

BAYESIAN VECTOR AUTOREGRESSIONS WITH STOCHASTIC VOLATILITY¹

By Harald Uhlig

February 2, 1996

Abstract

This paper proposes a Bayesian approach to a vector autoregression with stochastic volatility, where the multiplicative evolution of the precision matrix is driven by a multivariate beta variate. Exact updating formulas are given to the nonlinear filtering of the precision matrix. Estimation of the autoregressive parameters requires numerical methods: an importance-sampling based approach is explained here.

1 Introduction

This paper introduces Bayesian vector autoregressions with stochastic volatility. In contrast to multivariate autoregressive conditional heteroskedasticity (ARCH), the stochastic volatility setup here models the error precision matrix as an unobserved component with shocks drawn from a multivariate beta distribution. This allows the interpretation of a sudden large movement in the data as the result of a draw from a distribution with a randomly increased but unobserved variance. Exploiting a conjugacy between Wishart distributions and multivariate singular beta distributions, the integration over the unobserved shock to the precision matrix can be performed in closed form, leading to a generalization of the standard Kalman-Filter formulas to the nonlinear filtering problem at hand. Estimating the autoregressive parameters requires numerical methods, however. The paper focusses on an importance-sampling based approach.

Bayesian vector autoregressions have been studied and popularized by e.g. Litterman (1979), Doan, Litterman and Sims (1984) and Doan's RATS Manual (1990). ARCH models have been introduced by Engle (1982), see the review in Bollerslev, Chou and Kroner (1992). Stochastic volatility models provide an alternative approach to model time variation in the size of fluctuations. The stochastic volatility model used here is similar to Shephard (1994), whose model is a univariate, non-Bayesian and non-autoregressive special case of the model proposed here. In contrast to other Bayesian approaches to stochastic volatility, see Jacquier, Polson and Rossi (1994), the method here results in exact updating formulas for the posterior in the sense that the integration over the unobserved shocks to the precision matrices is done in closed form. The conjugacy result needed for this step is established in Uhlig (1994b).

For simplicity, the main ideas are explained in section 2 for the univariate case with the general case presented in section 3. Section 4 discusses how to analyze the posterior numerically. Section 5 concludes. Appendix A lists some of the distributions used and fixes notation. Appendix B contains the proofs and one additional theorem. Appendix C proposes a prior.

2 A Simple Case

Consider the following simple version of the model studied in this paper:

$$(1) \quad y_t = \beta y_{t-1} + h_t^{-1/2} \epsilon_t, \text{ with } \epsilon_t \sim \mathcal{N}(0, 1),$$

$$(2) \quad h_{t+1} = h_t \vartheta_t / \lambda, \text{ with } \vartheta_t \sim \mathcal{B}_1((\nu + 1)/2, 1/2),$$

where all ϑ_t 's and ϵ_t 's drawn independently, where $t = 1, \dots, T$ denotes time, $y_t \in \mathbf{R}$, $t = 0, \dots, T$ is data and observable, $\lambda > 0$, $\nu > 0$ are parameters and $\mathcal{B}_1(p, q)$ denotes the (one-dimensional) beta-distribution on the interval $[0, 1]$.

Equation (2) specifies the unobserved precision h_t of the innovation $h_t^{-1/2}\epsilon_t$ to be stochastic. The model thus belongs to the family of stochastic volatility models, see e.g. Jacquier-Polson and Rossie (1994). The model captures autocorrelated heteroskedasticity, a feature often found especially in financial data series. Another popular specification which does so is the ARCH-family of models. A GARCH(1,1) model, for example, replaces $h_t^{-1/2}$ in (1) with σ_t and replaces (2) with

$$(3) \quad \sigma_{t+1}^2 = \iota + \lambda\sigma_t^2 + \mu\sigma_t^2\epsilon_t^2,$$

where ι , λ and μ are parameters. It thus ties the innovation in the variance to the size of the current innovation $\sigma_t\epsilon_t$. Given ϵ_{t-1} and h_{t-1} or σ_{t-1} , an unusually large innovation in (2) can result from a randomly decreased h_t as well as a large ϵ_t , whereas the GARCH-model (3) only allows for an unusually large draw ϵ_t .

To analyze the system (1) to (2) in a Bayesian fashion, one needs to choose a prior density $\pi_0(\beta, h_1)$ for β and h_1 , given y_0 . The goal is to find the posterior density $\pi_T(\beta, h_{T+1})$ given data y_0, \dots, y_T . We restrict the choice of priors to be of the following form. Fix $\lambda > 0$ and $\nu > 0$ (for a more general treatment, see section 3). Choose $\bar{b}_0 \in \mathbf{R}$, $n_0 > 0$, $s_0^2 > 0$ and a function $g_0(\beta) \geq 0$ to describe a prior density proportional to

$$\pi_0(\beta, h_1) \propto g_0(\beta)f_{NG}(\beta, h_1 \mid \bar{b}_0, \lambda n_0, s_0, \nu),$$

where f_{NG} denotes the Normal-gamma density, see appendix A. The form of the prior allows for a flexible treatment of a root near or above unity via the function $g_0(\beta)$, see Uhlig (1994a).

Adapting the Bayesian updating formulas (12), (13), (14) and (15) derived below in section 3 to the simple model above results in

$$(4) \quad n_t = \lambda n_{t-1} + y_{t-1}^2$$

$$(5) \quad \bar{b}_t = \left(\lambda \bar{b}_{t-1} n_{t-1} + y_t y_{t-1} \right) / n_t$$

$$(6) \quad s_t = \lambda s_{t-1} + \frac{\lambda}{\nu} e_t^2 \left(1 - y_{t-1}^2 / n_t \right),$$

where

$$e_t = y_t - \bar{b}_{t-1} y_{t-1},$$

and

$$(7) \quad g_t(\beta) = g_{t-1}(\beta) \left((\beta - \bar{b}_t)^2 n_t + \frac{\nu}{\lambda} s_t \right)^{-1/2}$$

for $t = 1, \dots, T$. These deliver the posterior density

$$\pi_T(\beta, h_{T+1}) \propto g_T(\beta)f_{NG}(\beta, h_{T+1} \mid \bar{b}_T, \lambda n_T, s_T, \nu).$$

Equations (4) and (5) are the recursion formulas or Kalman Filter formulas for geometrically weighted least squares. Different observations receive different weights according to the size of s_t via equation (7). Equation (6) prescribes to find the “estimate” s_t of h_{t+1} essentially via a geometric lag on past squared residuals. Notice the formal similarity to GARCH: ignoring the term $(1 - y_{t-1}^2/n_t)$, equation (6) resembles equation (3) rewritten in terms of observables, using $\iota = 0$ and $\mu = \lambda/\nu$.

The key for proving the validity of these updating formulas here or in the next section is theorem 2 and its proof (see appendix B): as the unobserved shock ϑ_t occurs, one needs to do a “change of variable” from $d\beta dh_t d\vartheta_t$ to $d\beta dh_{t+1} dz_t$ for some suitably defined z_t . Thanks to the conjugacy between the beta and the gamma distribution, integration over dz_t can be performed in closed form, resulting in an integration constant depending on β and the data. This constant is captured by the function $g_t(\beta)$.

Shephard (1994) finds similar formulas with a classical interpretation for (1) to (2) without the autoregressive term βy_{t-1} . To include autoregressive terms, Shephard (1994) suggests approximate filtering formulas. In contrast, the Bayesian formulas here are exact. They do, however, require numerical techniques such as importance-sampling for the estimation of β . There is no treatment of the multivariate case in Shephard (1994).

For $\lambda = \nu/(\nu + 1)$ we have $\lambda/\nu = 1 - \lambda$ in equation (6). For $\lambda = (\nu + 1)/(\nu + 2)$ the precision h_t is a martingale $E[h_{t+1} | h_t] = h_t$ on the positive part of the real axis. Shephard (1994) suggests setting $\lambda = e^r$, where $r = E[\log \vartheta_t]$. This avoids the problem, that otherwise $h_t \rightarrow \infty$ a.s. or $h_t \rightarrow 0$ a.s. (see Nelson (1990)) and makes $\log h_t$ a random walk.

For $\nu \rightarrow \infty$ one obtains a model where h_1 is known a priori, $h_1 = s_0^{-1}$, and where $h_{t+1} = h_t(\nu + 1)/(\lambda(\nu + 2))$. In other words the model allows for the greater time variation in the precision, the smaller the parameter ν .

Figure 1 shows parts of the densities for λ/ϑ_t , which are the multiplicative disturbances of the variance $\sigma_t^2 \equiv h_t^{-1}$. It shows that (2) typically leads to a slight decrease in the innovation variance except for occasional and potentially large increases.

3 The General Model

Consider the VAR(k)-model with time-varying error precision matrices

$$(8) \quad Y_t = B_{(0)}C_t + B_{(1)}Y_{t-1} + B_{(2)}Y_{t-2} + \dots + B_{(k)}Y_{t-k} + \mathcal{U}(H_t^{-1})'\epsilon_t, \text{ with } \epsilon_t \sim \mathcal{N}(0, I_m),$$

$$(9) \quad H_{t+1} = \mathcal{U}(H_t)'\Theta_t\mathcal{U}(H_t) / \lambda, \text{ with } \Theta_t \sim \mathcal{B}_m((\nu + c + km)/2, 1/2),$$

where $t = 1, \dots, T$ denotes time, Y_t , $t = 1 - k, \dots, T$, size $m \times 1$, is observable data, and C_t , size $c \times 1$, denotes deterministic regressors such as a constant and a time trend. The coefficient matrix $B_{(0)}$ is of size $m \times c$, the coefficient matrices $B_{(i)}$, $i = 1, \dots, k$ are

of size $m \times m$, $\nu > m - 1$ and $\lambda > 0$ are parameters and all $\epsilon_t, t = 1, \dots, T$, size $m \times 1$, and $\Theta_t, t = 1, \dots, T$, size $m \times m$ are independently distributed. $\mathcal{U}(H)$ denotes the upper Cholesky factor of a positive definite matrix H and $\mathcal{B}_m(p, q)$ denotes the multivariate beta-distribution. This distribution has been chosen for Θ_t to exploit a conjugacy between that distribution and the Wishart distribution. The distribution $\mathcal{B}_m(p, q)$ is traditionally only defined for $p > (m - 1)/2$ and $q > (m - 1)/2$, see Muirhead (1982). This definition has been extended along with the conjugacy results by Uhlig (1994b) to allow for values $q = n/2$ for any integer $n \geq 1$, see appendix A for the details pertinent to this paper. Equation (9) is one of two rather natural generalizations of the multiplication of two real numbers in equation (2) in order to guarantee the symmetry of the resulting matrix H_{t+1} . The other natural generalization switches H_t and Θ_t in (9): it turns out that only (9) works for the proof of the updating formulas.

Let $X_t = [C_t Y'_{t-1} Y'_{t-2} \dots Y'_{t-k}]'$ and $B = [B_{(0)} B_{(1)} B_{(2)} \dots B_{(k)}]$. Rewrite the first equation (8) more concisely as

$$Y_t = BX_t + \mathcal{U}(H_t^{-1})'\epsilon_t \text{ with } \epsilon_t \sim \mathcal{N}(0, I_m),$$

Let $l = c + km$. Consider the following algorithm, which generalizes the Kalman Filter to the nonlinear filtering problem at hand by including an exact updating formula for the mean precision matrix. This algorithm finds the posterior²

$$\pi_T(\nu, \lambda, B, H_{T+1}) \propto \frac{\psi_T(\nu, \lambda)}{\kappa_{NW}(\lambda N_T, S_T, \nu, m)} g_T(B) f_{NW}(B, H_{T+1} | \bar{B}_T, \lambda N_T, S_T, \nu)$$

in ν, λ, B and H_{T+1} for a prior $\pi_0(\nu, \lambda, B, H_1)$ of the form

$$\pi_0(\nu, \lambda, B, H_1) \propto \frac{\psi_0(\nu, \lambda)}{\kappa_{NW}(\lambda N_0, S_0, \nu, m)} g_0(B) f_{NW}(B, H_1 | \bar{B}_0, \lambda N_0, S_0, \nu),$$

given the initial observations $Y_t, t = 1 - k, \dots, 0$. If ν and λ are treated as known, prior and posterior are

$$(10) \quad \pi_0(B, H_1) \propto g_0(B) f_{NW}(B, H_1 | \bar{B}_0, \lambda N_0, S_0, \nu)$$

$$(11) \quad \pi_T(B, H_{T+1}) \propto g_T(B) f_{NW}(B, H_{T+1} | \bar{B}_T, \lambda N_T, S_T, \nu).$$

The General Method

1. *To fix the prior, choose a $m \times l$ -matrix \bar{B}_0 , a positive definite $l \times l$ -matrix N_0 and a positive definite $m \times m$ -matrix S_0 . Choose a measurable function $g_0(B) \geq 0$. Fix ν and λ or choose a measurable function $\psi_0(\nu, \lambda) \geq 0$, $\nu > m - 1$, $\lambda > 0$.*
2. *For each $t = 1, \dots, T$ calculate $e_t = Y_t - \bar{B}_t X_t$ and*

$$(12) \quad N_t = \lambda N_{t-1} + X_t X_t'$$

$$(13) \quad \bar{B}_t = (\lambda \bar{B}_{t-1} N_{t-1} + Y_t X_t') N_t^{-1}$$

$$(14) \quad S_t = \lambda S_{t-1} + \frac{\lambda}{\nu} e_t (1 - X_t' N_t^{-1} X_t) e_t'$$

$$(15) \quad g_t(B) \equiv g_{t-1}(B) | (B - \bar{B}_t) N_t (B - \bar{B}_t)' + \frac{\nu}{\lambda} S_t |^{-1/2}$$

$$(16) \quad \psi_t(\nu, \lambda) \equiv \frac{\Gamma_m((\nu + l + 1)/2)}{\Gamma_m((\nu + l)/2)} \lambda^{m(l+\nu)/2} \psi_{t-1}(\nu, \lambda)$$

In equation (16), $\Gamma_m(\cdot)$ is the multivariate gamma function, defined in Muirhead (1982), Definition 2.1.10. In practice, one has to recalculate step 2 for each ν , λ and B in the domain of the functions $\psi_0(\nu, \lambda)$ and $g_0(B)$. Note that N_t^{-1} can be computed numerically cheaply via

$$N_t^{-1} = \left(N_{t-1}^{-1} - N_{t-1}^{-1} X_t X_t' N_{t-1}^{-1} / (X_t' N_{t-1}^{-1} X_t + \lambda) \right) / \lambda,$$

as can be verified directly or with rule (T8), p. 324 in Leamer (1978). The proof that the formulas above correctly calculate the posterior follows directly from the two theorems below.

Theorem 1 *Let a prior for $\nu > m - 1$, $\lambda > 0$, the $m \times l$ coefficient matrix matrix B and the $m(m + 1)/2$ distinct elements of the precision matrix H be given by a density proportional to*

$$\frac{\psi(\nu, \lambda)}{\kappa_{NW}(N, S, \nu, m)} g(B) f_{NW}(B, H | \bar{B}, N, S, \nu),$$

where N and S are positive definite and $\psi(\nu, \lambda) \geq 0$, $g(B) \geq 0$ are measurable functions. Suppose additionally, that there is one observation of data X and Y (where X is $l \times 1$ and Y is $m \times 1$), obeying $Y = BX + \mathcal{U}(H^{-1})' \epsilon$, with $\epsilon \sim \mathcal{N}(0, I_m)$. Then the posterior for ν , λ , B and H is given by a density proportional to

$$\frac{\psi(\nu, \lambda)}{\kappa_{NW}(\tilde{N}, \tilde{S}, \nu + 1, m)} g(B) f_{NW}(B, H | \tilde{\bar{B}}, \tilde{N}, \tilde{S}, \nu + 1),$$

where \tilde{N} and \tilde{S} are positive definite and where $\tilde{N} = N + X X'$ and $\tilde{\bar{B}} = (\bar{B} N + Y X') \tilde{N}^{-1}$ as well as

$$\tilde{S} = \frac{\nu}{\nu + 1} S + \frac{1}{\nu + 1} e (1 - X' \tilde{N}^{-1} X) e',$$

where $e = (Y - \bar{B} X)$.

The proof is in appendix B.

Theorem 2 *Let a prior for $\nu > m - 1$, $\lambda > 0$, the $m \times l$ coefficient matrix B and the $m(m+1)/2$ distinct elements of the precision matrix H be given by a density proportional to*

$$\frac{\psi(\nu, \lambda)}{\kappa_{NW}(N, S, \nu + 1, m)} g(B) f_{NW}(B, H \mid \bar{B}, N, S, \nu + 1).$$

Suppose additionally, that there is one unobserved shock to the precision matrix obeying $\tilde{H} = \mathcal{U}(H)' \Theta \mathcal{U}(H) / \lambda$, and $\Theta \sim \mathcal{B}_m((\nu + l)/2, 1/2)$, where $\mathcal{B}_m(p, q)$ is the multivariate beta-distribution, see appendix A. Then the posterior density in ν , λ , B and \tilde{H} is proportional to

$$\frac{\tilde{\psi}(\nu, \lambda)}{\kappa_{NW}(\tilde{N}, \tilde{S}, \nu, m)} \tilde{g}(B) f_{NW}(B, \tilde{H} \mid \bar{B}, \tilde{N}, \tilde{S}, \nu),$$

where $\tilde{N} = \lambda N$, and $\tilde{S} = \lambda^{\frac{\nu+1}{\nu}} S$, as well as

$$\tilde{g}(B) \equiv g(B) \mid (B - \bar{B})N(B - \bar{B})' + \frac{\nu}{\lambda} \tilde{S} \mid^{-1/2},$$

and

$$\tilde{\psi}(\nu, \lambda) \equiv \frac{\Gamma_m((\nu + l + 1)/2)}{\Gamma_m((\nu + l)/2)} \lambda^{m(l+\nu)/2} \psi(\nu, \lambda).$$

The proof is in appendix B. The updating formulas also show that the ordering of the variables does not matter in equation (9) despite the use of the Cholesky-factorization. For suppose that one instead considers the transformed data $\tilde{Y}_t = AY_t$, where A is an orthogonal matrix. Define

$$\check{A} = \begin{bmatrix} I_{c,c} & 0_{c,km} \\ 0_{km,c} & I_k \otimes A \end{bmatrix}.$$

The coefficient matrix for the transformed system is $\tilde{B} = AB\check{A}'$ and the list of regressors is $\tilde{X}_t = \check{A}X_t$. Thus, modify the prior accordingly as well, using $\tilde{\tilde{B}}_0 = A\tilde{B}_0\check{A}'$, $\tilde{\tilde{N}}_0 = \check{A}N_0\check{A}'$, $\tilde{\tilde{S}}_0 = AS_0A'$ and $\tilde{\tilde{g}}_0(\tilde{\tilde{B}}) = g_0(A'\tilde{\tilde{B}}\check{A}')$. It is then easy to check that the updating formulas (12), ..., (16) applied to the transformed system simply result in the transformation of the updated parameters of the original system. In particular $\tilde{\tilde{S}}_t = AS_tA'$, implying that reordering the system does not matter for the method proposed above. This is somewhat surprising, since in general, $\mathcal{U}(AHA') \neq A\mathcal{U}(H)A'$ for arbitrary positive semidefinite matrices H .

4 Numerical Analysis

To use the method, one first needs to select a prior. For the analysis of macroeconomic time series, we suggest using a modification of the Minnesota random walk prior, see

Doan, Litterman and Sims (1984). That is, we suggest setting the prior mean \bar{B}_0 to correspond to a random walk specification and to choose, say, $\nu = 20$ for quarterly data and $\lambda = \nu/(\nu + 1)$. More details are in appendix C. Having selected a prior, the method above delivers the posterior, which needs to be analyzed numerically. We focus entirely on the case, where λ and ν are treated as known and where $g_0(B) \equiv 1$, i.e. where the prior is given by equation 10. Numerical methods are needed, since the posterior (11) is proportional to a Normal-Wishart distribution scaled with the function $g_T(B)$. We use importance-based sampling to analyze the posterior, see Geweke (1989). First, integrate over H_{T+1} to find the marginal posterior

$$(17) \quad \begin{aligned} \log(\pi_{T,\text{marg}}(B)) &= \text{const} - \frac{1}{2} \sum_{t=1}^T \log | (B - \bar{B}_t)N_t(B - \bar{B}_t)' + \frac{\nu}{\lambda}S_t | \\ &\quad - \frac{l + \nu}{2} \log | (B - \bar{B}_T)N_T(B - \bar{B}_T)' + \frac{\nu}{\lambda}S_T | . \end{aligned}$$

Conditional on the coefficient matrix B , the precision matrix H_{T+1} has a Wishart distribution $\mathcal{W}_m(l + \nu, \Omega)$, where

$$(18) \quad \Omega^{-1} = \lambda(B - \bar{B}_T)N_T(B - \bar{B}_T)' + \nu S_T.$$

Note that the marginal posterior is a product of generalized multivariate t-distribution and hence similar to the kind of distributions occurring in the study of common parameters, see Box and Tiao (1973), Chapter 9. While it can have multiple peaks in principle, we proceed under the assumption that there is a single peak: multiple peaks or a very flat single peak should be interpreted as indicators of misspecification (e.g. for a break in the sample).

Find the maximum of (17) with the following modified Newton-Raphson method. Let

$$J = -\frac{l + \nu}{2}N_T \otimes \left(\frac{\nu}{\lambda}S_T\right)^{-1} - \frac{1}{2} \sum_{t=1}^T N_t \otimes \left(\frac{\nu}{\lambda}S_t\right)^{-1} ,$$

be the sum of all second derivatives of the individual pieces of (17) evaluated at their individual maximum. Set $B^{(0)} \equiv \bar{B}_T$ and iterate on

$$\text{vec}(B^{(n)}) = \text{vec}(B^{(n-1)}) - J^{-1} \frac{\partial}{\partial \text{vec}(B)} \log(\pi_{T,\text{marg}}(B^{(n-1)})).$$

until convergence, where the gradient can be computed with theorem 3 in appendix B.

Let B^* denote the maximum of the posterior and let J^* be the Hessian of (17) at $\text{vec}(B^*)$ (see theorem 3 in appendix B). Use as importance sampling density a t-distribution centered at B^* with Hessian J^* , whose degrees of freedom are chosen to ensure fatter tails than those of the marginal posterior $\pi_{T,\text{marg}}(B)$: choose ν^* with $0 < \nu^* < T + l + \nu - ml$, preferably close to the upper bound. Choose the simulation

sample size n . Generate i.i.d. draws $B_i, i = 1, \dots, n$ from the multivariate-t importance sampling density

$$I(B) \propto \left(1 + \text{vec}(B - B^*)' \frac{-J^*}{\nu^* + ml} \text{vec}(B - B^*) \right)^{-(\nu^* + ml)/2}.$$

Given B_i , draw H_i from a Wishart distribution according to (18). Calculate the weight function, whose logarithm is given by

$$\log(w(B_i, H_i)) \equiv \text{const} + \log(\pi_{T, \text{marg}}(B_i)) - \log(I(B_i))$$

and where the constant should be chosen so that $w(B^*, S_T^{-1}) \approx 1$ for numerical reasons. Now one can proceed as in Geweke (1989) and approximately calculate $\bar{\phi} = E_{\pi_T}[\phi(B, H_{T+1})]$ for a given measurable function of interest $\phi(B, H_{T+1})$ (provided that $\text{Var}_{\pi_T}[\phi(B, H_{T+1})] < \infty$) via the weighted simulation sample average

$$\bar{\phi}_n = \frac{\sum_{i=1}^n \phi(B_i, H_{T+1,i}) w(B_i, H_{T+1,i})}{\sum_{i=1}^n w(B_i, H_{T+1,i})}.$$

The method has been successfully applied in Uhlig (1996). For the purposes here, it may be interesting to highlight a numerical issue that arose in an application of the method to a four-variable system with $k = 5$ lags. With a constant and a time trend, \bar{B} now contains 88 entries. We chose a prior as specified in appendix C. To analyze the posterior, we proceeded as described above, using $\nu^* = 72$ and $n = 4000$ random draws. When looking at one-dimensional slices, there is almost no difference between the logarithm of the importance sampling density and the density of the posterior. Nonetheless, the small remaining differences pile up quickly due to the high dimensionality of the problem. Figure 2 shows a scatter plot of the logarithm of the weights for the 4000 random draws. The weights can differ by orders of magnitude: examining the raw numbers shows that the draw with the largest weight receives 5.3% of the sum of all weights, the 109 most heavily weighted draws constitute 50% of the mass and the 741 draws with the highest weights make up 90%. While these numbers indicate a substantial unevenness in the weight distribution, they also indicate that inference based on these draws is nonetheless likely to be sensible since several hundred draws rather than just a few will “matter” for final results. In 88 dimensions as here, it is, in general, easily possible that importance sampling assigns practically the entire weight to a single point: thus, figure 2 looks actually rather good and indicates that the procedure described above should work in practice.

5 Conclusion

This paper introduced Bayesian vector autoregressions with stochastic volatility, deriving in closed form the Bayesian posterior, when the error precision matrix is stochas-

tically time-varying. The key to the proof was a recent result concerning the conjugacy between singular multivariate beta distributions and Wishart distributions, see Uhlig (1994b), making it possible to integrate out the disturbance to the precision matrix. Posterior-based inference requires numerical methods: the paper examines an importance-sampling based approach.

*Center for Economic Research, Tilburg University, Postbus 90153, 5000 LE Tilburg,
THE NETHERLANDS.*

Appendix

A Some distributions and their properties

This appendix has the purpose to fix the notation and to review some useful facts. Additional information can be found in Zellner (1971), Leamer (1978), Muirhead (1982) and Uhlig (1994b).

In the proof of Theorem 2 below, we need the density of the rank-1 singular Wishart distribution. Uhlig (1994b) has shown this density to be

$$(19) \quad f_W(Z | \Omega, 1) = \frac{\pi^{(-m+1)/2} |\zeta|^{-m/2}}{2^{m/2} \Gamma\left(\frac{1}{2}\right) |\Omega|^{1/2}} \exp\left(-\frac{1}{2} \text{tr}(\Omega^{-1} Z)\right),$$

with respect to the volume element (see Uhlig, 1994b) on the space of rank-1 positive semidefinite $m \times m$ -matrices, where ζ is the unique nonzero eigenvalue of Z .

The singular multivariate beta distribution $\mathcal{B}_m(p, 1/2)$ needed in equation (9) is defined and its density calculated in Uhlig (1994b). Let $\check{\Theta} = I_m - \Theta$. The density for $\check{\Theta}$ with respect to the volume element (see Uhlig, 1994b) on the subspace of positive semidefinite rank-1 matrices, where $I_m - \check{\Theta}$ is positive semidefinite as well, is given by

$$(20) \quad f_{B,m,p,1/2}(\check{\Theta}) = \pi^{(-m+1)/2} \frac{\Gamma_m(p+1/2)}{\Gamma(1/2)\Gamma_m(p)} \check{\vartheta}^{-m/2} |I_m - \check{\Theta}|^{p-(m+1)/2}$$

where $\check{\vartheta}$ is the unique nonzero eigenvalue of $\check{\Theta}$. One can show that $E[\Theta] = p/(p+q)I_m$ using the conjugacy with the Wishart distribution established in Uhlig (1994b).

To describe priors and posteriors, Normal-gamma distributions and Normal-Wishart distributions are needed. The Normal-gamma density for a coefficient $\beta \in \mathbf{R}$ and a precision h is given by

$$f_{NG}(\beta, h | \bar{b}, n, s, \nu) = \kappa_{NG}(n, s, \nu) h^{(\nu-1)/2} \exp\left(-\frac{1}{2}(\beta - \bar{b})^2 nh - \frac{\nu}{2} sh\right)$$

where

$$\kappa_{NG}(n, s, \nu) = \frac{n^{1/2} (\nu s/2)^{\nu/2}}{(2\pi)^{1/2} \Gamma(\frac{\nu}{2})}$$

is the appropriate integrating constant. The Normal-gamma distribution specifies, that the precision h follows a gamma distribution $\Gamma(s, \nu)$ and that, conditional on h , the coefficient β follows a Normal distribution $\mathcal{N}(\bar{b}, (nh)^{-1})$.

Let N be $l \times l$ and positive definite, let S be $m \times m$ and positive definite, let \bar{B} be a $m \times l$ -dimensional matrix and let $\nu \geq m$. The Normal-Wishart distribution for a $m \times l$

-dimensional coefficient matrix B and a $m \times m$ precision matrix H is given by the density

$$(21) \quad f_{NW}(B, H \mid \bar{B}, N, S, \nu) = \kappa_{NW}(N, S, \nu, m) \mid H \mid^{(l+\nu-m-1)/2} \exp\left(-\frac{1}{2} \text{tr}\left((B - \bar{B})N(B - \bar{B})' + \nu S\right) H\right)$$

where

$$(22) \quad \kappa_{NW}(N, S, \nu, m) = \frac{\mid N \mid^{m/2} \left(\frac{\nu}{2}\right)^{m\nu/2} \mid S \mid^{\nu/2}}{(2\pi)^{(lm)/2} \Gamma_m\left(\frac{\nu}{2}\right)}$$

is the appropriate integrating constant and $\text{vec}(\cdot)$ denotes columnwise vectorization. The Normal-Wishart distribution specifies that the precision matrix H follows a Wishart distribution $\mathcal{W}_m(\nu, (\nu S)^{-1})$ with $(E[H])^{-1} = S$, and that, conditional on H , the coefficient matrix B in its vectorized form $\text{vec}(B')$ follows a Normal Distribution $\mathcal{N}\left(\text{vec}(\bar{B}'), H^{-1} \otimes N^{-1}\right)$. This definition is a slight generalization of Leamer (1978). The Normal-Wishart distribution is popular in traditional Bayesian multivariate regression analysis, see Zellner (1971).

B Proofs

Proof of Theorem 1. The proof proceeds by directly calculating the densities. Note first, that \tilde{N} and \tilde{S} are indeed positive definite. With equation (21), the posterior density, obtained by multiplying the prior with the likelihood, is proportional to

$$\begin{aligned} \pi(\nu, \lambda, B, H) &\propto \psi(\nu, \lambda) g(B) \mid H \mid^{(l+\nu-m)/2} \\ &\quad \exp\left(-\frac{1}{2} \text{tr}\left((Y - BX)(Y - BX)' + (B - \bar{B})N(B - \bar{B})' + \nu S\right) H\right) \\ &\propto \psi(\nu, \lambda) g(B) \mid H \mid^{(l+\nu-m)/2} \\ &\quad \exp\left(-\frac{1}{2} \text{tr}\left((B - \tilde{B})\tilde{N}(B - \tilde{B})' + (\nu + 1)\tilde{S} + \phi(B)\right) H\right) \end{aligned}$$

where

$$\begin{aligned} \phi(B) &= (B - \bar{B})N(B - \bar{B})' + (Y - BX)(Y - BX)' - (B - \tilde{B})\tilde{N}(B - \tilde{B})' \\ &\quad - (Y - \bar{B}X)(1 - X'\tilde{N}^{-1}X)(Y - \bar{B}X)' \end{aligned}$$

It remains to show that $\phi(B) \equiv 0$ or, equivalently, that

1. $\phi(\bar{B}) = 0$ and that
2. $\frac{\partial \phi}{\partial B_{ij}} \equiv 0$ for all i, j .

The first part follows with

$$\left(\bar{B} - \tilde{B}\right) \tilde{N} \tilde{N}^{-1} \tilde{N} \left(\bar{B} - \tilde{B}\right)' = \left(Y - \bar{B}X\right) X' \tilde{N}^{-1} X \left(Y - \bar{B}X\right)'$$

The second part is easy to see.

Q.E.D.

Proof of Theorem 2. For notation see appendix A, Muirhead (1982) and Uhlig (1994b). We first give the proof for $m \geq 2$. The idea is to follow the proof of Theorem 7 in Uhlig (1994b) “in reverse” and additionally keeping track of terms coming from the Normal density in the Normal-Wishart distribution. Recall the density for $\check{\Theta} = I_m - \Theta$ on the space of rank-1 positive semidefinite matrices, given in equation (20). The posterior in $B, H, \check{\Theta}, \nu$ and λ is

$$\begin{aligned} \pi(\nu, \lambda, B, H, \check{\Theta}) &\propto \psi(\nu, \lambda) g(B) | H |^{(l+\nu-m)/2} \\ &\exp\left(-\frac{1}{2} \text{tr}\left(\left(B - \bar{B}\right) N \left(B - \bar{B}\right)' + (\nu + 1)S\right) H\right) \\ &\frac{\Gamma_m((\nu + l + 1)/2)}{\Gamma_m((\nu + l)/2)} \check{\nu}^{-m/2} | \Theta |^{(\nu+l-m-1)/2} d\nu \wedge d\lambda \wedge (dB) \wedge (dH) \wedge (d\check{\Theta}) \end{aligned}$$

where $\check{\nu}$ is the unique nonzero eigenvalue of $\check{\Theta}$. Let $\check{H} = \mathcal{U}(H)' \Theta \mathcal{U}(H)$ and $Z = \mathcal{U}(H)' \check{\Theta} \mathcal{U}(H)$. Note that $\check{H} + Z = H$ and that Z is positive semidefinite and of rank 1. Let ζ denote³ the unique nonzero eigenvalue of Z . It follows from Theorem 4 in Uhlig (1994b), that

$$(d\check{H}) \wedge (dZ) = (dH) \wedge (dZ) = (\zeta/\check{\nu})^{m/2} | \mathcal{U}(H) | (dH) \wedge (d\check{\Theta}).$$

Note that $\tilde{H} = \check{H}/\lambda$ and $(d\tilde{H}) = \lambda^{-m(m+1)/2} (d\check{H})$. Exploiting $| \mathcal{U}(H) | = | H |^{1/2}$, it follows that

$$\begin{aligned} \pi(\nu, \lambda, B, \tilde{H}, Z) &\propto \tilde{\psi}(\nu, \lambda) g(B) \zeta^{-m/2} | \tilde{H} |^{(l+\nu-m-1)/2} \\ &\exp\left(-\frac{1}{2} \text{tr}\left(\left(B - \bar{B}\right) \tilde{N} \left(B - \bar{B}\right)' + \nu \tilde{S}\right) \tilde{H} + R^{-1} Z\right) d\nu \wedge d\lambda \wedge (dB) \wedge (d\tilde{H}) \wedge (dZ) \end{aligned}$$

where

$$\tilde{\psi}(\nu, \lambda) = \frac{\Gamma_m((\nu + l + 1)/2)}{\Gamma_m((\nu + l)/2)} \lambda^{m(l+\nu)/2} \psi(\nu, \lambda)$$

and $R^{-1} = \left(B - \bar{B}\right) N \left(B - \bar{B}\right)' + (\nu/\lambda) \tilde{S}$. Using the density given in equation (19) for $n = 1$ and integrating over (dZ) , one obtains the marginal posterior

$$\begin{aligned} \pi(\nu, \lambda, B, \tilde{H}) &\propto \tilde{\psi}(\nu, \lambda) g(B) | R |^{1/2} | \tilde{H} |^{(l+\nu-m-1)/2} \\ &\exp\left(-\frac{1}{2} \text{tr}\left(\left(B - \bar{B}\right) \tilde{N} \left(B - \bar{B}\right)' + \nu \tilde{S}\right) \tilde{H}\right) (dB) \wedge (d\tilde{H}) \wedge d\nu \wedge d\lambda. \end{aligned}$$

For $m = 1$, the same calculations go through verbatim. In that case, the formulas for the densities of singular multivariate Wishart- and beta-distributions collapse to the usual densities of univariate gamma- and beta-distributions with $\check{\vartheta} = \check{\Theta} = 1 - \Theta$, $\zeta = Z$ and $\mathcal{U}(H) = H^{1/2}$, $H \in \mathbf{R}_+$. **Q.E.D.**

Theorem 3 Let \bar{B} be $m \times l$, and N be $l \times l$ and positive semidefinite and Z be $m \times m$ and positive definite. For $B \in \mathbf{R}^{m \times l}$, define

$$f(B) \equiv \log \left(| (B - \bar{B})N(B - \bar{B})' + Z | \right).$$

For $1 \leq i, \tilde{i} \leq m$, $1 \leq j, \tilde{j} \leq l$, define E_{ij} to be the $m \times l$ matrix with a 1 as its (i, j) -entry and zeros everywhere else, and

$$\begin{aligned} A &= \left((B - \bar{B})N(B - \bar{B})' + Z \right)^{-1} \\ C_{ij} &= E_{ij}N(B - \bar{B})' + (B - \bar{B})NE_{ij} \\ D_{ij, \tilde{i}\tilde{j}} &= E_{ij}NE'_{\tilde{i}\tilde{j}} + E_{\tilde{i}\tilde{j}}NE'_{ij}. \end{aligned}$$

Then,

$$\frac{\partial f}{\partial B_{ij}} = (\text{vec}(A))' \text{vec}(C_{ij})$$

and

$$\frac{\partial^2 f}{\partial B_{ij} \partial B_{\tilde{i}\tilde{j}}} = (\text{vec}(A))' \text{vec}(D_{ij, \tilde{i}\tilde{j}}) - (\text{vec}(A C_{\tilde{i}\tilde{j}} A))' \text{vec}(C_{ij})$$

In particular, at $B = \bar{B}$,

$$\frac{\partial^2 f}{\partial \text{vec}(B) \partial \text{vec}(B)'} = N \otimes Z^{-1}$$

Proof of Theorem 3. This follows from basic rules of calculus involving matrices. To provide some details, define $h : \mathbf{R} \rightarrow \mathbf{R}^{lm}$ by

$$h(x) \equiv \text{vec} \left((B - \bar{B} + xE_{ij})N(B - \bar{B} + xE_{ij})' + Z \right).$$

Define the linear function $\text{mat} : \mathbf{R}^{lm} \rightarrow \mathbf{R}^{l \times m}$ to be the inverse of $\text{vec}(\cdot)$ and let $r : \mathbf{R}^{lm} \rightarrow \mathbf{R}$, $r(v) = \log(|\text{mat}(v)|)$. Calculate

$$\frac{\partial f}{\partial B_{ij}} = \frac{\partial}{\partial v} r|_{v=h(0)} \frac{\partial}{\partial x} h|_{x=0} = (\text{vec}(A))' \text{vec}(C_{ij}).$$

To find the second derivative, use the rule

$$\frac{\partial M^{-1}}{\partial t} = -M^{-1} \frac{\partial M}{\partial t} M^{-1}$$

to show that

$$\frac{\partial}{\partial B_{ij}} \left((B - \bar{B})N(B - \bar{B})' + Z \right)^{-1} = -A C_{ij} A$$

and thus the result with some additional algebra.

Q.E.D.

C Prior Selection

The selection of a prior in Bayesian time series analysis has recently been subject to much debate, see Phillips' (1991) critique of Sims and Uhlig (1991), his discussants, and the summary in Uhlig (1994a). For the analysis of macroeconomic time series we suggest the following modification of the "Minnesota prior" in Doan, Litterman and Sims (1984) or Doan's RATS manual (1990). The main difference is that we choose values for ν and λ and that we include a constant and a time trend in the regression.

Use logarithms of the levels of the time series except for series expressed in per cent (like interest rates), which are used in their raw form. Include a constant and a time trend, $C_t = [1 \ t]'$ and $c = 2$.

Treat ν and λ as fixed at $\nu = 20$ for quarterly data and $\nu = 60$ for monthly data: this allows for a reasonable amount of time variation. Set $\lambda = \nu/(\nu + 1)$. Sensitivity analysis with respect to ν and λ is advisable. Set $g_0(B) \equiv 1$. Let S_0 be the diagonal matrix of the average squared residuals from AR(1) univariate regressions for each included data series. This, of course, amounts to a first pass through the data, which, strictly speaking, is not legitimate. However, this "loss of m degrees of freedom" should not be big for coming up with a reasonable starting point S_0 in most practical applications. Include between one and two years of lags (e.g. 5 lags for quarterly data). Let \bar{B}_0 be the random walk specification

$$\bar{B}_0 = [0_{m,2} \ I_m \ 0_m \ \dots \ 0_m].$$

Choose N_0 to be block diagonal. The first block is of size 2×2 with

$$\begin{bmatrix} N_0(1,1) & N_0(1,2) \\ N_0(2,1) & N_0(2,2) \end{bmatrix} = \begin{bmatrix} \zeta_3 & -\zeta_3^2/2 \\ -\zeta_3^2/2 & \zeta_3^3/3 \end{bmatrix}.$$

The second block is of size $(km) \times (km)$ and diagonal with

$$N_0(2 + m(l-1) + i, 2 + m(l-1) + i) = Y_{0,i}^2 \zeta_1 l^{\zeta_2}, \quad i = 1, \dots, m, \quad l = 1, \dots, k.$$

Here, $\zeta_1 \geq 0$, $\zeta_2 \geq 0$ and $\zeta_3 \geq 0$ are hyperparameters and Y_0 is the date-0 data vector, where we assume that all $Y_{0,i} \neq 0, i = 1, \dots, m$. The motivation for this particular form of the prior can be seen from the updating equation (12) for N_t at $\lambda = 1$ and from the interpretation of N_t as precision along rows of B , see equation (21) given in appendix A: ζ_1 and ζ_3 correspond roughly to the number of presample dummy observations, starting from a flat prior: given $N = 0$ at time $t = -\zeta_3$, and ζ_3 artificially created and added observations (dummy observations) for $t = -\zeta_3 + 1, \dots, 0$, for which the linear relationships hold exactly at the prior mean \bar{B}_0 , the updating formulas will result in $N_0(i, j), I = 1, 2, j = 1, 2$ as defined above. A similar argument can be given for ζ_1 . ζ_2 determines how “fast” higher lags are excluded. We suggest using $\zeta_1 = 5$, $\zeta_2 = 2$ and $\zeta_3 = 8$. Note that we have $\zeta_2 = 2 > 0$ here, seemingly in contrast to standard BVAR methodology as in Doan, Litterman and Sims (1984). The reason is that the elements of N denote precision here, not variance as usual. Our choices for the hyperparameters are quite weak, since the prior imposed that way is swamped in the first few observations, see equation (12). Rescaling of a time series is automatically taken care of with this prior in practice: writing an interest rate of 3.1% as 0.031 or as 3.1, say, results in the same inference.

References

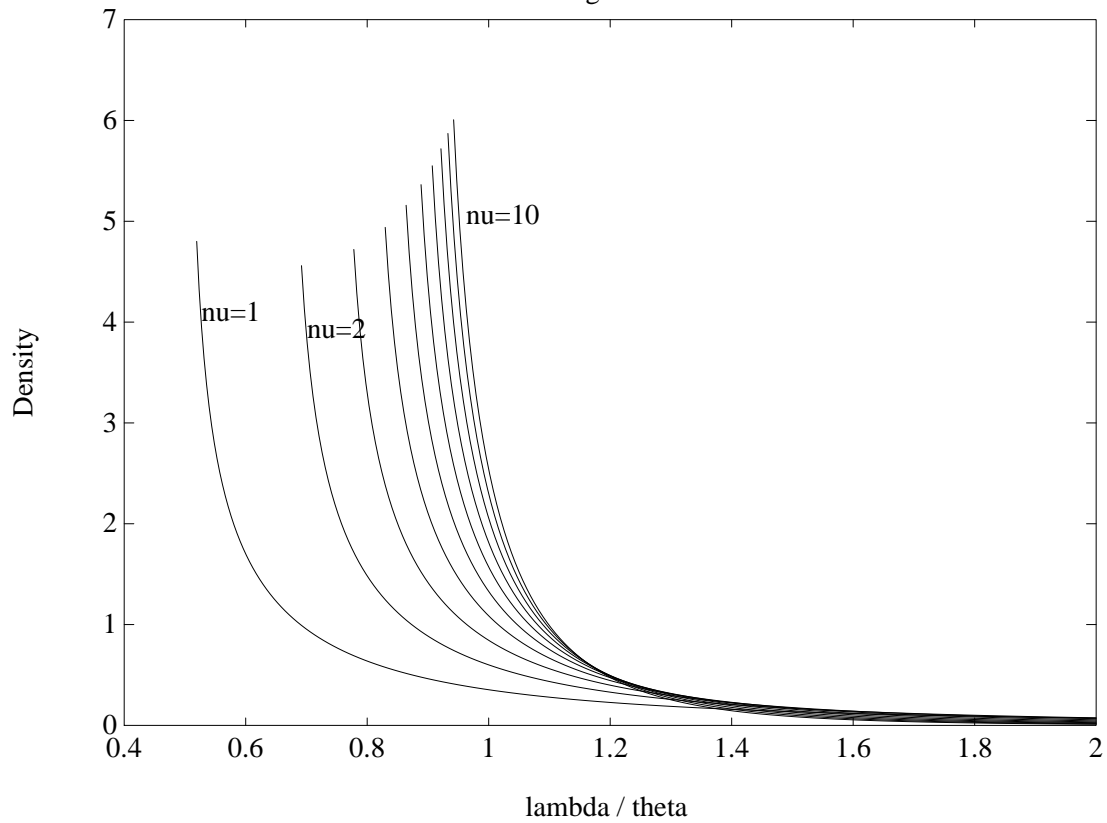
- [1] BOLLERSLEV, T., R. Y. CHOU AND K. F. KRONER (1992): "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5-59
- [2] BOX, G. E. P. AND G. C. TIAO (1973): *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons.
- [3] DOAN, T., R. LITTERMAN, AND C. A. SIMS (1984): "Forecasting and Conditional Projection using Realistic Prior Distributions," *Econometric Reviews*, 3, 1-100.
- [4] DOAN, T. (1990): *RATS Manual*. VAR Econometrics, 1800 Sherman Ave. Suite 612, Evanston, IL 60201.
- [5] ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1008.
- [6] GEWEKE, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317-1339.
- [7] JACQUIER, E., N. G. POLSON AND P. E. ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371-389.
- [8] LEAMER, E. E. (1978): *Specification Searches - ad hoc Inference with Nonexperimental Data*. New York: John Wiley & Sons.
- [9] LITTERMAN, R. (1979): "Techniques of forecasting using vector autoregressions", Federal Reserve Bank of Minneapolis, working paper 115.
- [10] MUIRHEAD, R. J. (1982): *Aspects of multivariate statistical theory*. New York: John Wiley & Sons.
- [11] NELSON, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318-334.
- [12] PHILLIPS, P. C. B. (1991): "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6, 333-364.
- [13] SHEPHARD, N. G. (1994): "Local Scale Models: State Space Alternative to Integrated GARCH Processes," *Journal of Econometrics*, 60, 181-202.

- [14] SIMS, C. A. AND H. UHLIG (1991): “Understanding Unit Rooters: A Helicopter Tour,” *Econometrica*, 59, 1591-1600.
- [15] UHLIG, H. (1994A): “What Macroeconomists Should Know About Unit Roots: a Bayesian Perspective,” *Econometric Theory*, 10, 645-671.
- [16] UHLIG, H. (1994B):, “On Singular Wishart and Singular Multivariate Beta Distributions,” draft, Princeton University, *The Annals of Statistics*, 22, 395-405.
- [17] Uhlig, H. (1996), “The Changing Effect of Monetary Policy: a Stochastic Volatility Approach,” draft, CentER, Tilburg University.
- [18] ZELLNER, A. (1971): *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons.

Footnotes

1. I am deeply indebted to Neil Shephard, who found a mistake - now fixed - in Theorem 2 in an early version of this paper, as well as to Chris Sims. I am grateful to Gregory Chow, Robert Engle, Jum-Keung Kwan, Robin Lumsdaine, Dan Nelson, Jecek Osiewalski, Pierre Perron, Peter Rossi, Peter Schotman, Mark Steel, Mark Watson and Arnold Zellner as well as seminar participants at various seminars for comments on earlier versions of the paper. I am grateful to Peter Rossi for allowing me to use his computer. I am grateful to three careful referees and the co-editor for helpful comments. I am grateful to Werner Hildenbrand and the SFB 303 at the University of Bonn as well as the University of Chicago for their hospitality and the opportunity to complete this research.
2. See Appendix A, equations (21), (22) for the definition of the Normal-Wishart density f_{NW} and the integrating constant κ_{NW} .

Figure 2



Figures

FIGURE 1 - Distribution for λ/θ where $\lambda = \nu/(\nu + 1)$. λ/θ is the multiplicative, unobserved shock to the variance σ_t^2 in the univariate model, see equation (2).

Figure 4

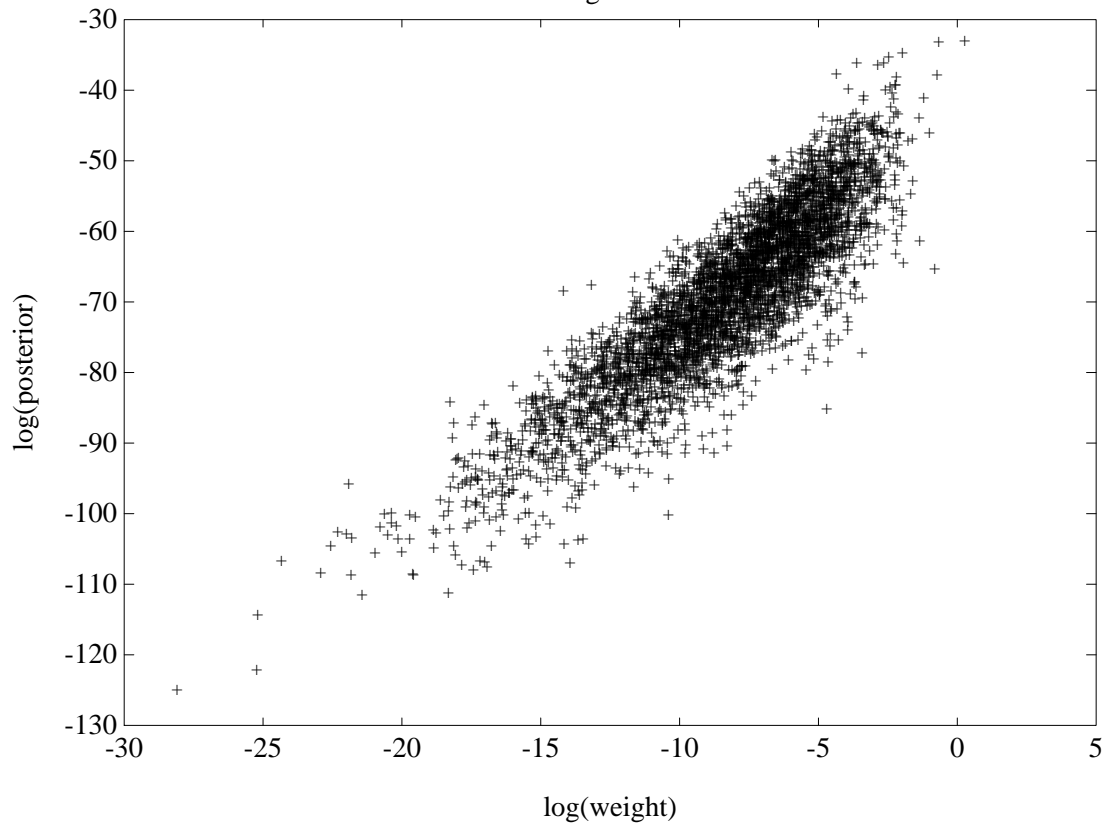


FIGURE 2 - Scatterplot for the 4000 generated draws from the importance sampling density.