

Research Support  
University of Michigan Business School

February 1998

UNIFORM SCALE MIXTURE MODELS WITH  
APPLICATIONS TO BAYESIAN INFERENCE

WORKING PAPER #98005

BY

ZHAOHUI QIN

UNIVERSITY OF MICHIGAN, DEPARTMENT OF STATISTICS

PAUL DAMIEN

UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

STEPHEN WALKER

IMPERIAL COLLEGE, LONDON, ENGLAND

# Uniform Scale Mixture Models With Applications to Bayesian Inference

Zhaohui Qin<sup>1</sup>  
Paul Damien<sup>2</sup>  
Stephen Walker<sup>3</sup>

<sup>1</sup> Department of Statistics, University of Michigan, Ann Arbor, MI. 48109  
USA.

<sup>2</sup> University of Michigan Business School, Ann Arbor, MI. 48109 USA.

<sup>3</sup> Department of Mathematics Imperial College, 180 Queen's Gate, London  
SW7 2BZ, UK.

## SUMMARY

We show that the scale mixture of uniform family generalises the scale mixture of normal family. We also show that inference for the former family is more straightforward and can easily accommodate both mean and variance regression functions. Illustrative analysis for the autocorrelated-heteroscedastic regression model is provided using data obtained from 14 US banks.

*Key words:* Scale mixture of uniforms, Uniform distribution, Gibbs sampler, Autocorrelation, Heteroscedasticity, Truncated densities, Share price, Economic value added, Net operating profit, Market capitalization.

## 1 Introduction

We start this paper by considering the scale mixture of normal family of distributions (Andrews and Mallows, 1974; Karim and Paruolo, 1996). As is

well known these provide generalisations to the normal family; for example, the exponential power and Student  $t$  distributions. Their practical use is in providing probability density functions with heavier tails compared to the normal, useful for modeling robustness (Box and Tiao, 1973; Choy and Smith, 1997). The heaviness of tails is measured by kurtosis. The normal density has zero kurtosis whereas a scale mixture of normal density can have positive kurtosis (leptokurtic).

Consider the density function for  $X$ , defined on  $(-\infty, +\infty)$ , given by

$$f_X(x) = \int N(x|\mu, \sigma^2\lambda)\pi(\lambda)d\lambda,$$

where  $N(\mu, \sigma^2)$  denotes a normal density with mean  $\mu$  and variance  $\sigma^2$  and  $\pi(\cdot)$  is a density defined on  $(0, +\infty)$ . The exponential power arises when  $\pi(\cdot)$  is an inverse positive stable density and the Student  $t$  arises when  $\pi(\cdot)$  is an inverse gamma density. However, a surprise is why interest has focused on these two alone. After all, if

$$E[\lambda] = 1 \quad \text{and} \quad \text{var}[\lambda] = \tau/3$$

then

$$E[X] = \mu, \quad \text{var}[X] = \sigma^2 \quad \text{and} \quad \kappa(X) = \tau,$$

where  $\kappa(X)$  denotes the kurtosis. Note then that the kurtosis must be positive.

Our work is based on the fact that we can express the normal distribution as a scale mixture of a uniform distribution. Therefore, any distribution which has a scale mixture of normal representation, also has a scale mixture of uniform representation. Moreover, the scale mixture of uniform family includes densities with negative kurtosis (platykurtic).

Here we present a result which does not appear to be widely known, at least if known, its relevance for making statistical inference. In the following  $G$  represents a gamma distribution and  $U$  a uniform distribution. We state without proof:

**Theorem 1.** If  $X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v})$  and  $V \sim G(3/2, 1/2)$  then  $X \sim N(\mu, \sigma^2)$ .

## 2 Scale mixtures of uniform distributions

We introduce the scale mixture of uniform family by considering the scale mixture of normal family and using the result of Theorem 1. Suppose then

$$f_X(x) = \int N(x|\mu, \sigma^2\lambda)\pi(\lambda)d\lambda.$$

We can write this in a three level model, given by

$$X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v}),$$

$$V \sim G(3/2, \sqrt{\lambda}/2)$$

$$\lambda \sim \pi(.).$$

We can combine the last two to give

$$f_V(v) = \frac{\sqrt{v}}{2\sqrt{2}} \int \lambda^{3/4} \exp(-v\sqrt{\lambda}/2) \pi(\lambda)d\lambda.$$

The family of distributions we will be looking at in this paper has the representation:

$$X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v}),$$

$$V \sim f_V(.).$$

Here we note that the Student  $t$  and exponential power distributions arise as scale mixtures of uniform distributions. The Student  $t$ , with  $\alpha$  degrees of freedom, has the representation

$$X|\xi \sim N(\mu, \sigma^2/\xi),$$

$$\xi \sim G(\alpha/2, \alpha/2).$$

We can write the first level as the scale mixture of a uniform:

$$X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v}),$$

$$V|\xi \sim G(3/2, \xi/2).$$

We can then combine  $f(v|\xi)$  and  $f(\xi)$ , integrate over  $\xi$ , to obtain the marginal distribution of  $V$  and hence obtain the scale mixture of a uniform representation:

**Theorem 2.** If  $V$  has density given up to proportionality by

$$f_V(v) \propto \frac{\sqrt{v}}{(\alpha + v)^{(\alpha+3)/2}}$$

and

$$X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v}),$$

then  $X$  has a Student  $t$  distribution with mean  $\mu$ , scale parameter  $\sigma$  and  $\alpha$  degrees of freedom.

We state the result for the exponential power distribution:

**Theorem 3.** If  $V$  has density proportional to

$$f(v) \propto v^{1/\tau-1/2} \exp(-v^{1/\tau}),$$

so  $V^{1/\tau} \sim G(1 + \tau/2, 1)$ , and

$$X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v}),$$

then

$$f_X(x) \propto \exp\left(-\left|\frac{x - \mu}{\sigma}\right|^{2/\tau}\right),$$

where  $\tau \in (0, 2]$ .

This characterisation of the exponential power distribution appears to be more tractable than the alternative scale mixture of a normal characterisation (West, 1987), which is only valid for  $\tau \in (1, 2]$ . We can obtain an interesting result by combining the result of West with ours:

**Theorem 4** (West, 1987). If  $X|\lambda \sim N(0, \lambda)$ , here  $\lambda$  denotes  $1/(\text{variance})$ , and  $f(\lambda) \propto \lambda^{-1/2} p_{1/\tau}(\lambda)$ , ( $1 < \tau < 2$ ), where  $p_a(\cdot)$  denotes the density of the positive stable distribution with index  $a$  ( $0 < a < 1$ ), then  $f_X(x) \propto \exp(-|x|^{2/\tau})$ .

We can now insert the uniform and gamma mixture to replace the normal, leading to the following 3 stage mixture:

$$X|[V = v] \sim U(-\sqrt{v}, +\sqrt{v})$$

$$V|\lambda \sim G(3/2, \lambda/2)$$

and

$$f(\lambda) \propto \lambda^{-1/2} p_{1/\tau}(\lambda).$$

Combining the last two stages implies:

$$v^{1/\tau-1/2} \exp(-v^{1/\tau}) \propto v^{1/2} \int_{\lambda=0}^{\infty} \lambda^{3/2} \exp(-0.5\lambda v) \lambda^{-1/2} p_{1/\tau}(\lambda) d\lambda.$$

Therefore,

**Theorem 5.** If  $V|\lambda$  has the exponential distribution with mean  $2/\lambda$  and  $f(\lambda) = p_{1/\tau}(\lambda)$ , ( $1 < \tau < 2$ ), then  $f_V(v) \propto v^{1/\tau-1} \exp(-v^{1/\tau})$ , a Weibull distribution.

There does not seem to be any reason why we should just consider the Student  $t$  and exponential power distributions.

Consider the general family of distributions:

$$X|[V = v] \sim U(\mu - \sigma\sqrt{v}, \mu + \sigma\sqrt{v})$$

$$V \sim f_V.$$

Then the following hold:  $E[X] = \mu$ ,  $\text{var}[X] = \sigma^2 E[V]/3$  and

$$\kappa[X] = \frac{9E[V^2]}{5[E(V)]^2} - 3.$$

So the mean and variance of  $V$  determine the variance and kurtosis of  $X$ . To obtain  $\text{var}[X] = \sigma^2$  and  $\kappa[X] = \tau$  we require  $E[V] = 3$  and  $\text{var}[V] = 5\tau + 6$ . Note that we must have  $\tau > -6/5$  which is the kurtosis for the uniform density. A particular distribution which satisfies these requirements is given by

$$f_V = G(9\alpha, 3\alpha),$$

where  $\alpha = (5\tau + 6)^{-1}$ . This new family of distributions has parameters  $(\mu, \sigma, \tau)$ , with mean  $\mu$ , variance  $\sigma^2$  and kurtosis  $\tau$ . We recover the normal distribution when  $\tau = 0$  ( $\alpha = 1/6$ ).

The scale mixture of a uniform family coincides with the class of unimodal, symmetric distributions:

**Theorem 6.** If  $f_X$  is a unimodal, symmetric density about 0 and  $f'_X(x)$  exists for all  $x$  then

$$f_X(x) = 1/2 \int_{v>x^2} f_V(v) dv / \sqrt{v},$$

where  $f_V(v) = -f'_X(\sqrt{v})$ .

Therefore, we can write  $X|[V = v] \sim U(-\sqrt{v}, +\sqrt{v})$  with  $V \sim f_V$ , provided  $f_V$  is a density on  $(0, \infty)$ . Note that  $-\int_{v=0}^{\infty} f'_X(\sqrt{v})dv = 1$  which follows from

$$1 = \int_{-\infty}^{+\infty} f_X(x)dx = 1/2 \int_{v=0}^{\infty} -f'_X(\sqrt{v})/\sqrt{v} \left[ \int_{x=-\sqrt{v}}^{+\sqrt{v}} dx \right] dv,$$

and  $f_V(v) \geq 0$  iff  $f_X(x)$  is unimodal.

We have already seen that the scale mixture of uniform family generalises the scale mixture of normal family to include platykurtic shapes. In this section we demonstrate one other advantage of using the scale mixture of uniform family which is the easy analysis involving a variance regression model. Suppose we wish to model data on covariates  $Z$  and  $W$ , such that  $E[X_i] = Z_i\beta$  and  $\log \text{var}[X_i] = 2W_i\theta$ , where  $\beta = (\beta_1, \dots, \beta_p)$  and  $\theta = (\theta_1, \dots, \theta_q)$ . This is difficult to implement using the normal model, even within a Gibbs sampling strategy. However, for the uniform model it is analysis via the Gibbs sampler is remarkably easy. The model is given by

$$X_i|[V_i = v_i] \sim U(Z_i\beta - \exp(W_i\theta)\sqrt{v_i}, Z_i\beta + \exp(W_i\theta)\sqrt{v_i})$$

$$V_i \sim \text{iid } f_V(\cdot).$$

The condition for the variance is satisfied provided we constrain  $E[V] = 3$ . For the implementation of a Gibbs sampler we require the full conditional distributions. We will concentrate on those for  $\beta_k$  and  $\theta_k$  and the  $V_i$ . There will be a parameter associated with  $f_V$  but the full conditional for this parameter will be based on the  $V_i$  being iid from  $f_V$  and so should not pose any problem. If  $\pi(\cdot)$  represent the priors, assumed independent, then the full conditional for  $\beta_k$  is

$$f(\beta_k|\dots) \propto \pi(\beta_k)I(\beta_k \in A_k),$$

where

$$A_k = (\max_{i:Z_{ik} \neq 0} \{X_i - \exp(W_i\theta)\sqrt{v_i} - \gamma_{ik}\}, \min_{i:Z_{ik} \neq 0} \{X_i + \exp(W_i\theta)\sqrt{v_i} - \gamma_{ik}\}),$$

and  $\gamma_{ik} = \sum_{l \neq k} Z_{il}\beta_l$ . The full conditional for  $\theta_k$  is

$$f(\theta_k|\dots) \propto \exp\left(-\left[\sum_i W_{ik}\right]\theta_k\right) \pi(\theta_k)I(\theta_k \in B_k),$$

where

$$B_k = \left( \max_{i:W_{ik}>0} \left\{ 0.5 \log \xi_i - \sum_{l \neq k} W_{il} \theta_l \right\}, \min_{i:W_{ik}<0} \left\{ -0.5 \log \xi_i + \sum_{l \neq k} W_{il} \theta_l \right\} \right),$$

where  $\xi_i = (X_i - Z_i \beta)^2 / V_i$ . Therefore, if priors are normal, then these full conditionals are simply truncated normals. The full conditional for  $V_i$  is

$$f_{V_i}(v | \dots) \propto \exp(-v/2) I(v > (X_i - Z_i \beta)^2 \exp(-2W_i \theta)).$$

These straightforward full conditionals mean that implementation of this model via the Gibbs sampler is 'automatic', in that no rejection and/or Metropolis steps need to be tuned each time a new data set is analysed. Therefore, we have provided the basis for routine mean/variance regression models. This could have dramatic impact for financial time series data which typically includes the model of volatility via variance regression; for example, the stochastic volatility model (Jacquier et al., 1995).

### 3 The autocorrelated-heteroscedastic model

We now discuss the likelihood (expressed as a scale mixture of uniforms), prior, and posterior conditional distributions for the autoregressive-heteroscedastic model. In particular, the simplifications for the Gibbs sampler resulting from the scale mixture of uniforms representation is noted. Following Zellner (1987), chapter 4, we consider the following model.

- Likelihood:

$$y_{it} = \alpha + \beta x_{it} + \epsilon_{it},$$

$$\epsilon_{it} = \rho_i \epsilon_{it-1} + u_{it}, \quad i = 1, \dots, N, t = t_0 + 1, \dots, T$$

where  $u_{it} \sim N(0, \sigma_u^2)$ , these two equations is equivalent to

$$y_{it} = \rho_i y_{it-1} + \beta(x_{it} - \rho_i x_{it-1}) + \alpha(1 - \rho_i) + u_{it}$$



- Scale Mixture:

$$Y_{it}|V_{it} \sim U(\mu_{it} - V_{it}^{1/2}, \mu_{it} + V_{it}^{1/2}),$$

where

$$\mu_{it} = \rho_i y_{it-1} + \beta(x_{it} - \rho_i x_{it-1}) + \alpha(1 - \rho_i), \quad i = 1, \dots, N, t = t_0 + 1, \dots, T,$$

where  $t_0$  is the initial time period. And

$$V_{it}|\lambda \sim G(3/2, \lambda/2),$$

where  $\lambda = \sigma_u^{-2}$

- Prior Distributions:

$$\pi(\alpha) \sim N(\mu_\alpha, \sigma_\alpha^2),$$

$$\pi(\beta) \sim N(\mu_\beta, \sigma_\beta^2),$$

$$\pi(\rho_i) \sim N(\mu_i, \sigma_i^2),$$

$$\pi(\lambda) \sim G(\alpha_\lambda, \beta_\lambda).$$

- Posterior Conditional Distributions:

i)  $V_{it}$

$$P(V_{it}|\alpha, \beta, \rho, \lambda, Y_{it}) = G(1, \lambda/2) I_{[(\mu_{it} - Y_{it}), \infty)},$$

$i = 1, \dots, N$ , and  $t = t_0 + 1 \dots, T$ .

ii)  $\alpha$

$$P(\alpha|Y, V, \beta, \rho, \lambda) = \pi(\alpha) \prod_{i,t} P(Y_{it}, V_{it}, \beta, \rho_i, \lambda|\alpha)$$

$$\propto \pi(\alpha) \prod_{i,t} I_{[\mu_i - V_{it}^{1/2}, \mu_i + V_{it}^{1/2}]}(Y_{it})$$

Let  $a_i = 1 - \rho_i$ ,  $b_{it} = \rho_i y_{it-1} + \beta(x_{it} - \rho_i x_{it-1})$ ,  $i = 1, \dots, N$  and  $t = t_0 + 1, \dots, T$ , such that  $\mu_i - V_{it}^{1/2} < Y_{it} < \mu_i + V_{it}^{1/2}$  is equivalent to

$$\max \left\{ \frac{Y_{it} - b_{it} - V_{it}^{1/2}}{a_i} \right\} < \alpha < \min \left\{ \frac{Y_{it} - b_{it} + V_{it}^{1/2}}{a_i} \right\}$$

for all  $i$  and  $t$ . And the posterior distribution for  $\alpha$  is truncated normal:

$$N(\mu_\alpha, \sigma_\alpha^2) I \left[ \max \left\{ \frac{Y_{it} - b_{it} - V_{it}^{1/2}}{a_i} \right\}, \min \left\{ \frac{Y_{it} - b_{it} + V_{it}^{1/2}}{a_i} \right\} \right].$$

iii)  $\beta$

$$\begin{aligned} P(\beta|Y, V, \alpha, \rho, \lambda) &= \pi(\beta) \prod_{i,t} P(Y_{it}, V_{it}, \alpha, \rho_i, \lambda|\beta) \\ &\propto \pi(\beta) \prod_{i,t} I_{[\mu_i - V_{it}^{1/2}, \mu_i + V_{it}^{1/2}]}(Y_{it}) \end{aligned}$$

Let  $c_{it} = (x_{it} - \rho_i x_{it-1})$ ,  $d_{it} = \rho_i y_{it-1} + \alpha(1 - \rho_i)$ ,  $i = 1, \dots, N$  and  $t = t_0 + 1, \dots, T$ , such that  $\mu_i - V_{it}^{1/2} < Y_{it} < \mu_i + V_{it}^{1/2}$  is equivalent to

$$\max \left\{ \frac{Y_{it} - d_{it} - V_{it}^{1/2}}{c_{it}} \right\} < \beta < \min \left\{ \frac{Y_{it} - d_{it} + V_{it}^{1/2}}{c_{it}} \right\}$$

for all  $i$  and  $t$ . The posterior distribution for  $\beta$  is truncated normal:

$$N(\mu_\beta, \sigma_\beta^2) I \left[ \max \left\{ \frac{Y_{it} - d_{it} - V_{it}^{1/2}}{c_{it}} \right\}, \min \left\{ \frac{Y_{it} - d_{it} + V_{it}^{1/2}}{c_{it}} \right\} \right].$$

iv)  $\rho_i$ ,  $i = 1, \dots, N$

$$\begin{aligned} P(\rho_i|Y, V, \alpha, \beta, \sigma_u) &= \pi(\rho_i) \prod_t P(Y_{it}, V_{it}, \alpha, \beta, \sigma_u|\rho_i) \\ &\propto \pi(\rho_i) \prod_t I_{[\mu_i - V_{it}^{1/2}, \mu_i + V_{it}^{1/2}]}(Y_{it}) \end{aligned}$$

Let  $e_{it} = (y_{it-1} - \beta x_{it-1} - \alpha)$ ,  $f_{it} = \alpha + \beta x_{it}$ ,  $i = 1, \dots, N$  and  $t = t_0 + 1, \dots, T$ , such that  $\mu_i - V_{it}^{1/2} < Y_{it} < \mu_i + V_{it}^{1/2}$  is equivalent to

$$\frac{Y_{it} - f_{it} - V_{it}^{1/2}}{e_{it}} < \rho_i < \frac{Y_{it} - f_{it} + V_{it}^{1/2}}{e_{it}}$$

for all  $i$  and  $t$ . And the posterior distribution for  $\rho_i$  is truncated normal:

$$N(\mu_i, \sigma_i^2) I \left[ \max \left\{ \frac{Y_{it} - f_{it} - V_{it}^{1/2}}{e_{it}} \right\}, \min \left\{ \frac{Y_{it} - f_{it} + V_{it}^{1/2}}{e_{it}} \right\} \right].$$

v)  $\lambda$

$$\begin{aligned} P(\lambda|Z, V, \alpha, \beta, \rho) &= \pi(\lambda) \prod_{i,t} P(Z_{it}, V_{it}, \alpha, \beta, \rho_i|\lambda) \\ &\propto \pi(\lambda) \prod_{it} P(V_{it}|\lambda) \propto \lambda^{\alpha\lambda-1} e^{-\beta\lambda} \lambda^{3N(T-t_0)/2} e^{-\frac{\lambda}{2} \sum_{it} V_{it}} \\ &= G(\alpha\lambda + 3N(T-t_0)/2, \beta\lambda + \sum_{it} V_{it}/2). \end{aligned}$$

Since most of the conditional distributions above are truncated normals, a new and simple way of sampling truncated normal distributions developed by Damien and Walker (1997) is described in the Appendix.

## 4 Data Analysis

### *Simulated Data Analysis*

Two scenarios for the model developed in the previous section were studied: the *explosive* ( $\rho = 1.25$ ) case; and the *non-explosive* ( $\rho = .50$ ) case. Following Zellner, without loss of generality, we omitted the intercept term and set the slope parameter,  $\beta$  to equal 3.0. The prior distributions for  $\beta$  and  $\rho$ , respectively, were Normal(5.0, 4) and Normal(0.8, 1.0). Under each scenario, we simulated 50 samples. The Gibbs sampler detailed in the text was implemented. To ensure that the final samples were not autocorrelated, we took every 100th iterate after executing the chain for 100,000 iterations, and which took approximately 1/2 an hour on a Sun Sparc station, using FORTRAN. Based on the 1000 samples, we estimated the posterior parameter estimates for  $\beta$  and  $\rho$  under the two scenarios; the standard deviations are given in parentheses.

*Non-explosive*  $\beta = 3.03(0.04)$ ;  $\rho = 0.56(.09)$ .

*Explosive*  $\beta = 2.98(0.08)$ ;  $\rho = 1.25(2.0E - 14)$ .

### *Bank Data Analysis*

In recent years, following the pioneer work of Stern & Stewart Company, the financial metric, Market Value Added (MVA), is being recognized as a robust indicator of the performance of a firm (Stewart, 1991, 1994; Uyemura

et al. 1991, 1996; Copeland et al. 1996). Based on the concept of Economic Value Added (EVA), a cash-based model founded on the economic theory of residual income, MVA has been recognized as one of the most useful measures to assess the performance of a company. A growing number of FORTUNE 500 firms and investment analysts now subscribe to the *Stern Stewart Performance 1000* index. In companies such as Microsoft, General Electric, Coca-Cola, Philip-Morris, Whirlpool, Quaker Oats, Berkshire Hathaway, etc. senior management has made the pursuit of their MVA ranking as the primary goal; FORTUNE magazine now reports on the performance of companies using MVA as one of the financial indicators of a firm's viability; see, for example, Dempster (1997).

What is MVA? The wealth of shareholders is maximized by maximizing the difference between a company's total value (TV) and the total capital (TC) investors have vested in it. Hence,

$$MVA = TV - TC;$$

TV is also known Market Capitalization (MC). MC is defined as the product of the number of outstanding shares and the share price of a firm. It is this financial measure that is of interest in our context; i.e., MC forms the dependent variable in our model; see, also, Chen et al. (1997).

So what influences or drives MC (and whence MVA)? Clearly there are several macroeconomic factors (inflation, commodity prices, treasury bill return rates, foreign exchange rates, etc.) and microeconomic factors (price-to-earnings ratios, corporate announcements, debt-level ratios, analyst recommendations, etc.) that will likely influence MC. In fact, Stewart (1994) notes 164 factors (or measures) that are likely to influence the MC of a firm. But he also goes on to note that only 5 to 10 factors are really used in practice; furthermore, all such factors can be collapsed into a single influencing factor — Economic Value Added (EVA) — that Stern Stewart have trademarked.

What is EVA? EVA is simply a firm's expected performance discounted at the cost of capital. By definition, there is no *unique* method to calculating the EVA of a firm. The accounting protocols vary substantially among firms in an industry and across industries; see, for example, Imhoff et al. (1991), Uyemura et al (1996); Black et al. (1997 a,b). From a statistical perspective, EVA is positively correlated to MC (or MVA). It is evident that EVA from one time period will likely be affected by past EVA (autocorrelation); also, note there is substantial variability across firms within an industry, such as

the banking industry; hence the econometrician will likely entertain a random effects model (heteroscedastic influences); see, for example, Zellner (1971).

The EVA concept has usually been implemented as a management tool within an organization; see, for example, Baciadore et al (1998). Since projects that generate a larger EVA will lead to an increase in wealth, management can be awarded or reprimanded based on whether or not they have added or reduced value to the firm. Only recently, EVA is also being recognized as a marketing tool. This challenge was posed by *Citicorp Securities* to an MBA students team at the University of Michigan, which resulted in the research reported in Black et al. (1997 a, b). Larger banks provide services to customers of smaller banks for a fee. Before entering into a contract, the service provider would like to assess the feasibility of its client partner. Reversing this, a larger bank can market its services to smaller banks based on its assessment of the potential client's EVA. In this sense, EVA becomes a marketing tool. Also, EVA assessments can pave the way for mergers and takeovers.

A general EVA model for all banking institutions was developed in Black et al. (1997 a, b). The details of financial accounting protocols need not detain us. On the other hand we will concern ourselves with the questions: given a firm's EVA and MC is it possible to develop a statistical model that will allow us to predict the firm's future MC? how can we factor prior knowledge to better forecast the future worth of a firm? can an iterative, and straightforward mathematical system be developed so that a bank can routinely update its data base and forecasts?

As noted in Black et al. (1997 a b), obtaining data to calculate EVA for a given year is made substantially difficult by the inconsistency in the raw data needed to calculate EVA. Two sources (SNL and One Source) contain electronically stored bank profile data. Neither of the sources carry all past data needed to compute EVA. To verify whether these two sources of data could be combined for different years, for a specific bank, a consistency check must be carried out by comparing accounting information detailed in a firm's 10K report. The latter itself is a source of substantial variability for various reasons such as year-to-year changes in itemizing components in the balance sheet. The difficulty is enhanced the farther one goes back in the past.

We originally targeted 26 banks that were representative in terms of the US banking industry. These banks had varying asset structures ("large" to "small") and somewhat different internal accounting protocols. After close analysis, only 14 banks' data proved to be "clean" and reasonably consistent

for four years, 1993 to 1996. Thus the cross-sectional collection of 14 banks data over four years netted in 56 data points for MC and EVA. Fortunately, the 14 banks were representative of the US banking industry as a whole. Further, we dropped a “large” bank from this set of 14 banks. The motivation was to find out how well the model would predict these out-of-sample bank’s market capitalization (or share price).

We chose not to include any macroeconomic variables in the model based on the fact that the US economy was fairly stable with respect to the banking industry during this time period.

We refrained from including any microeconomic variables in the model *directly* mainly because the bank sponsoring the research was keen on studying the effect of only EVA on MC. Also, we reached an agreement that this would not be a one-off study; both the data-base and the statistical models would be updated periodically. Also, somewhat to our surprise, we found statistical practice in corporate finance to be somewhat archaic. We felt that suggesting more sophisticated methods at the outset may be off-putting to our customer. Instead, we were able to make a compelling argument in favor of increasing the sophistication in stages by demonstrating the value in using — to borrow a phrase of Arnold Zellner — *sophisticatedly simple* statistical procedures.

#### *Prior parameters*

Based on discussions with the bank’s analysts, and the segment of data that was “messy” we entertained the following prior choices for the parameters in the model developed in the previous section.  $\alpha \sim N(7.86e9, (3.0d8)^2)$ ;  $\beta \sim N(3.5e3, (8.0e2)^2)$ ; and  $\lambda \sim \text{Gamma}(5.0e - 22, 0.1)$ . Having classified the banks into “large” and “small” groups, based on their asset base, resulted in  $\rho_{large} \sim N(1.873, .70)$  and  $\rho_{small} \sim N(0.968, .70)$ .

#### *Posterior Estimates*

As an illustration, the posterior distributions, along with summary statistics, for  $\beta$  and the  $\rho$  parameters corresponding to the “largest” and “smallest” bank are given in Figure 1.

The sponsoring bank was interested in forecasting the market capitalization (or stock price) of another institution in its peer group. We used the  $\rho$  (see, Figure 1) corresponding to the largest bank in the pool of 13 banks as a proxy, and the EVA of the out-of-sample bank to forecast the latter’s stock price for the years 1994-1996; these predictions appear in Figure 2.

The posterior mean values from these predictions when compared to the actual values appear to be reasonable. We note here that these predictions will likely be more robust as more (clean) data is included in the analysis. Also, factoring in other variables would also help better assess the financial viability of a banking institution over time. The sponsoring bank, as noted earlier, have embarked on these as part of a long-term strategy within their econometrics unit.

## 5 Discussion

The scale mixture of uniform family appears to be a very useful way to encapsulate more realistic assumptions (such as heavy tailed behaviour) in a variety of modeling contexts. In this paper, theory pertaining to this new family was developed. A Bayesian illustrative analysis within the context of modeling market value of financial institutions was provided using data obtained from 14 banks. Further extensions of the general idea presented in this paper, in a variety of economic contexts, will be reported elsewhere.

## Acknowledgments

This research was supported by a fellowship from IBM Canada, and an EPSRC Realising Our Potential Award. Thanks to Andre McKoy, Vice-President of Securities, at *Citicorp, Inc.* for sponsoring and participating in the University of Michigan Business School MAP project, which led to this research.

## References

- Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society, Series B* **36**, 99-102.
- Baciadore, J., Boquist, J.A., Milbourn, T.T., and Thakor, A.V. (1996). A search for the best financial performance measure. *Journal of Applied Corporate Finance*. To appear.
- Black, F., Martinez, S., Ou, C., Reinhard, S., Rodriquez, A., Savage, T.W.E.,

- and Wilton, S. (1997). Economic value added model for banks, *Multidisciplinary Action Project Report*, The University of Michigan Business School, Ann Arbor.
- Black, F., Martínez, S., Ou, C., Reinhard, S., Rodriguez, A., Savage, T.W.E., Wilton, S., and Damien, P. (1997). Economic value added model to assess bank performance. *Submitted for Publication*.
- Box, G.E.P. and Tiao, G.C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley, Massachusetts.
- Chen, T., Magnus, E., Marks, A. and Damien, P. (1997). Statistical insights into the performance of EVA. *Submitted for publication*.
- Choy, S.T.B. and Smith, A.F.M. (1997). On robust analysis of a normal location parameter. *Journal of the Royal Statistical Society, Series B* 59, 463-474.
- Copeland, T., Koller, T., and Murrin, J. (1996). *Valuation, Measuring and Managing the Value of of Companies*. John Wiley & Sons, Inc. New York.
- Damien, P., Wakefield, J.C. and Walker, S.G. (1997). Gibbs sampling for Bayesian nonconjugate models using auxilliary variables, *Journal of the Royal Statistical Society, B*. To appear.
- Damien, P. and Walker, S.G. (1997). Sampling truncated normal, gamma and beta densities. Revised for *Statistics and Computing*.
- Dempster, M. (1997). EVA: get it while its hot. *Monroe Street Journal*. October 20, 1997, page 17. University of Michigan Business School Publication.
- Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer-Verlag, New York.
- George, E.I. and McCulloch, R. (1993). Variable selection via Gibbs sampling, *Journal of the American Statistical Association*, 88, 881-889.
- Imhoff, E.A., Lipe, R.C., and Wright, D.W. (1991). Operating leases: im-



pect of constructive capitalization, *Accounting Horizons*, March issue.

Jacquier, E., Polson, N.G., and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, **12**, 371-389.

Karim, A.M. and Paruolo, P. (1996). Two mixed normal densities from cointegration analysis. *Econometrica*, **65**, 671-680.

Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion), *Journal of the Royal Statistical Society, B* **34**, 1-41.

Stewart, G.B. (1991). *The Quest for value*. Harper Collins, New York.

Stewart, G.B. (1994). EVA: fact and fantasy. *Journal of Applied Corporate Finance*, **7**, 2, 71-84.

Uyemura, D.G. and Kantor, C.C. (1997). Mind & Muscle, *Banking Strategies*, Volume 73, January/February issue, 57-59.

Uyemura, D.G., Kantor, C.C., Pettit, J.M., and Stern & Stewart Co. (1992). EVA for banks: value creation, risk management, and profitability measurement, *Journal of Applied Corporate Finance*, **9**.

Zaik, E., Walter, J., Kelling, G. and James, C. (1996). RAROC at Bank of America: from theory to practice, *Journal of Applied Corporate Finance*, **9**.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika* **74**, 646-648.

Zellner, A. (1987). *An Introduction to Bayesian inference in Econometrics*. Robert E. Kreiger Publishers, Florida.

## Appendix

While the focus of this paper was on the use of EVA to predict MC, it may be desirable to include macroeconomic variables in the model formulation over time as the “clean” data base gets updated. This would then lead to posterior conditional distributions that will be proportional to a truncated multivariate normal. Sampling from such a distribution is accomplished as follows; see, Damien & Walker (1997).

$$f_{X_1, \dots, X_p}(x_1, \dots, x_p) \propto \exp\left(-1/2(x - \mu)' \Sigma^{-1}(x - \mu)\right) I(x \in A),$$

where we assume that the bounds for  $x_i$  given  $x_{-i}$  are available and given by, say,  $(a_i, b_i)$ . Therefore

$$f_{X_i|X_{-i}}(x_i|x_{-i}) \propto \exp\left(-1/2(x_i - \nu_i)^2/\sigma_i^2\right) I(x_i \in (a_i, b_i)),$$

for  $i = 1, \dots, p$ , are the full conditionals, and  $\nu_i = \mu_i - \sum_{j \neq i} (x_j - \mu_j) e_{ij}/e_{ii}$  and  $\sigma_i^2 = 1/e_{ii}$ , where  $e_{ij}$  is the  $ij$ th element of  $\Sigma^{-1}$ .

However, since we are already in a Gibbs sampler it seems appropriate to implement the auxiliary variable idea (Damien et al. 1997). We do not need to introduce  $p$  latent variables, one is sufficient.

We define the joint density of  $(X_1, \dots, X_p, Y)$  by

$$f_{X_1, \dots, X_p, Y}(x_1, \dots, x_p, y) \propto \exp(-y/2) I(y > (x - \mu)' \Sigma^{-1}(x - \mu)) I(x \in A).$$

The full conditional distributions are given by

$$f_{X_i|X_{-i}, Y}(x_i|x_{-i}, y) \propto I(x_i \in A_i),$$

where

$$A_i = (a_i, b_i) \cap B_i,$$

and  $B_i$  is the set  $\{x_i|x_{-i} : (x - \mu)' \Sigma^{-1}(x - \mu) < y\}$  and so the bounds for  $B_i$  are obtained by solving a quadratic equation. The full conditional for  $Y|X$  is clearly a truncated exponential distribution which can be sampled using the cdf inversion technique.

Therefore we have a Gibbs sampler which runs on  $p + 1$  full conditionals which can all be sampled directly using uniform variates.

In the context of the analyses in this paper, the above algorithm reduces to the following special case of sampling a univariate truncated normal. Suppose we wish to sample from the density given by

$$f_X(x) \propto \exp(-x^2/2) I(x \in (a, b)).$$

Again, we introduce the latent variable  $Y$  which has joint density with  $X$  given by

$$f_{X,Y}(x, y) \propto I_{(0, \exp(-x^2/2))}(y) I(x \in (a, b)),$$

leading to the new full conditionals:

$$Y|(X = x) \sim U(0, \exp(-x^2/2))$$

$$X|(Y = y) \sim U(\max\{a, -\sqrt{-2 \log y}\}, \min\{b, \sqrt{-2 \log y}\}).$$

The algorithm extends the Gibbs loop by one more full conditional, which is a uniform distribution. The new full conditional for  $X$  is also uniform.

There are alternative ways to sample truncated normal distributions; see, for example, Devroye (1986). However, typically, these methods require rejection and “tuning”; these factors, of course, depend on the likelihood function so that each time the database is updated maximization routines will have to be engineered. The algorithm proposed above bypasses this difficulty. The statistician implementing the model in a bank simply runs the Gibbs sampler like a “black-box” method after merely updating the prior parameters. This obviates specialized training needed to search for dominating densities, maximizing functions, or worry about “tuning” each time the likelihood/prior gets updated. Of course, alternative methods may be more efficient, but we found that simplicity in coding and execution was the preferred choice. In any case, Damien and Walker (1997) demonstrate that the algorithm discussed above is *at least* as efficient as other methods.

Yet another reason for preferring a Gibbs sampler that involves no rejection sampling and/or Metropolis-Hastings detours is for the reason given in the first paragraph in this Appendix. We noted in the text that it is very unlikely that the number of independent variables, in this context, will ever exceed six or seven. Model selection in such cases is easily done within the Gibbs loop using the method of George and McCulloch (1992). These authors’ method

is very elegant when the number of dimensions are few. Combining their idea with the method in Damien et al. (1997) will result in an almost “black-box” implementation strategy within a bank’s statistical services unit.

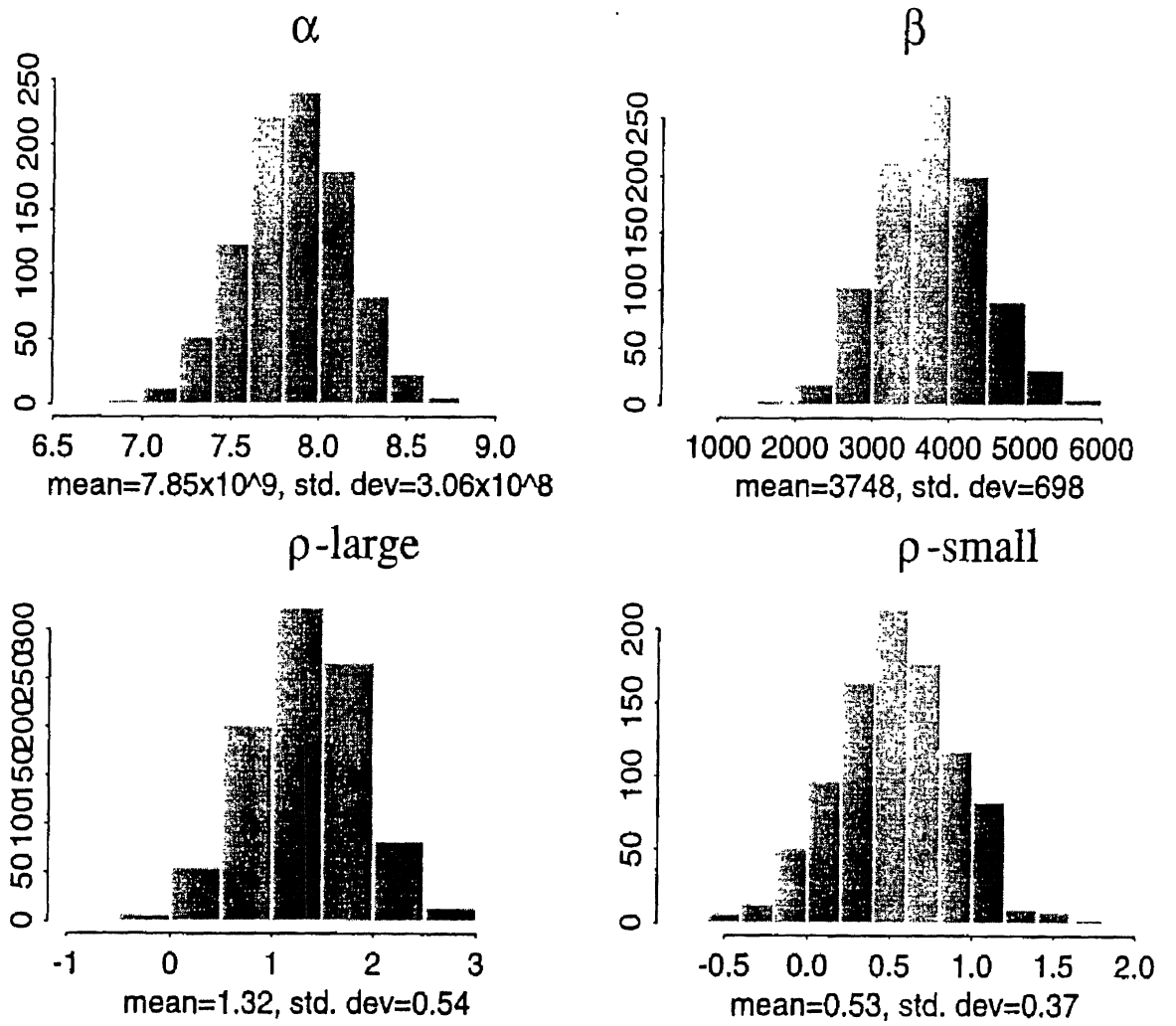


Figure 1: Posterior distributions of  $\alpha$ ,  $\beta$  and  $\rho$ .

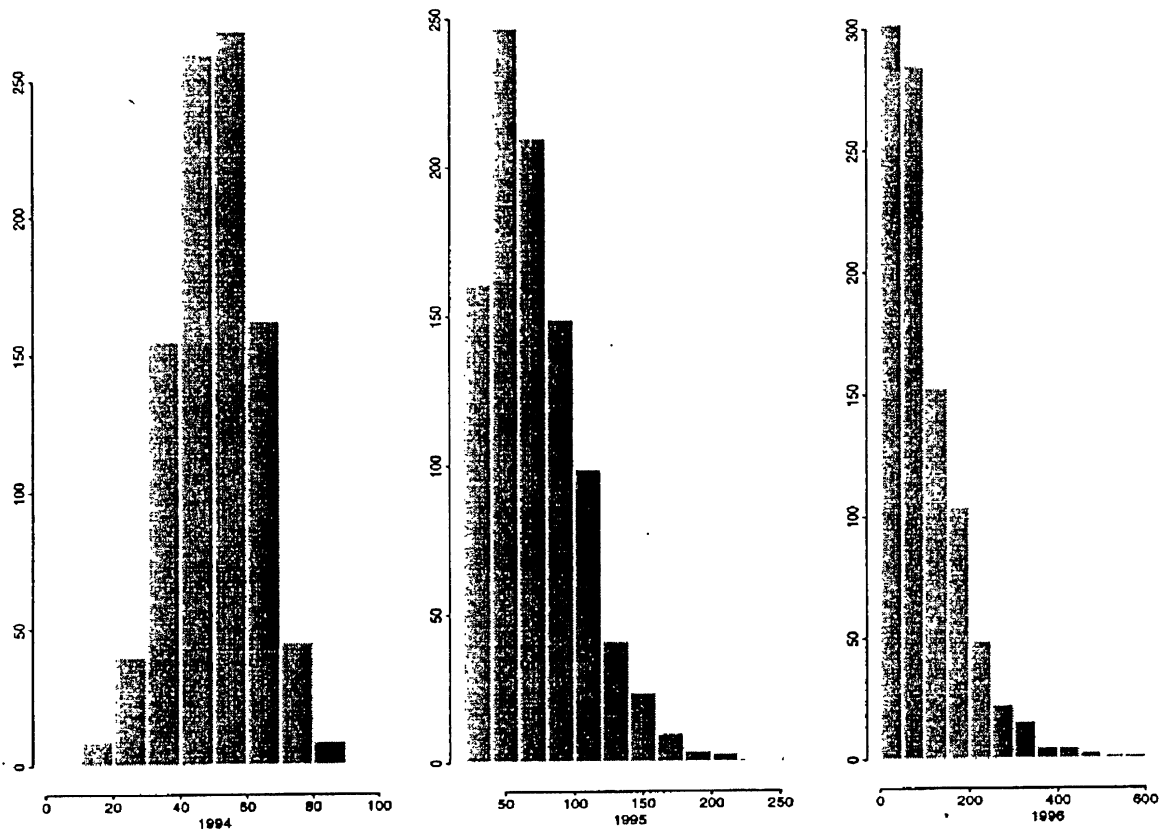


Figure 2: Posterior predictive distributions of a “large” bank.  
 1994: Actual Price: 41.38; Predictive mean=50.4, std. dev=12.6,  
 1995: Actual Price: 67.25; Predictive mean=72.0, std. dev=34.3,  
 1996: Actual Price: 103.0; Predictive mean=103.1, std. dev=85.3.