapiro, 1986, Do we reject too often? Small sample properties of tests of rational expectations models, *Economic Letters* 20, 139–145.
- Menzly, Lior, Tano Santos, and Pietro Veronesi, 2004, Understanding predictability, *Journal of Political Economy* 112, 1–47.
- Nelson, Charles R., and Myung Jig Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.
- Poirier, Dale J., 1995, *Intermediate Statistics and Econometrics: A Comparative Approach* (MIT Press: Cambridge, MA).
- Schwert, G. William, 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115–1153.
- Scruggs, John T., 1998, Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach, *Journal of Finance* 53, 575–603.
- Shanken, Jay, and Ane Tamayo, 2004, Dividend yield, risk, and mispricing: A Bayesian analysis, Working Paper, Emory University.

- Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457–498.
- Sims, Christopher A., 1988, Bayesian skepticism on unit root econometrics, *Journal of Economic Dynamics and Control* 12, 463–474.
- , and Harald Uhlig, 1991, Understanding unit rooters: A helicopter tour, *Econometrica* 59, 1591–1599.
- Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Taylor, Stephen J., 1986, *Modelling Financial Time Series* (John Wiley: Chichester).
- , 1994, Modeling stochastic volatility: A review and comparative study, *Mathematical Finance* 4, 183–204.
- Torous, Walter, Rossen Valkanov, and Shu Yan, 2005, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.
- Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (J. Wiley and Sons, Inc.: New York).



**Table I**  
**Posterior Estimates of Parameters: Homoskedastic VAR Models**

This table reports summary statistics describing the Bayesian posterior densities for parameters of the homoskedastic VAR model:

$$\begin{aligned} \mathbf{z}_{t+1} &= \mathbf{A}\mathbf{z}_t + \mathbf{w}_{t+1} \\ \mathbf{w}_{t+1} &\sim \mathcal{N}_k(\mathbf{0}, \Sigma). \end{aligned}$$

The sample period is 1952:1–2002:12. The mean and standard deviation (in parentheses) of the posterior density are reported for each parameter. The posterior densities are based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain.

Prior 1: Normal-Wishart; stationarity not imposed;  $\mathbf{z}_0$  deterministic.

Prior 2: Normal-Wishart; stationarity imposed;  $\mathbf{z}_0$  deterministic.

Prior 3: Normal-Wishart; stationarity imposed;  $\mathbf{z}_0$  stochastic.

Prior 4: Normal-Wishart; stationarity imposed;  $\mathbf{z}_0$  stochastic; features of interest prior.

Dependent Variable	$h_t$	$\mathbf{A}$			$\Sigma$			$R^2$
		$(D/P)_t$	$rb_t$	$h_{t+1}$	$(D/P)_{t+1}$	$rb_{t+1}$		
<i>Panel A: Prior 1</i>								
$h_{t+1}$	0.026 (0.040)	0.351 (0.150)	-0.594 (0.144)	16.657 (0.950)	-0.617 (0.037)	-0.256 (0.127)	0.030 (0.004)	
$(D/P)_{t+1}$	-0.001 (0.002)	0.986 (0.006)	0.033 (0.006)	-0.617 (0.037)	0.029 (0.002)	0.010 (0.005)	0.978 (0.000)	
$rb_{t+1}$	0.013 (0.008)	-0.030 (0.028)	0.756 (0.027)	-0.256 (0.127)	0.010 (0.005)	0.597 (0.034)	0.558 (0.002)	
<i>Panel B: Prior 2</i>								
$h_{t+1}$	0.027 (0.040)	0.354 (0.147)	-0.596 (0.144)	16.655 (0.948)	-0.617 (0.037)	-0.257 (0.128)	0.030 (0.004)	
$(D/P)_{t+1}$	-0.001 (0.002)	0.986 (0.006)	0.033 (0.006)	-0.617 (0.037)	0.029 (0.002)	0.010 (0.005)	0.978 (0.000)	
$rb_{t+1}$	0.013 (0.008)	-0.030 (0.028)	0.756 (0.027)	-0.257 (0.128)	0.010 (0.005)	0.597 (0.034)	0.558 (0.002)	
<i>Panel C: Prior 3</i>								
$h_{t+1}$	0.027 (0.040)	0.316 (0.139)	-0.598 (0.143)	16.674 (0.952)	-0.618 (0.038)	-0.260 (0.127)	0.030 (0.004)	
$(D/P)_{t+1}$	0.000 (0.002)	0.988 (0.006)	0.034 (0.006)	-0.618 (0.038)	0.029 (0.002)	0.010 (0.005)	0.978 (0.000)	
$rb_{t+1}$	0.013 (0.008)	-0.025 (0.027)	0.756 (0.027)	-0.260 (0.127)	0.010 (0.005)	0.597 (0.034)	0.558 (0.002)	
<i>Panel D: Prior 4</i>								
$h_{t+1}$	0.026 (0.040)	0.312 (0.139)	-0.599 (0.144)	16.680 (0.953)	-0.618 (0.037)	-0.260 (0.127)	0.030 (0.004)	
$(D/P)_{t+1}$	0.000 (0.002)	0.988 (0.006)	0.034 (0.006)	-0.618 (0.037)	0.029 (0.002)	0.010 (0.005)	0.978 (0.000)	
$rb_{t+1}$	0.013 (0.008)	-0.026 (0.027)	0.756 (0.027)	-0.260 (0.127)	0.010 (0.005)	0.597 (0.034)	0.558 (0.002)	

**Table II**  
**Variance Decomposition for Real Stock Returns: Homoskedastic Models**

This table reports summary statistics describing the Bayesian posterior densities for components of the variance of unexpected real stock returns and other features of interest. The sample period is 1952:1–2002:12. The mean, median (in braces), standard deviation (in parentheses), and 5% and 95% quantiles (in brackets) are reported for each parameter. The posterior densities are based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain.

Prior 1: Normal-Wishart; stationarity not imposed;  $\mathbf{z}_0$  deterministic.

Prior 2: Normal-Wishart; stationarity imposed;  $\mathbf{z}_0$  deterministic.

Prior 3: Normal-Wishart; stationarity imposed;  $\mathbf{z}_0$  stochastic.

Prior 4: Normal-Wishart; stationarity imposed;  $\mathbf{z}_0$  stochastic; features of interest prior.

Share of $\text{Var}(\eta)$ due to:				
$\text{Var}(\eta_d)$	$\text{Var}(\eta_h)$	$-2\text{Cov}(\eta_d, \eta_h)$	$\text{Corr}(\eta_d, \eta_h)$	$P_h$
<i>Panel A: Prior 1</i>				
9.686	10.210	-18.895	-0.027	7.964
{0.235}	{0.735}	{0.045}	{-0.057}	{7.074}
(1753)	(1757)	(3510)	(0.320)	(24.450)
[0.195, 0.447]	[0.329, 1.407]	[-0.770, 0.327]	[-0.498, 0.549]	[4.295, 12.539]
<i>Panel B: Prior 2</i>				
0.267	0.786	-0.054	-0.033	7.519
{0.234}	{0.737}	{0.046}	{-0.058}	{7.065}
(0.165)	(0.381)	(0.471)	(0.312)	(2.633)
[0.194, 0.421]	[0.331, 1.371]	[-0.715, 0.326]	[-0.496, 0.528]	[4.294, 12.196]
<i>Panel C: Prior 3</i>				
0.283	0.821	-0.104	-0.001	7.587
{0.239}	{0.754}	{0.026}	{-0.032}	{7.098}
(0.191)	(0.443)	(0.554)	(0.340)	(2.785)
[0.196, 0.492]	[0.302, 1.518]	[-0.911, 0.330]	[-0.505, 0.597]	[4.141, 12.561]
<i>Panel D: Prior 4</i>				
0.261	0.739	0.000	-0.068	7.252
{0.236}	{0.711}	{0.070}	{-0.090}	{6.844}
(0.095)	(0.302)	(0.307)	(0.304)	(2.493)
[0.195, 0.406]	[0.291, 1.274]	[-0.582, 0.337]	[-0.520, 0.468]	[4.076, 11.720]

**Table III**  
**Posterior Estimates of Parameters: VAR-MSV Models**

This table reports summary statistics describing the Bayesian posterior densities for parameters of the VAR-MSV model:

$$\begin{aligned} \mathbf{z}_{t+1} &= \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{V}_{t+1}^{1/2}\boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1} &\sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}) \\ \mathbf{V}_{t+1} &= \text{diag}\{\exp(v_{1,t+1}), \dots, \exp(v_{k,t+1})\} \\ v_{j,t+1} &= \mu_j + \phi_j(v_{j,t} - \mu_j) + \sigma_j\zeta_{j,t+1} \end{aligned}$$

The sample period is 1952:1–2002:12. The mean and standard deviation (in parentheses) are reported for each parameter. The posterior densities are based on 50,000 Gibbs sampler iterations from a suitably constructed Markov chain.

Prior 1: Normal-MSV; stationarity not imposed;  $\mathbf{z}_0$  deterministic.

Prior 2: Normal-MSV; stationarity imposed;  $\mathbf{z}_0$  deterministic.

Prior 3: Normal-MSV; stationarity imposed;  $\mathbf{z}_0$  stochastic.

Prior 4: Normal-MSV; stationarity imposed;  $\mathbf{z}_0$  stochastic; features of interest prior.

Dependent Variable	$h_t$	<b>A</b>			<b>B</b>			$\exp(\mu_j/2)$	$\phi_j$	$\sigma_j$	$R^2$
		$(D/P)_t$	$rb_t$	$h_{t+1}$	$(D/P)_{t+1}$	$rb_{t+1}$					
<i>Panel A: Prior 1</i>											
$h_{t+1}$	0.008 (0.042)	0.264 (0.149)	-0.527 (0.146)	1	0	0	3.799 (0.268)	0.917 (0.040)	0.221 (0.056)	0.029 (0.005)	
$(D/P)_{t+1}$	0.000 (0.002)	0.993 (0.006)	0.028 (0.005)	-0.034 (0.001)	1	0	0.058 (0.013)	0.975 (0.013)	0.223 (0.044)	0.977 (0.000)	
$rb_{t+1}$	0.007 (0.006)	0.005 (0.020)	0.784 (0.028)	-0.011 (0.006)	0.233 (0.360)	1	0.554 (0.096)	0.960 (0.018)	0.282 (0.047)	0.556 (0.003)	
<i>Panel B: Prior 2</i>											
$h_{t+1}$	0.009 (0.042)	0.297 (0.130)	-0.523 (0.147)	1	0	0	3.799 (0.279)	0.916 (0.041)	0.221 (0.056)	0.029 (0.004)	
$(D/P)_{t+1}$	0.000 (0.002)	0.991 (0.005)	0.028 (0.005)	-0.034 (0.001)	1	0	0.059 (0.015)	0.975 (0.012)	0.222 (0.044)	0.978 (0.000)	
$rb_{t+1}$	0.007 (0.006)	0.002 (0.019)	0.784 (0.028)	-0.011 (0.006)	0.230 (0.362)	1	0.550 (0.097)	0.960 (0.017)	0.282 (0.046)	0.557 (0.003)	
<i>Panel C: Prior 3</i>											
$h_{t+1}$	0.008 (0.042)	0.299 (0.128)	-0.525 (0.146)	1	0	0	3.804 (0.261)	0.916 (0.041)	0.221 (0.055)	0.029 (0.004)	
$(D/P)_{t+1}$	0.000 (0.002)	0.991 (0.005)	0.028 (0.005)	-0.034 (0.001)	1	0	0.059 (0.015)	0.975 (0.013)	0.225 (0.045)	0.978 (0.000)	
$rb_{t+1}$	0.007 (0.006)	0.005 (0.018)	0.784 (0.028)	-0.010 (0.006)	0.154 (0.361)	1	0.538 (0.093)	0.960 (0.017)	0.283 (0.047)	0.557 (0.002)	
<i>Panel D: Prior 4</i>											
$h_{t+1}$	0.008 (0.042)	0.295 (0.129)	-0.526 (0.146)	1	0	0	3.833 (0.275)	0.917 (0.041)	0.222 (0.056)	0.029 (0.004)	
$(D/P)_{t+1}$	0.000 (0.002)	0.991 (0.005)	0.028 (0.005)	-0.034 (0.001)	1	0	0.054 (0.010)	0.973 (0.013)	0.224 (0.045)	0.978 (0.000)	
$rb_{t+1}$	0.007 (0.006)	0.005 (0.018)	0.783 (0.028)	-0.010 (0.006)	0.157 (0.363)	1	0.537 (0.090)	0.960 (0.017)	0.282 (0.046)	0.557 (0.002)	

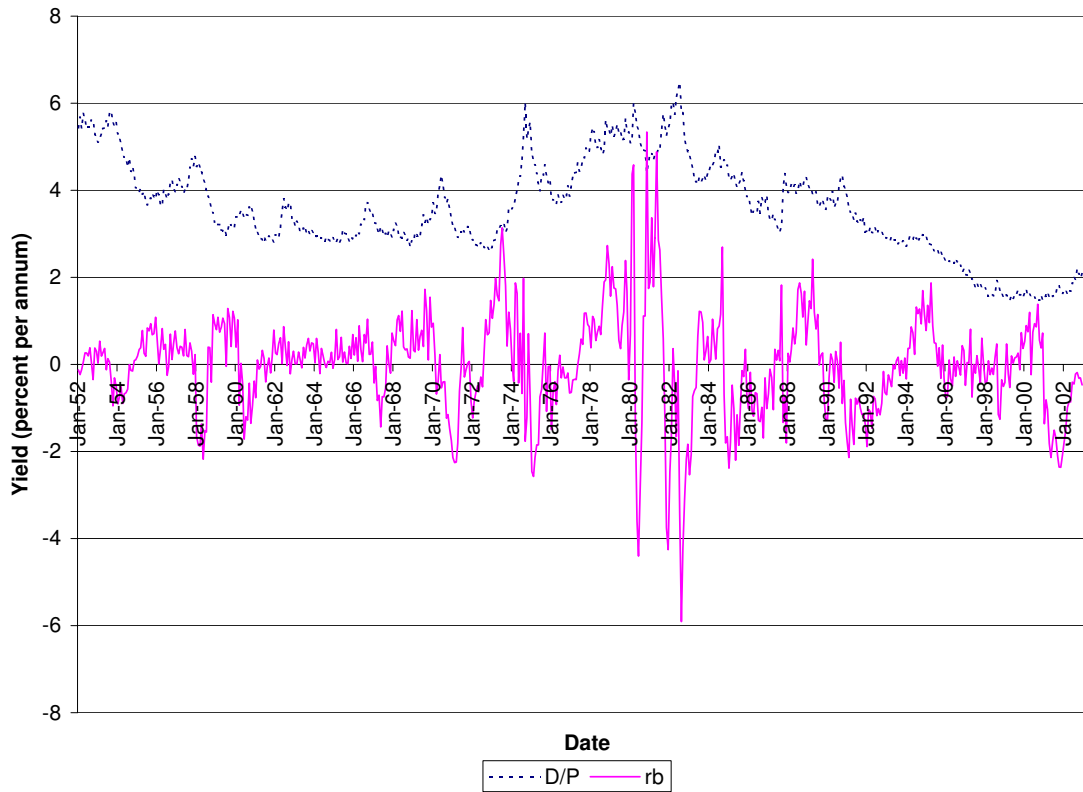


Figure 1: **Predictive Variables.** Time-series plots of the dividend-price ratio and relative bill rate (the one-month T-bill yield less its twelve-month moving average) for the period 1952:1-2002:12.

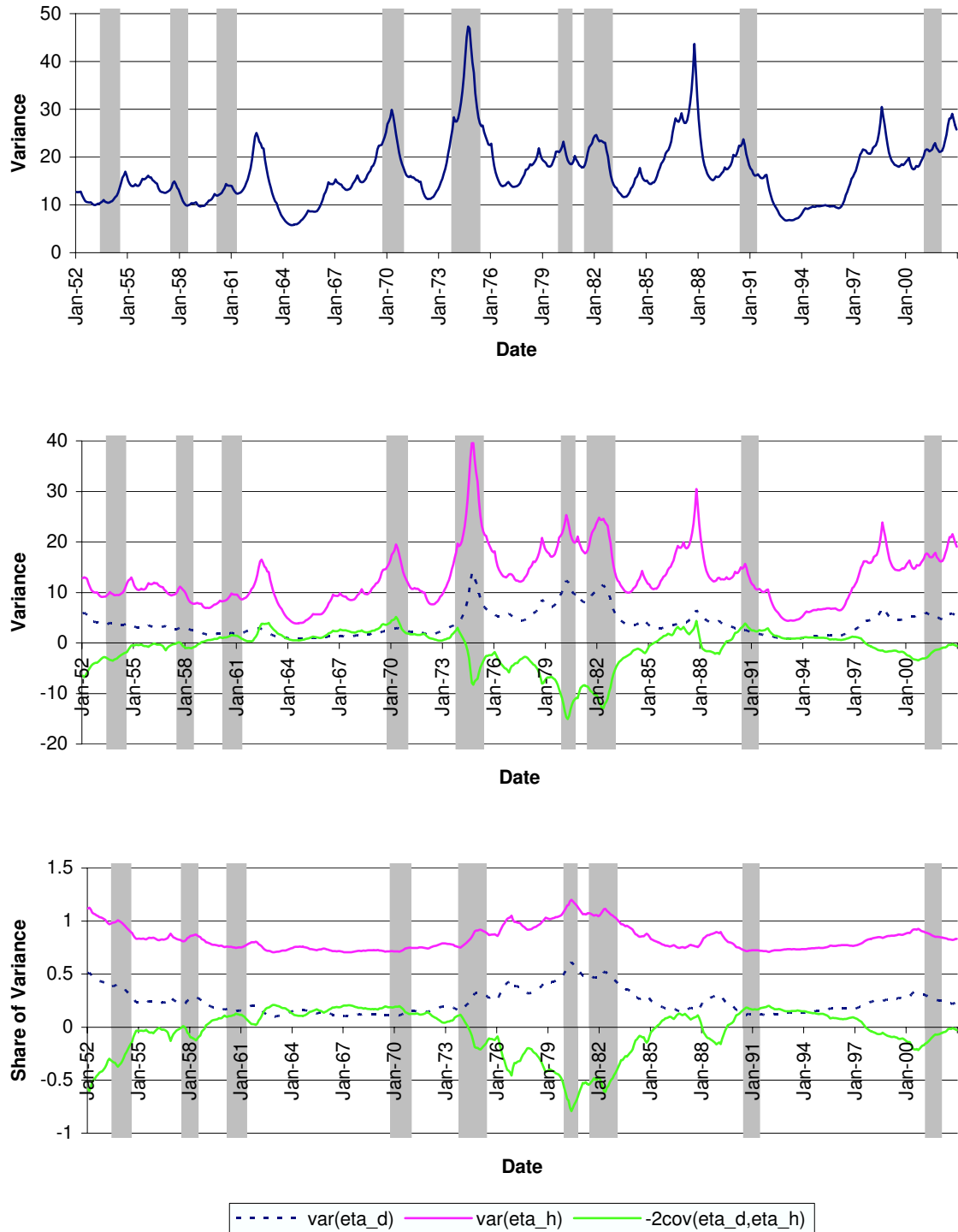


Figure 2: **Variance Decomposition for Unexpected Real Stock Returns.** Top: Time-series plot of the variance of unexpected real stock returns,  $\text{Var}(\eta_{t+1})$ . Middle: Plots of  $\text{Var}(\eta_{d,t+1})$ ,  $\text{Var}(\eta_{h,t+1})$ , and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$ . Bottom: Plots of shares of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{d,t+1})$ ,  $\text{Var}(\eta_{h,t+1})$ , and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$ . The sample period is 1952:1-2002:12. Vertical gray bars indicate NBER recessions.

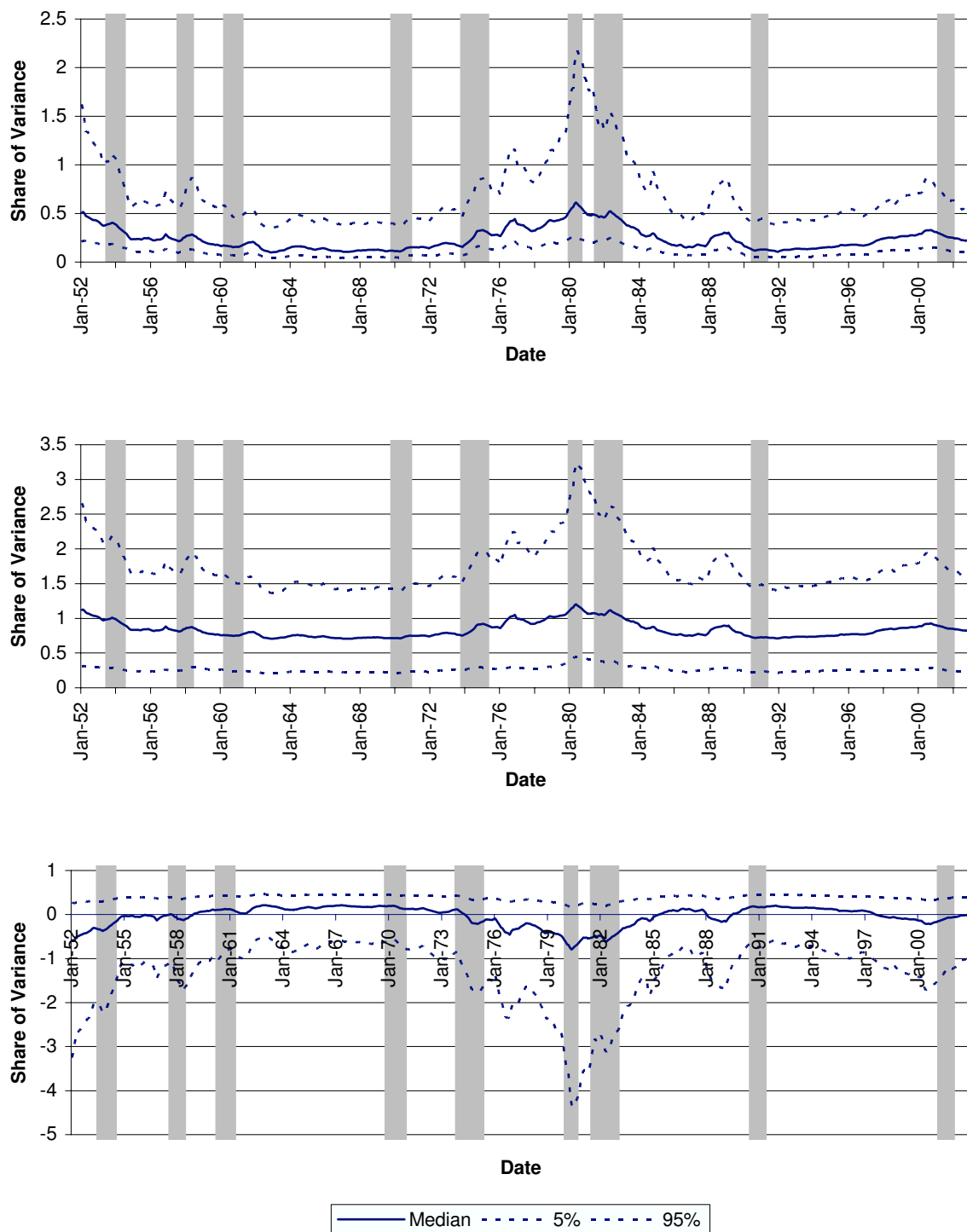


Figure 3: **Posterior Densities of Variance Component Shares.** Plots of median, 5% and 95% quantiles of posterior densities for shares of the variance of  $\text{Var}(\eta_{t+1})$  attributable to  $\text{Var}(\eta_{d,t+1})$  (top),  $\text{Var}(\eta_{h,t+1})$  (middle), and  $-2\text{Cov}(\eta_{d,t+1}, \eta_{h,t+1})$  (bottom). The sample period is 1952:1-2002:12. Vertical grey bars indicate NBER recessions.

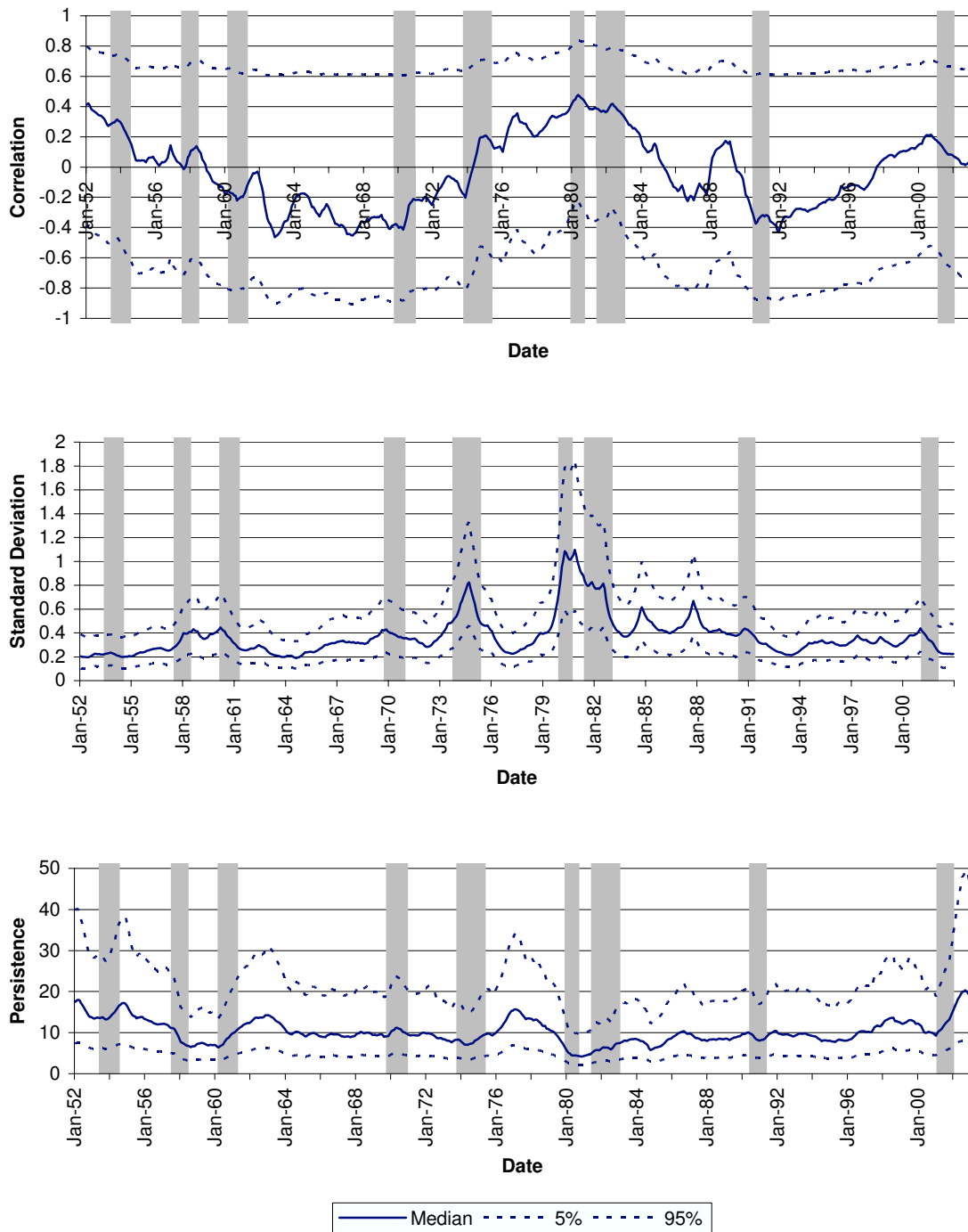


Figure 4: **Posterior Densities of Additional Features of Interest.** Plots of median, 5% and 95% quantiles of posterior densities for  $\text{Corr}(\eta_{d,t+1}, \eta_{h,t+1})$  (top),  $\sigma(u_{t+1})$  (middle), and  $P_{h,t+1}$  (bottom). The sample period is 1952:1-2002:12. Vertical gray bars indicate NBER recessions.