Journal of Forecasting
J. Forecast. **24**, 39–59 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.933

Forecasting Stock Prices Using a Hierarchical Bayesian Approach

JUN YING,[1] LYNN KUO[2]* AND GIM S. SEOW[2]
[1] *Indiana University School of Medicine, USA*
[2] *University of Connecticut, USA*

ABSTRACT

The Ohlson model is evaluated using quarterly data from stocks in the Dow Jones Index. A hierarchical Bayesian approach is developed to simultaneously estimate the unknown coefficients in the time series regression model for each company by pooling information across firms. Both estimation and prediction are carried out by the Markov chain Monte Carlo (MCMC) method. Our empirical results show that our forecast based on the hierarchical Bayes method is generally adequate for future prediction, and improves upon the classical method.

KEY WORDS autoregression; hierarchical mixture priors; MCMC; prediction; simultaneous estimation

INTRODUCTION

Recent development in the security valuation literature has provided a model that relates the stock price to its book value and expected future earnings. It includes the work of Bernard (1995), Feltham and Ohlson (1995), Lang and Lundholm (1996) and Ohlson (1991, 1995). These studies develop a logically consistent framework for thinking about equity valuation using accounting data. The primary objectives of this paper are to empirically evaluate the adequacy of the security valuation model and to use it to forecast stock prices.

The security valuation model has been developed based on a single firm. The empirical literature in both accounting and finance is based primarily on classical statistical techniques. In this paper, we apply an innovative statistical method, a hierarchical Bayesian (HB) approach that allows improved estimation of the regression coefficients by sharing information across firms. Using 14 years of quarterly stock price data, accounting book values and expected future earnings for 28 companies included in the Dow Jones Industrial Average, we show that the forecast based on the HB model is consistently superior to those obtained using the classical approach.

The original Ohlson model proposes that the stock price is a linear function of the company's book value per share and expected excess earnings per share for the following four periods with normally distributed innovation terms. Each company has its own coefficients; we use $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots,$

* Correspondence to: Lynn Kuo, Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, USA.
E-mail: lynn@stat.uconn.edu

$\beta_{i6})' = (\beta_{i,1}, \ldots, \beta_{i,6})'$ to denote the regression coefficients of the intercept, book value, each of the expected excess earnings for the following four periods for the ith company with $i = 1, \ldots, n$. The model can be described as follows for all $t = 0, \ldots, T$;

$$y_{it} = \beta_{i,1} + \beta_{i,2} v_{it} + \left(\sum_{k=1}^{4} \beta_{i,k+2} w_{i,t+k} \right) + u_{it} \tag{1}$$

where y_{it} denotes the ith company's stock price per share at time t; v_{it} denotes the book value per share of stock i at time t; $w_{i,t+k}$ denotes the expected excess earnings per share of stock i in the kth period after time t; and u_{it} is the innovation term for y_{it}. In fact, Ohlson (1991) proposes the expected excess earnings as

$$w_{i,t+k} = E[s_{i,t+k} - r_t v_{i,t+(k-1)}] \tag{2}$$

where $s_{i,t+k}$ denotes the earnings per share of stock i in the $(t + k)$th period for $k = 1, \ldots, 4$ and r_t is the discount rate at time t. The expected excess earnings are, however, not available in terms of financial accounting data. We use

$$w_{i,t+k} = E(s_{i,t+k}) - r_t v_{i,t+(k-1)} \tag{3}$$

instead of (2). In this paper, we use $\boldsymbol{x}_{it} = (x_{i1t}, \ldots, x_{i6t})' = (1, v_{it}, w_{i,t+1}, w_{i,t+2}, w_{i,t+3}, w_{i,t+4})'$ to denote the vector of the predictors of the ith company at time t.

We employ a HB approach to inferences in the Ohlson model. A general form of HB model is presented in Lindley and Smith (1972). In this paper, we extend the HB approach in two distinct aspects. First, each regression coefficient $\boldsymbol{\beta}_i$ is modelled as a mixture of normal distributions with unknown hyperparameters. Thus, the hierarchical setup has two sets of parameters to be estimated. One is the set of parameters of interest and the other is the set of hyperparameters that model the parameters. Second, the innovation terms u_{it} are modelled in the following ways: (a) u_{it} has a first-order autoregressive structure (AR(1)) for each company i, that is, $\text{corr}(u_{i,t}, u_{i,t-1}) = \rho_i$ for $t = 1, \ldots, T$; and (b) u_{it} are allowed to have heterogeneous variance among the companies, that is, $\text{var}(u_{it}) = \sigma_i^2/(1 - \rho_i^2)$ for $i = 1, \ldots, n$ and for all t. The unknown parameters (σ_i^2, ρ_i) are modelled with known proper priors. In fact, the AR(1) structure can easily be extended to a more general autoregressive moving average (ARMA) structure.

A widely used HB structure for the regression coefficients is to assume that $\boldsymbol{\beta}_i$ are i.i.d. from a normal distribution with its hyperparameters being modelled by the hyperpriors. The use of this hierarchical setup can 'borrow strength' across the different individuals and bring 'shrinkage' effects to the posterior mean of the regression coefficients. However, our data analysis shows that the estimates of the regression coefficients in some companies are quite far away from the majority of other companies. The use of a normal distribution for the i.i.d. regression coefficients is not adequate and will cause 'over-shrinkage' to these estimates. Thus, we need a model that is flexible enough to accommodate these 'outliers' while at the same time adjusting the estimates for shrinkage effects. A natural choice would be a mixture of two normal distributions. The first component of the mixture retains the company's individual mean; and the second component of the mixture shares the common mean among different companies. Müller and Rosner (1997) provide a detailed discussion on hierarchical mixture priors.

Bayesian HB methodology has been applied to the analysis of variance–covariance matrices of the innovation terms u_{it}. For example, Gelfand *et al*. (1990), Wakefield *et al*. (1994) and Rosenberg *et al*. (1999) model u_{it} as i.i.d. random variables over i and t. Gelfand and Sfiridis (1996) and Kasim and Raudenbush (1998) extend the homogeneous variance models to the heterogeneous variance models, while keeping the conditional independent structure of the innovation terms within each company. Kasim and Raudenbush (1998) relax the conditional independence assumption and assume a compound symmetry (constant correlation) structure for the innovation terms within each company. The variances are assumed to be different, but the correlation coefficients are the same among the companies.

In order to model the innovation terms as a function of times, Albert and Chib (1993) and Chib (1993) apply a Bayesian framework to the autoregressive models in their data analysis with homogeneity of variance. Chib and Greenberg (1995) extend the HB method from the autoregressive process to the ARMA(p, q) process.

Preliminary data analysis (not presented here) shows that (a) the variance differs widely among the 28 companies, and (b) for each company the stock prices follow an AR(1) process. Moreover, the autocorrelation coefficients vary substantially among the companies. Therefore, we extend the methodologies in Albert and Chib (1993) and Chib (1993) to accommodate the heterogeneity of variances and autocorrelations.

The advantages of the hierarchical formulation can be summarized. (a) We retain the firm-specific feature of the whole assembly. Each firm's specific regression coefficients will provide us with information on how sensitive the firm's stock price is to the changes of book values and expected excess earnings. (b) We provide a more information-efficient modelling for the data and a full synthesis on all the data for the regression coefficients. Because the regression coefficients for the ith company are not only affected by the data from the ith company, but also affected by other companies' data (perhaps to a lesser degree), this full synthesis provides a more satisfactory solution to modelling our data. (c) This full synthesis is known to provide estimates that have smaller mean squared errors than the usual least squared or classical estimates (Stein, 1955).

On the methodology for inference, we will employ the MCMC algorithm (Gelfand and Smith, 1990; Chen *et al*., 2000) to provide point and interval estimates for the unknown parameters. The MCMC algorithm simulates random parameters (or blocks of parameters) from their full conditional distribution given the data by constructing a Markov chain, so the stationary distribution of the Markov chain is the desirable posterior distribution. The sampling-based approach circumvents the difficulties in evaluating multidimensional integrals needed in Bayesian inference. It allows us to obtain any features, such as mean, variance, quantiles and histograms, of the posterior distribution.

In addition to estimation, we are also concerned with the issues of model adequacy and model selection. We compare the out-of-sample performance of our method to that of the classical approach for two timeframes: one is from the 40th quarter to the 54th quarter, the other is the 52nd quarter (the 3rd quarter of 1997). The criteria we use for comparison include the predictive mean squared error (PMSE), the predictive mean absolute error (PMAE), the predictive mean absolute relative deviation (PMARD) and the coverage probability for the forecast intervals. Our results show that the HB method based on the mixture priors for $\boldsymbol{\beta}_i$ has better forecasting power than the HB method with a single normal prior, that further improves upon the classical method for both timeframes. The improvement is more dramatic for the 52nd quarter. The two sets of priors in the HB method are also compared by using the prequential pseudo-Bayes factor (PPBF). The result supports the mixture prior assumption.

MODEL

The Ohlson model assumes that the error terms are conditionally independent among companies and an AR(1) structure in time periods within each company. We can express the Ohlson model using the following expressions:

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta}_i + u_{it} \\ u_{it} &= \rho_i u_{i,t-1} + \varepsilon_{it} \\ \varepsilon_{it} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\ \sigma_i^2) \quad \text{for } t = 1,\ \ldots,\ T \end{aligned} \tag{4}$$

where y_{it} is the observation of company i at time t, $\mathbf{x}_{it} = (x_{i1t}, \ldots, x_{iKt})'$ is the vector of K predictors for the ith company at the tth period, with $x_{i1t} = 1$, $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{iK})'$ is the vector of intercept and slope coefficients of the predictors. The AR(1) structure is described by the innovation term. The residue ε_{it} is independent of $\varepsilon_{i't}$ for $i \neq i'$. The observation at time 0 is considered to be stationary, that is, for $i = 1, \ldots, n$,

$$y_{i0} \sim \mathcal{N}\left(\mathbf{x}_{i0}'\boldsymbol{\beta}_i,\ \frac{\sigma_i^2}{1-\rho_i^2}\right) \tag{5}$$

Conditioning on ρ_i, it is easy to apply a change of variable technique to (4). Let $\mathbf{y}_i^* = (y_{i1}^*, \ldots, y_{iT}^*)'$, $\mathbf{X}_i^* = (\mathbf{x}_{i1}^{*'}, \ldots, \mathbf{x}_{iT}^{*'})'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, where $y_{it}^* = y_{it} - \rho_i y_{i,t-1}$ and $\mathbf{x}_{it}^* = \mathbf{x}_{it} - \rho_i \mathbf{x}_{i,t-1}$ for $t = 1, \ldots, T$. Equation (4) can be written as

$$\begin{aligned} \mathbf{y}_i^* &= \mathbf{X}_i^*\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\varepsilon}_i &\sim \mathcal{N}_T(0,\ \sigma_i^2\mathbf{I}) \end{aligned} \tag{6}$$

We use the notation $\mathcal{N}_T(\boldsymbol{\mu},\ \Sigma)$ to denote a T-variate normal distribution with mean $\boldsymbol{\mu}$ and variance–covariance matrix Σ. We use $\mathbf{I}$ to denote the identity matrix of rank T.

To construct HB estimators, we use the following hierarchical model. At the first stage, the observations $\mathbf{y}_i^*$ and y_{i0} are described by the parameters $\{\boldsymbol{\beta}_i,\ \sigma_i^2,\ \rho_i)$ for each $i = 1, \ldots, n$. This relationship can be expressed in the likelihood function $\mathcal{L}$ for $\{\mathbf{y}_i^*,\ y_{i0};\ i = 1, \ldots, n\}$. Under the assumption that the observations are conditionally independent among the companies, we have

$$\prod_{i=1}^{n} \mathcal{L}\left(\boldsymbol{\beta}_i,\ \sigma_i^2,\ \rho_i \middle| \mathbf{y}_i^*,\ y_{i0}\right) = \prod_{i=1}^{n}\left[\mathcal{N}_T\left(\mathbf{y}_i^* \middle| \mathcal{X}_i^*\boldsymbol{\beta}_i,\ \sigma_i^2\mathbf{I}\right) \times \mathcal{N}\left(y_{i0} | \mathbf{x}_{i0}'\boldsymbol{\beta}_i,\ \frac{\sigma_i^2}{1-\rho_i^2}\right)\right] \tag{7}$$

At the second stage, we provide the prior distributions for the parameters $\{\boldsymbol{\beta}_i,\ \sigma_i^2,\ \rho_i;\ i = 1, \ldots, n\}$, and we let $\boldsymbol{\beta}_i$, σ_i^2 and ρ_i be conditionally independent. In particular, we model $\boldsymbol{\beta}_i$ as a mixture of two normal distributions:

$$\pi(\boldsymbol{\beta}_i | \mu_i,\ \Sigma_i,\ \eta, \Omega,\ w_i) = w_i \mathcal{N}_K(\boldsymbol{\beta}_i | \mu_i,\ \Sigma_i) + (1 - w_i)\mathcal{N}_K(\boldsymbol{\beta}_i | \eta,\ \Omega)_S \tag{8}$$

where $\{\mu_i, \Sigma_i, \omega_i; i = 1, \ldots, n\}$ and $\{\eta, \Omega\}$ are the unknown hyperparameters. The first component that allows individual means brings more dispersion among $\{\boldsymbol{\beta}_i; i = 1, \ldots, n\}$, and thus also prevents over-shrinkage of the estimates for the outliers. The second component that shares the same mean provides a mechanism for sharing information across firms, as is usually done in the HB method. The prior distribution for σ_i^2 is assumed to be an inverse gamma, that is, $\sigma_i^2 \sim \mathcal{IG}(a_i, b_i)$ with mean $\dfrac{b_i}{a_i - 1}$ and variance $\dfrac{b_i^2}{(a_i - 1)^2 (a_i - 2)}$. Both a_i and b_i are known. We assume a truncated normal for the correlation coefficient ρ_i, that is, $\rho_i \sim \mathcal{N}(\rho_{0i}, \sigma_{\rho_i}^2)\, I\,(-1 \le \rho_i \le 1)$, where the hyperparameters ρ_{0i} and $\sigma_{\rho_i}^2$ are known. In order to make inferences on ρ_i primarily from the data rather than from the prior, we usually set $\sigma_{\rho_i}^2$ sufficiently large. The prior at the second stage can be written as

$$\prod_{i=1}^{n} [\pi(\boldsymbol{\beta}_i | u_i, \Sigma_i, \eta, \Omega, \omega_i) \times \pi(\sigma_i^2 | a_i, b_i) \times \pi(\rho_i | \rho_{0i}, \sigma_{\rho_i}^2)] =$$

$$\prod_{i=1}^{n} \{[\omega_i \mathcal{N}_K(\boldsymbol{\beta}_i | \mu_i, \Sigma_i) + (1 - \omega_i) \mathcal{N}_K(\boldsymbol{\beta}_i | \eta, \Omega)] \times \mathcal{IG}(a_i, b_i) \times \mathcal{N}(\rho_{0i}, \sigma_{\rho_i}^2) 1(-1 \le \rho_i \le 1)\} \quad (9)$$

The unknown hyperparameters $\{\mu_i, \Sigma_i, \omega_i; i = 1, \ldots, n\}$ and $\{\eta, \Omega\}$ are modelled at the third stage. We assume they are conditionally independent of each other and let $\mu_i \sim \mathcal{N}_K(\mathbf{m}, \mathbf{M})$, $\eta \sim \mathcal{N}_K(\mathbf{h}, \mathbf{H})$, $\Sigma_i^{-1} \sim W_K(\upsilon, \mathbf{V})$ (Wishart distribution with υ degree of freedom and scale matrix V), $\Omega^{-1} \sim W_K(\mathrm{q}, \mathbf{Q})$ and $\omega_i \sim Be(r, s)$ (Beta distribution) for $i = 1, \ldots, n$. The parameters $\{\mathbf{m}, \mathbf{M}, \mathbf{h}, \mathbf{H}, \upsilon, \mathbf{V}, q, \mathbf{Q}, r, s\}$ are all assumed to be known. A schematic diagram for the ith firm is given in Figure 1.

Under the circumstance that $\omega_i = 0$ for all i, we have the non-mixture model in which all $\boldsymbol{\beta}_i$ $(i = 1, \ldots, n)$ share the same mean and variance across firms.

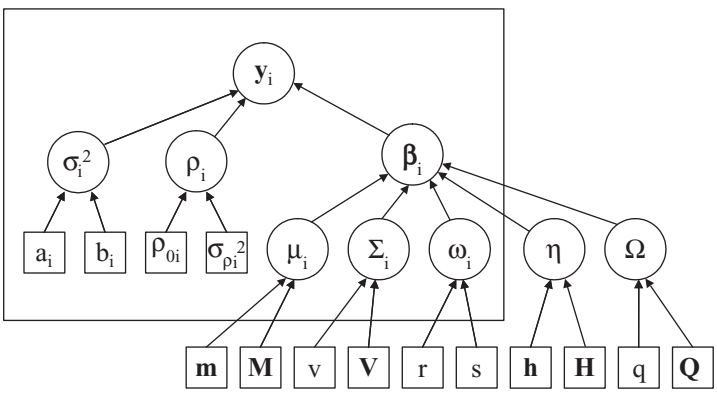

Figure 1. Schematic diagram for the ith firm $(i = 1, \ldots, n)$. The circles contain the unknown parameters and hyperparameters, the squares contain the known parameters. All other firms in addition to the ith firm share the same parameters $\{\eta, \Omega, \mathbf{m}, \mathbf{M}, \ldots, \mathbf{q}, \mathbf{Q}\}$

GIBBS SAMPLING

We first summarize the full conditional distributions for the parameters $\{\boldsymbol{\beta}_i, \sigma_i^2, \rho_i; i = 1, \ldots, n\}$.

(a) Note

$$\boldsymbol{\beta}_i|. \sim \omega_i^* \mathcal{N}_K\left(\boldsymbol{\beta}_i|\mu_i^*, \Sigma_i^*\right) + \left(1-\omega_i^*\right)\mathcal{N}_K\left(\boldsymbol{\beta}_i|\eta_i^*, \Omega_i^*\right) \tag{10}$$

where

$$\Sigma_i^* = \left[\Sigma_i^{-i} + \frac{\mathbf{X}_i^{*\prime}\mathbf{X}_i^* + (1-\rho_i^2)\mathbf{x}_{i0}\mathbf{x}_{i0}'}{\sigma_i^2}\right]^{-1}$$

$$\mu_i^* = \Sigma_i^*\left[\Sigma_i^{-i}\mu_i + \frac{\mathbf{X}_i^{*\prime} y_i^* + (1-\rho_i^2)y_{i0}\mathbf{x}_{i0}}{\sigma_i^2}\right]$$

$$\Omega_i^* = \left[\Omega^{-1} + \frac{\mathbf{X}_i^{*\prime}\mathbf{X}_i^* + (1-\rho_i^2)\mathbf{x}_{i0}\mathbf{x}_{i0}'}{\sigma_i^2}\right]^{-1}$$

$$\eta_i^* = \Omega_i^*\left[\Omega^{-1}\eta + \frac{\mathbf{X}_i^{*\prime} y_i^* + (1-\rho_i^2)y_{i0}\mathbf{x}_{i0}}{\sigma_i^2}\right]$$

$$\omega_i^* = \frac{\omega_i A_i}{\omega_i A_i + (1-\omega_i)B_i}$$

with

$$A_i = \left(\frac{\left|\Sigma_i^*\right|}{|\Sigma_i|}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\mu_i'\Sigma_i^{-1}\mu_i - \mu_i^{*\prime}\Sigma_i^{*-1}\mu_i^*\right]\right\}$$

$$B_i = \left(\frac{\left|\Omega_i^*\right|}{|\Omega|}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\eta_i^{*\prime}\Omega^{-1}\eta - \eta'\Omega_i^{*-1}\eta_i^*\right]\right\}$$

Thus, $\boldsymbol{\beta}_i$ is updated by first generating a Bernoulli random variable $\omega_i \sim \mathcal{B}(1, \omega_i^*)$, and then generating $\boldsymbol{\beta}_i \sim \mathcal{N}_K(\boldsymbol{\beta}_i|\mu_i^*, \Sigma_i^*)$ if $\omega_i = 1$ or generating $\boldsymbol{\beta}_i \sim \mathcal{N}_K(\boldsymbol{\beta}_i|\eta_i^*, \Omega_i^*)$ if $\omega_i = 0$.

(b)

$$\sigma_i^2|. \sim IG\left(a_i^*, b_i^*\right) \tag{11}$$

where $a_i^* = a_i + \frac{T+1}{2}$, $b_i^* = b_i + \frac{1}{2}\left(y_i^* - \mathbf{X}_i^*\boldsymbol{\beta}_i\right)'\left(y_i^* - \mathbf{X}_i^*\boldsymbol{\beta}_i\right) + \frac{1-\rho_i^2}{2}\left(y_{i0} - \mathbf{x}_{i0}'^*\boldsymbol{\beta}_i\right)^2$.

(c) Simulating the correlation coefficients ρ_i. Let $\mathbf{u}_{i1} = (u_{i1}, \ldots, u_{iT})'$ and $\mathbf{u}_{i0} = (u_{i0}, \ldots, u_{i,T-1})$, the full conditional distribution of ρ_i can be written as

$$\begin{aligned}\pi(\rho_i|.) &\propto (1-\rho_i^2)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_{\rho_i}^{2*}}\left(\rho_i - \rho_{0i}^*\right)^2\right\} 1\ (-1 \le \rho_i \le 1) \\ &= \exp\left\{\frac{1}{2}\log(1-\rho_i^2) - \frac{1}{2\sigma_{\rho_i}^{2*}}\left(\rho_i - \rho_{0i}^*\right)^2\right\} 1\ (-1 \le \rho_i \le 1) \end{aligned} \tag{12}$$

where

$$\sigma_{\rho_i}^{2*} = \left[\frac{1}{\sigma_{\rho_i}^2} + \frac{\mu_{i0}'\mu_{i0} - \mu_{i0}^2}{\sigma_i^2}\right]^{-1}$$

$$\rho_{0i}^* = \sigma_{\rho_i}^{2*}\left[\frac{\rho_{0i}}{\sigma_{\rho_i}^2} + \frac{\mu_{i0}'\mu_{i0}}{\sigma_i^2}\right]$$

We cannot obtain a simple form of the conditional posterior density of ρ_i from which the sample value can be generated directly. However, it is easy to show that $\pi(\rho_i|.)$ is log-concave and hence we can use the adaptive rejection sampling method (Gilks and Wild, 1992) to update ρ_i.

We next describe a data augmentation step in the Gibbs sampler to facilitate the generation of the hyperparameters.

(d) Data augmentation step of simulating the auxiliary variables z_i. The prior mixtures of $\boldsymbol{\beta}_i$ in (8) usually make it difficult to sample the unknown hyperparameters $\{\mu_i, \Sigma_i, \omega_i; i = 1, \ldots, n\}$ and $\{\eta, \Omega\}$. Therefore, we introduce a latent variable z_i as in Dey *et al*. (1995). The latent variable z_i can be simulated independently from the Bernoulli distribution $\mathcal{B}(1, p_i)$ where

$$p_i = \frac{\omega_i \mathcal{N}_K(\boldsymbol{\beta}_i|\mu_i, \Sigma_i)}{\omega_i \mathcal{N}_K(\boldsymbol{\beta}_i|\mu_i, \Sigma_i) + (1-\omega_i)\mathcal{N}_K(\boldsymbol{\beta}_i|\eta, \Omega)} \tag{13}$$

We then consider the joint density of $(\boldsymbol{\beta}_i, z_i)$ that is useful in deriving the conditional density of the hyperparameters.

Finally, we provide the full conditional distributions for the hyperparameters $\{\mu_i, \Sigma_i, \omega_i; i = 1, \ldots, n\}$, $\{\eta, \Omega\}$ based on their priors and the joint density of $(\boldsymbol{\beta}_i, z_i)$.

(e)

$$\mu_i|. \sim \mathcal{N}_K\left(\mathbf{m}_i^*, \mathbf{M}_i^*\right) \tag{14}$$

where $\mathbf{M}_i^* = (\mathbf{M}^{-1} + z_i\Sigma_i^{-1})^{-1}$ and $\mathbf{m}_i^* = \mathbf{M}_i^*(\mathbf{M}^{-1}\mathbf{m} + z_i\Sigma_i^{-1}\boldsymbol{\beta}_i)$.

(f)

$$\eta|. \sim \mathcal{N}_K(\mathbf{h}^*, \mathbf{H}^*) \tag{15}$$

where $\mathbf{H}^* = (\mathbf{H}^{-1} + (n - z^+)\Omega^{-1})^{-1}$, $\mathbf{h}^* = \mathbf{H}^*\left[\mathbf{H}^{-1}\mathbf{h} + \Omega^{-1}\left(\sum_{i=1}^{n}(1 - z_i)\boldsymbol{\beta}_i\right)\right]$ and $z^+ = \sum_{i=1}^{n} z_i$.

(g)

$$\Sigma_i^{-1}|. \sim \mathcal{W}_K\left(\upsilon_i^*, \mathbf{V}_i^*\right) \tag{16}$$

where $\upsilon_i^* = \upsilon + z_i$ and $\mathbf{V}_i^* = [\mathbf{V}^{-1} + z_i(\boldsymbol{\beta}_i - \mu_i)(\boldsymbol{\beta}_i - \mu_i)']^{-1}$.

(h)

$$\Omega^{-1}|. \sim \mathcal{W}_K(q^*, \mathbf{Q}^*) \tag{17}$$

where $q^* = q + z^+$ and $\mathbf{Q}^* = \left[\mathbf{Q}^{-1} + \sum_{i-1}^{n}(1 - z_i)(\boldsymbol{\beta}_i - \eta)(\boldsymbol{\beta}_i - \eta)'\right]^{-1}$.

(i)

$$\omega_i|. \sim \mathcal{B}e\left(r_i^*, s_i^*\right) \tag{18}$$

where $r_i^* = r + z_i$ and $s_i^* = s + 1 - z_i$.

The detailed procedures for deriving the above full conditional distributions are given in Ying *et al*. (2001). All the distributions except (c) are in standard form and therefore it is straightforward to generate the random variates. As described below, the MCMC procedure consists of steps 1 through 3, performed iteratively. The steps in 1 and 3 consist of substeps that are carried out sequentially for a single chain. We can also replicate the Markov chain by drawing independent initial values of the parameters and hyperparameters.

1. Update the parameters $\{\boldsymbol{\beta}_i, \sigma_i^2, \rho_i; i = 1, \ldots, n\}$ given the hyperparameters and data by carrying out the following substeps 1.1 to 1.3 independently for each $i = 1, \ldots, n$.
   - 1.1. Generate $\boldsymbol{\beta}_i$ from the conditional mixture normal density functions in (10) with the method discussed in (a).
   - 1.2. Generate σ_i^2 directly from the inverse gamma distribution described in (11).
   - 1.3. Update ρ_i as in (12) by applying the adaptive rejection sampling method.
2. Generate the latent variable z_i from a Bernoulli distribution discussed in (d).
3. Generate the hyperparameters $\{\mu_i, \Sigma_i, \omega_i; i = 1, \ldots, n\}$ and $\{\eta, \Omega\}$ given the values of the parameters in steps 1 and 2:
   - 3.1. Generate ω_i from a Beta distribution given in (18).
   - 3.2. Generate μ_i and Σ_i from the full conditional distributions described in (14) and (16), respectively.
   - 3.3. Generate η and Ω^{-1} from the full conditional distributions described in (15) and (17), respectively.

The HB method with single prior on $\boldsymbol{\beta}_i$ can be considered as a special case of the HB method with mixture prior by letting both ω_i (hence ω_i^*) and z_i be fixed at 0. It is straightforward to see that the full conditional distributions of $\boldsymbol{\beta}_i$ in (10), η in (15) and Ω^{-1} in (17) will be adjusted with $\omega_i^* = 0$ in (10) and $z_i = 0$ in (15) and (17). Therefore, the MCMC steps to the single prior assumption are given as above by using the adjusted conditional densities in (10), (15) and (17) and skipping steps 2, 3.1 and 3.2.

FORECAST

Having obtained the posterior distribution of the parameters, we can use it to predict future stock prices. We will illustrate this by the one-step forecast. Let $\mathbf{Y}_{t+1} = (Y_{1,t+1}, \ldots, Y_{n,t+1})$ denote the random future observations at period $t + 1$ for the n companies, and D_t and θ_t denote all the observed data and the parameters respectively from $t - T + 1$ to t (a window of data of length T periods). Our prediction for the $(t + 1)$th period follows from the predictive density

$$f(\mathbf{Y}_{t+1}|D_t) = \int f(\mathbf{Y}_{t+1}|D_t,\ \boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t|D_t)d\boldsymbol{\theta}_t \tag{19}$$

The predictive density in (19) can be approximated by the following Monte Carlo integration from the Gibbs sampler:

$$\hat{f}(\mathbf{Y}_{t+1}|D_t) = \frac{2}{RL}\sum_{r=1}^{R}\sum_{l=\lfloor L/2 \rfloor+1}^{L} f(\mathbf{Y}_{t+1}|D_t,\ \boldsymbol{\theta}_t^{(l,r)}) \tag{20}$$

where $\theta_t^{(l,r)}$ denotes the sample drawn in the lth iteration and the rth replication of the MCMC given the data set D_t and $\lfloor \cdot \rfloor$ denotes the floor operator, the largest integer less than or equal to the argument. Since the stock prices at period $t + 1$ are independent among the companies given $\boldsymbol{\theta}_t$ and D_t, we have

$$f(\mathbf{Y}_{t+1}|D_t,\ \boldsymbol{\theta}_t) = \prod_{i=1}^{n} f(Y_{i,t+1}|D_{i,t},\ \boldsymbol{\theta}_{i,t})$$

where $\theta_{i,t} = (\boldsymbol{\beta}_{i,t},\ \rho_{i,t},\ \sigma_{i,t}^2)$. In particular, the density function $f(Y_{i,t+1}|D_{i,t},\ \theta_{i,t})$ is just the normal density function of $Y_{i,t+1}$ with mean $\mathbf{x}_{i,t+1}'\boldsymbol{\beta}_i + \rho_i\,(y_{i,t} - \mathbf{x}_{i,t}'\boldsymbol{\beta}_i)$ and variance $\sigma_{i,t}^2$. Let $(\boldsymbol{\beta}_i^{(l,r)},\ \rho_{i,t}^{(l,r)},\ \sigma_{i,t}^{2(l,r)})$ denote the sampled parameters from the lth iteration and the rth replication of the MCMC, then we can rewrite (20) as

$$\hat{f}(\mathbf{Y}_{t+1}|D_t) = \frac{2}{RL}\sum_{r=1}^{R}\sum_{l=\lfloor L/2 \rfloor+1}^{L}\left\{\prod_{i=1}^{n}[\mathcal{N}(Y_{i,t+1}|\mu_{i,t}^{(l,r)},\ \sigma_{i,t}^{2(l,r)})]\right\} \tag{21}$$

where $\mu_{i,t}^{(l,r)} = \mathbf{x}_{i,t+1}'\boldsymbol{\beta}_i^{(l,r)} + \rho_i^{(l,r)}(y_{i,t} - \mathrm{x}_{i,t}'\boldsymbol{\beta}_i^{(l,r)})$. Hence, the sampled stock price for company i at time $t + 1$, denoted by $y_{i,t+1}^{(l,r)}$, can easily be drawn from the normal distribution function described above

given the observed data ($y_{i,t}$, $x_{i,t}$) and the sampled parameters ($\boldsymbol{\beta}^{(l,r)}$, $\rho_{i,t}^{(l,r)}$, $\sigma_{i,t}^{2(l,r)}$). The mean of $y_{i,t+1}^{(l,r)}$ over replication r and iteration L, after the first half ($L/2$) iterations being 'burned in', denoted by $\hat{y}_{i,t+1}$, can be obtained by

$$\hat{y}_{i,t+1} = \frac{2}{RL} \sum_{r=1}^{R} \sum_{l=\lfloor L/2 \rfloor + 1}^{L} y_{i,t+1}^{(l,r)} \tag{22}$$

The 95% predictive interval for $Y_{i,t+1}$ can be computed from the 2.5% and 97.5% empirical quantiles of the values $y_{i,t+1}^{(l,r)}$, $r = 1, \ldots, R$ and $l = \lfloor L/2 \rfloor + 1, \ldots, L$.

We then forecast Y_{t+2} by moving the data window up to D_{t+1} (data from periods $t - T + 2$ to $t + 1$). Similarly, we repeat this process until we obtain all the forecasts from $\hat{y}_{i,T+1}$ to $\hat{y}_{i,T'}$, where T' is the ending period of the data set.

MODEL VALIDATION AND MODEL CHOICE

Both model adequacy and model selection issues are discussed in this section. Model adequacy is checked by comparing the observed $y_{i,t+1}$ to its 95% predictive interval for $Y_{i,t+1}$ based on the results from the previous section. A model is judged to be adequate for each company if about 95% of the intervals for $t = T + 1, \ldots, T'$ contain the actual observed values for each company.

We use the coverage probability, the uncoverage probability and the average 95% predictive interval length as criteria for model selection. They are all computed as summary statistics based on all companies and all periods from $T + 1$ to T' . The coverage (uncoverage) probability is the probability that the actual values are covered (uncovered) by the 95% predictive intervals; and the average 95% predictive interval length is the average interval length of these intervals. We prefer a model with high coverage probability and low average length.

We also apply PMSE, PMAE, PMARD, PPBF for model selection. They are expressed as follows:

$$\begin{aligned}
\text{PMSE} &= \frac{1}{n(T'-T)} \sum_{i=1}^{n} \sum_{t=T}^{T'-1} (y_{i,t+1} - \hat{y}_{i,t+1})^2 \\
\text{PMAE} &= \frac{1}{n(T'-T)} \sum_{i=1}^{n} \sum_{t=T}^{T'-1} |y_{i,t+1} - \hat{y}_{i,t+1}| \\
\text{PMARD} &= \frac{1}{n(T'-T)} \sum_{i=1}^{n} \sum_{t=T}^{T'-1} \frac{|y_{i,t+1} - \hat{y}_{i,t+1}|}{y_{i,t+1}} \\
\text{PPBF} &= \prod_{t=T}^{T'-1} \hat{f}(\mathrm{Y}_{t+1} | D_t)
\end{aligned} \tag{23}$$

where $\hat{f}(\mathbf{Y}_{t+1}|D_t)$ and $\hat{y}_{i,t+1}$ are defined in (21) and (22), respectively. The best model is the one with the smallest PMSE PMAE, PMARD or the biggest PPBF. Note the PPBF is different from the pseudo Bayes factor where the cross-validation idea is used. In fact, the PPBF evaluates the conditional joint predictive density of the data $\mathbf{y}_{T+1}, \ldots, \mathbf{y}_{T'}$, given the window of length T of $\mathbf{y}_1, \ldots, \mathbf{y}_T$.

All the above comparisons are focused on the predictive performance of our forecast from the period $T + 1$ to T'. We can also single out a particular period for comparison. For example, we will focus on the 52nd period where the forecast differences between the HB and the classical method are more pronounced.

NUMERICAL EXAMPLES

We now apply the HB approach to the Ohlson model with a real data set. The preliminary data set contains 30 companies that make up the Dow Jones Index on July 1, 1998 and are recorded from the third quarter of 1984 to the first quarter of 1998. We obtained the data from two different sources, Value Line and the Federal Reserve Bank in D.C. The stock price, expected earnings per share and book value are all from Value Line. We use the 1-year Treasury bill as the interest rate, and it is provided by the Federal Reserve Bank in D.C.

Value Line records the stock price and the expected earnings per share for the next four quarters. We select the data from the ending month of each quarter in the analysis (in fact, other months in the quarter would also be appropriate). The interest rate is also collected on a monthly basis, and we select the interest rate from the last month of each quarter as the quarterly data.

Once we have obtained the quarterly data for expected earnings per share, interest rate and book value, we can calculate the expected excess earnings per share for the next four quarters as described in (3).

We also exclude two companies, Traveler Group and Goodyear Tire, from our data set since Value Line does not have complete expected earnings per share for these two companies during the period of our analysis. In summary, our data set consists of 28 companies and 54 quarters for each company.

As a general rule, we recentre and rescale the covariates to reduce the correlations among the covariates in the likelihood surface. Moreover, we use the autocorrelation function, partial autocorrelation function and inverse autocorrelation function to check the structure of the innovation terms. The results show that for most of the companies during the period of analysis, the innovation term has an AR(1) structure. The Durbin–Watson test for AR(1) also supports our conclusion. Thus, we use the AR(1) structure in our numerical analysis.

The initial study of the data using the classical method based on the Ohlson model shows that the estimates of σ^2 and ρ for each firm vary widely across firms. For example, the results from the data of the first 39 quarters show that the range of $\hat{\rho}$ among firms is from 0.425 to 0.977, and the range of $\hat{\sigma}^2$ is from 0.954 to 45.92. This suggests that we need to incorporate heterogeneous variances and separate correlation coefficients for each company.

The estimated regression coefficients for the expected excess earnings per share in four quarters show some outliers among the 28 companies. A single prior on $\boldsymbol{\beta}_i$'s will cause the outliers to be shrunk too much. We consider a mixture of two normal distributions for the prior of $\boldsymbol{\beta}_i$'s, which allows individual companies that share the same mean to borrow strength from each other but also to keep their individual properties.

The classical method uses least square estimates to make inferences on the parameters in an AR(1) linear model. Most of the statistical software packages, for example PROC AUTOREG in SAS, provide estimates with the classical method. These estimates provide useful information for the prior choices for the parameters and hyperparameters in the HB method. In fact, we let $\mathbf{m} = \mathbf{h} = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\beta}}_i$

and $\mathbf{M} = \mathbf{H} = \frac{1}{n}\sum_{i=1}^{n} \hat{\Sigma}_i$, where $\hat{\boldsymbol{\beta}}_i$ is the vector of the classical estimates of the regression coefficients and $\hat{\Sigma}_i$ is the classical estimate of the variance–covariance matrix of the regression coefficients of the *i*th company.

We can also obtain $\hat{\sigma}_i^2$ from the classical method. If we let the prior variance of σ_i^2 be 100, large enough that the prior will not drive the conclusion, then we can derive $a_i = 2 + \frac{(\hat{\sigma}_i^2)^2}{100}$ and $b_i = \hat{\sigma}_i^2 \times (a_i - 1)$ from the mean and variance expression of the inverse gamma density. The mean of the prior of ρ_i is set to the estimate of the correlation coefficient $\hat{\rho}_i$; that is, $\rho_{i0} = \hat{\rho}_i$; and we let $\sigma_{\rho i}^2 = 10$ for all $i = 1, \ldots, n$.

For the known parameters in the prior of the Wishart distribution, we let $v = q = K = 6$ and $\mathbf{V} = \mathbf{Q} = \mathbf{I}_K$. Finally, we choose $r = s = 1$, so that the prior of ω_i is uniformly distributed between 0 and 1.

We also consider alternative priors for the prior sensitivity analyses. In particular, we assume $\mathbf{m} = \mathbf{h} = 0$, $\mathbf{M} = \mathbf{H} = 100\mathbf{I}_K$; $a_i = 2.01$ and $b_i = 20$ (to make the variance of σ_i^2 in the prior as large

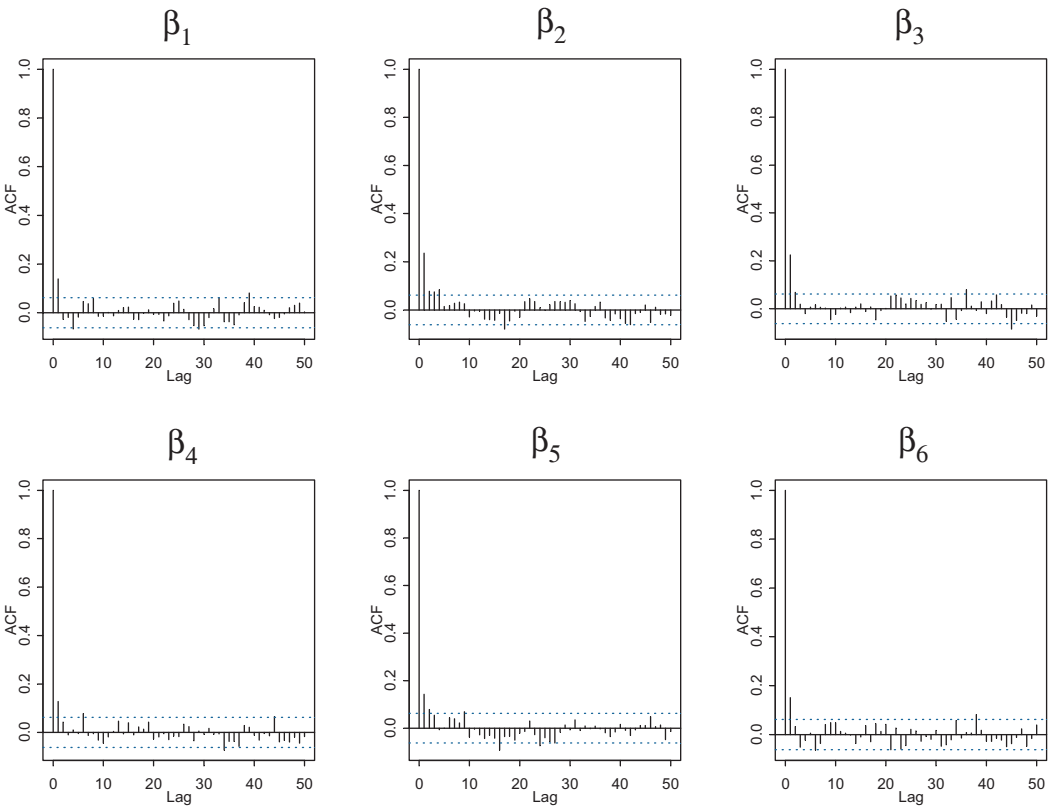

Figure 2. Autocorrelation plots. The autocorrelation plots of the Gibbs sampler of regression coefficients for Alcoa Inc. from the HB mixture prior method are given

as 39,212); $r = s = 1.5$ (to make the Beta prior of ω_i have more weight towards the centre than the uniform prior); and $\mathbf{V} = \mathbf{Q} = 100\mathbf{I}_K$ (to make the prior influence of Σ_i^{-1} and Ω^{-1} even weaker). The results (not presented here) show that our analysis is quite insensitive to the different prior choices.

All the parameter estimates are computed from the MCMC with 5000 iterations and two replications. Convergence of the Gibbs sampler is also assessed by the diagnostic procedures from CODA (Best *et al*., 1995) and Cowles and Carlin (1996).

The analyses based on the data from the first 39 quarters (that is from the third quarter of 1984 to the second quarter of 1993) are displayed in Figures 2 to 7.

Figure 2 shows the autocorrelation plots of the Gibbs sampler of regression coefficients for Alcoa Inc. from the HB mixture prior method. The autocorrelation drops to 0 quickly, suggesting an efficient Gibbs sampler. The same pattern exists for other companies.

We compare the HB estimates of the regression coefficients using the mixture (single) prior to the classical estimates in Figure 3 (Figure 4). We see the regression coefficients of the HB method in both figures shrink to each other as compared to those in the classical method by 'borrowing strength' from each other. However, the HB method using the single prior tends to over-shrink the regression coefficients, as shown in Figure 4.

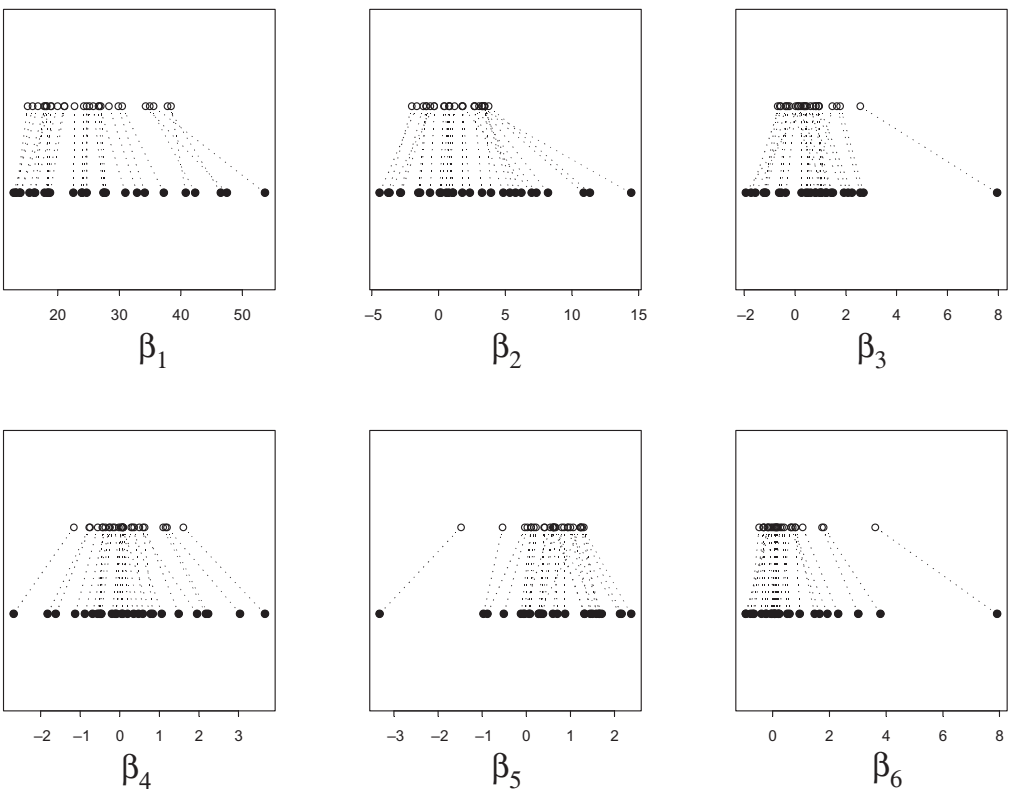

Figure 3. Shrinkage effect. The comparison between the HB method with mixture prior and the classical method for the point estimate of the regression coefficients for all the companies is given. The solid circles are from the classical method, the unfilled circles are from the HB method with mixture prior

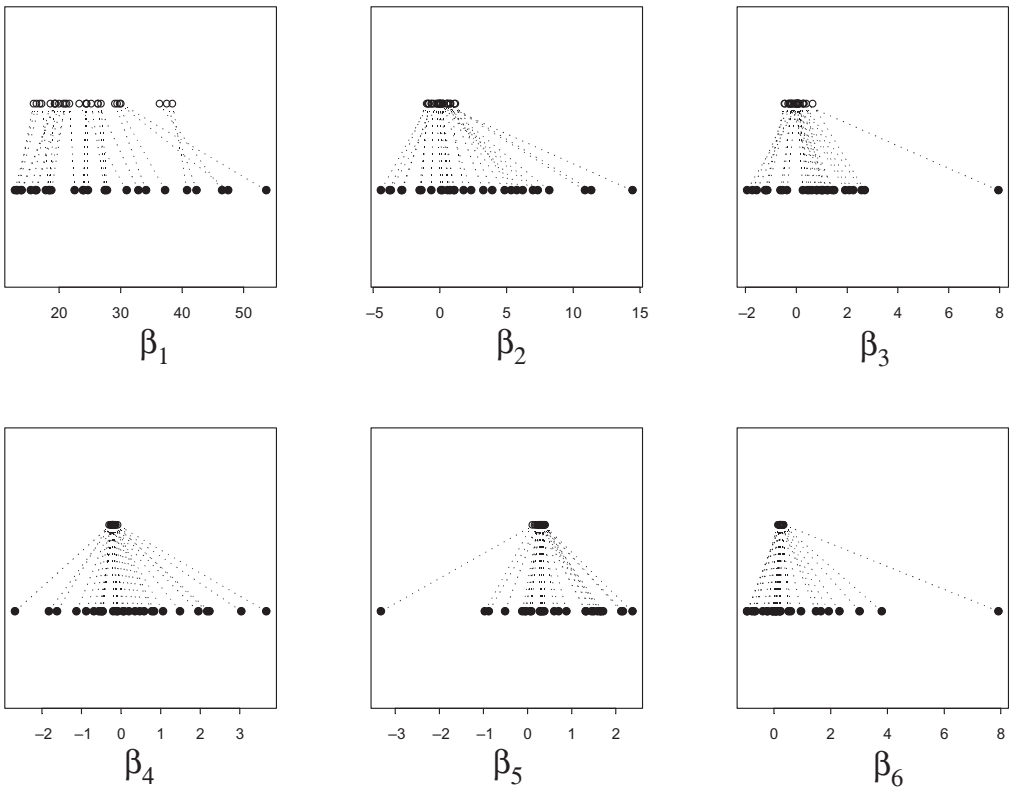

Figure 4. Shrinkage effect. The comparison between the HB method with single prior and the classical method for the point estimate of the regression coefficients for all the companies is given. The solid circles are from the classical method, the unfilled circles are from the HB method with single prior

We identify the outlying estimate of β_3 (corresponding to the expected 1st quarter excess earnings per share of stock) to be Coca Cola Co's. (KO) coefficient, and the outlying estimate of β_6 (corresponding to the expected 4th quarter excess earnings per share of stock) to be Procter & Gamble's (PG). During the first 39 quarters, the changes of stock prices from the lowest to the highest were \$56.97 for KO and \$46.03 for PG, much higher than the average of the 28 companies (\$27.66). At the same time, the changes of the expected 1st quarter excess earnings per share for KO (\$3.22) and the expected 4th quarter excess earnings per share for PG (\$3.34) were very close to the average of the 28 companies (\$3.23 and \$3.16, respectively). Therefore, the regression coefficient estimates (β_3 in KO and β_6 in PG) turn out to be much larger than others when the classical method is applied. Moreover, Figure 5 shows that the regression coefficients of these two companies are influenced by the outlying points with higher stock prices and relatively lower excess earnings per share, as we regress the stock prices on the corresponding covariate for each company. The slope estimates become smaller as the outliers are excluded from the estimation. The shrinkage effect from the HB method will bring similar effects by putting less weights on the outlying companies as expected. Similarly, the effect of the outlying company of ChevronTexaco Corp. (CVX) for β_5 can be smoothed out by the HB method.

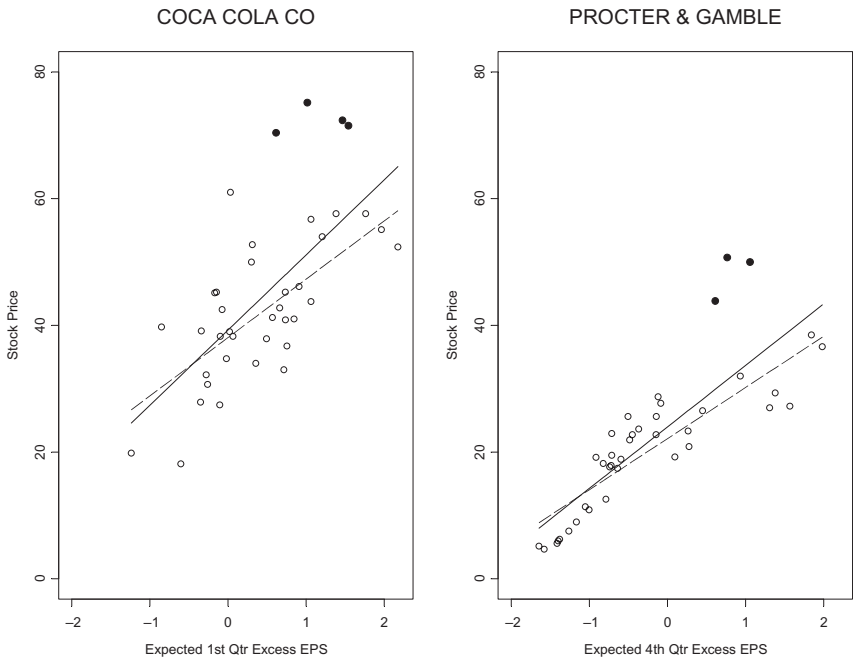

Figure 5. Outlying points. The solid lines are the simple regression lines when all the points are included in estimation. The dashed lines are the simple regression lines when the outlying points (solid points) are excluded from estimation

The 95% credible intervals of the regression coefficients with the HB method using the mixture prior as compared to that with the classical method for each company are shown in Figure 6. We conclude: (a) the HB method brings tighter credible intervals than the classical method; (b) the credible intervals in the HB method fluctuate less than those of the classical intervals across the companies. This is another indication of the effect of borrowing strength from each other in the HB method. Similarly, Figure 7 compares the HB estimates based on the single prior to that of the classical results. We see the credible intervals tend to be shrunk towards the same interval, not adaptive to the local variation of the regression coefficients.

One of the goals of our analysis is to compare different approaches to forecast the future stock prices. We produce the one-step-ahead prediction rule based on the estimates obtained respectively from HB and classical methods. In particular, we start with the first 39 quarters and predict the stock price in the 40th quarter. Then we move to the next window, the second 39 quarters (from the 2nd quarter to the 40th quarter), estimate the parameters and use them to predict the 41st quarter. As we move the windows of 39 quarters up by each quarter, we obtain 15 quarters of predicted stock prices for each company. Figure 8 (Figure 9) provides the 95% predictive intervals from the HB method using the mixture (single) prior and the classical method for six randomly picked companies. It can be seen that the predictive intervals made by the HB method are tighter than those from the classical method. Moreover, we notice that in some quarters, the classical method fails to forecast the real stock price with its 95% credible interval, while the HB method is able to contain the real future

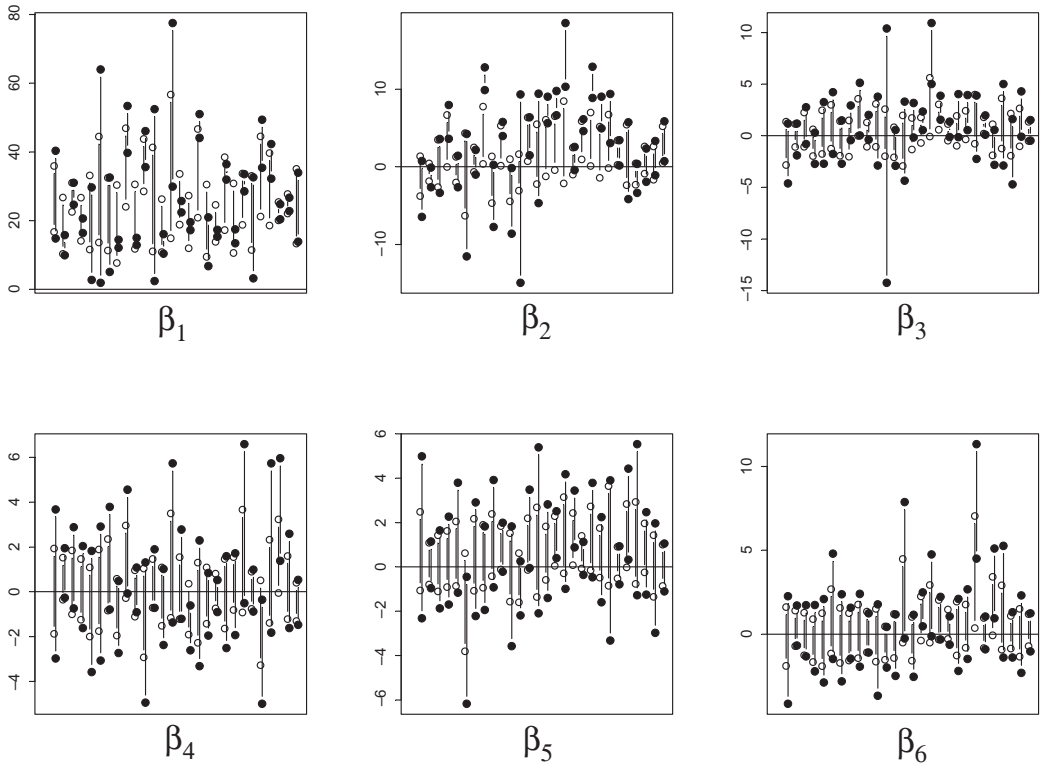

Figure 6. Credible intervals. 95% credible intervals for the regression coefficients from the HB method with the mixture prior and the classical method for all the companies are given. The solid circles are from the classical method and the unfilled circles are from the HB method with the mixture prior

observations in its intervals. This indicates that the HB method is more accurate in prediction than the classical method in these situations.

Table I exhibits the predictive performance among the classical method, the HB method with the single prior and the HB method with the mixture prior for the time period of the 40th quarter to the 54th quarter. Table II performs the same comparisons except only for the 52nd quarter. Both tables show HB methods are more accurate than the classical method in prediction. From Table I, among a total of 420 predictions, the HB method using the mixture prior contains all the real stock prices with its 95% predictive intervals, the HB method using the single prior misses five of the real stock prices from its 95% predictive intervals, and the classical method fails to contain 20 of the real stock prices. From Table II, the classical intervals fail to contain three out of 28 companies and none for any of the HB methods. Moreover, the average lengths of the classical intervals from both tables are longer than those of the HB mixture method. The latter intervals are also longer than those of the HB single method.

We use PMSE, PMAE and PMARD for model selection. Both tables show the HB mixture prior method is the best among three methods. The comparison of log(PPBF) between HB mixture and single prior methods also suggests that the former is the preferred method.

Table I. Model comparison

	Classical	HB single	HB mixture
Uncov.	0.048 (20/420)	0.012 (5/420)	0 (0/420)
Ave. Lgt.	13.97	11.36	12.17
PMSE	39.20	30.84	28.92
PMAE	4.56	4.15	4.01
PMARD (%)	10.28	9.27	8.83
log(PPBF)	—	–2.23	–2.17

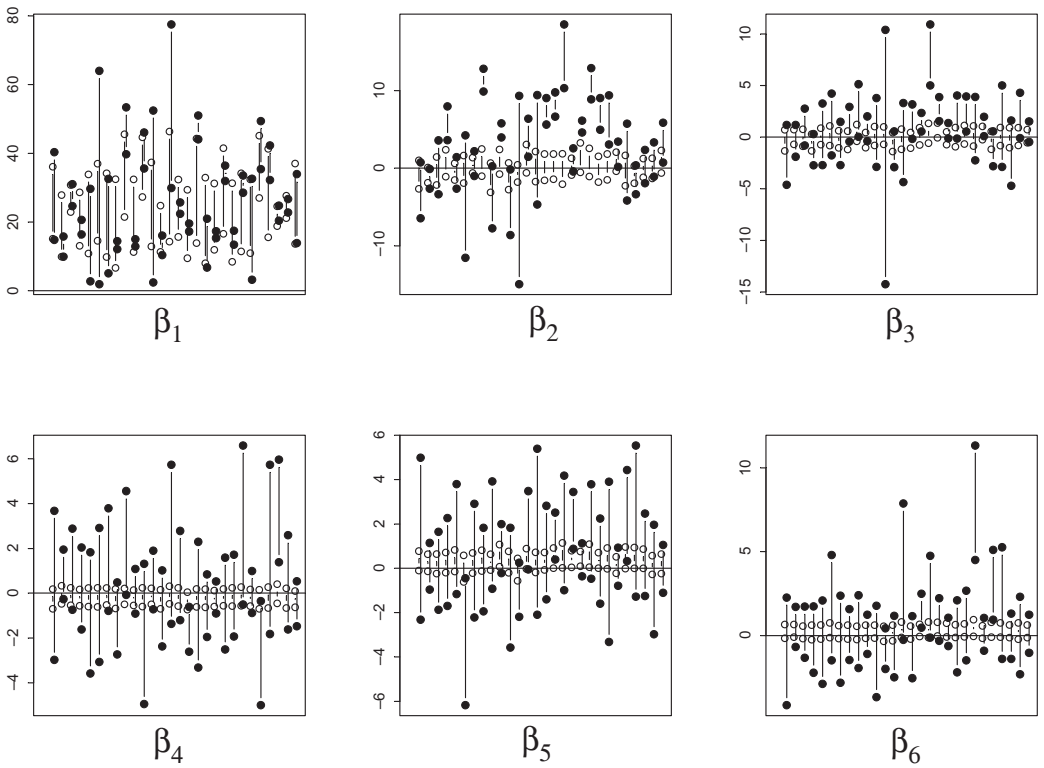

Figure 7. Credible intervals. 95% credible intervals for the regression coefficients from the HB method with the single prior and the classical method for all the companies are given. The solid circles are from the classical method and the unfilled circles are from the HB method with the single prior

CONCLUSIONS

The HB method described in this article has provided superior estimates to the classical method when applied to the Ohlson model. The improvement was obtained by pooling information across companies and borrowing strength from each other. The mixture of two normal distributions on the prior for the regression coefficients helps us to avoid the undesirable overshrinkage problem.

Table II. Model comparison for the 3rd quarter of 1997

	Classical	HB single	HB mixture
Uncov.	0.107 (3/28)	0 (0/28)	0 (0/28)
Ave. Lgt.	15.64	12.63	14.63
PMSE	82.15	45.26	38.6
PMAE	7.05	5.36	4.98
PMARD (%)	12.83	9.37	8.72
log(PPBF)	—	–2.39	–2.24

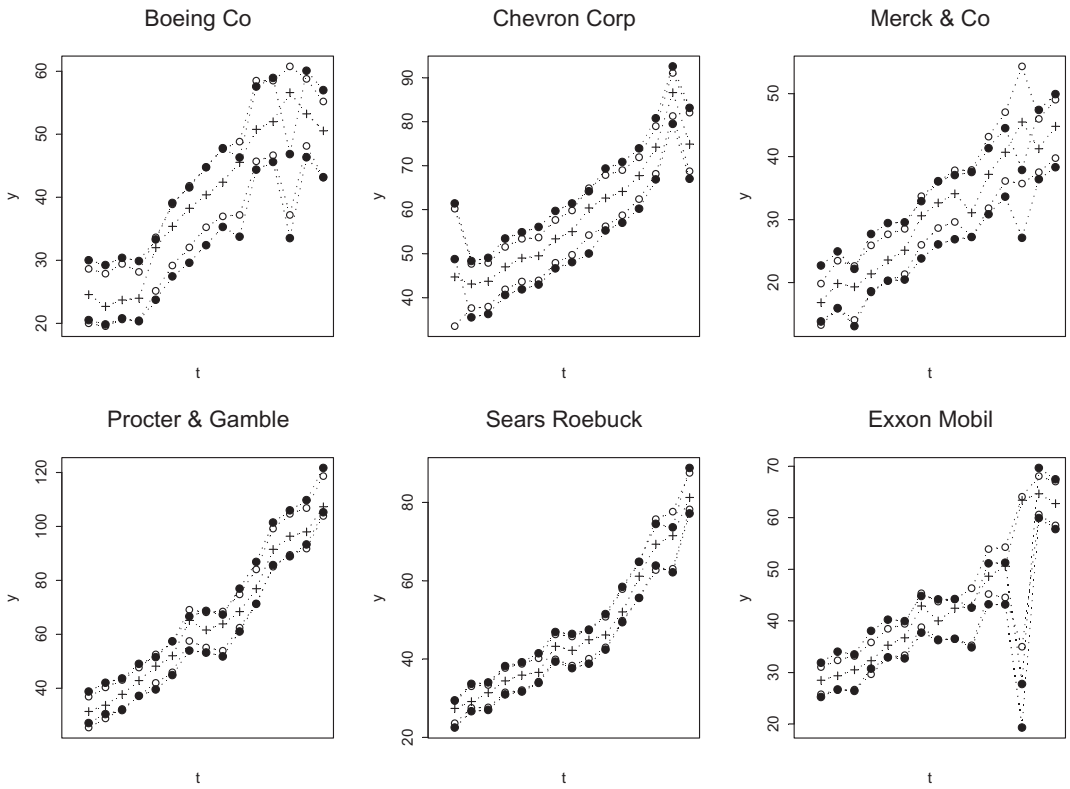

Figure 8. Forecasting. 95% predictive intervals based upon one-step-ahead forecasting for six companies in 15 quarters from the third quarter 1994 to the first quarter 1998 are plotted. The solid circles are from the classical method, the unfilled circles are from the HB method with the mixture prior, and the cross signs indicate the real stock prices

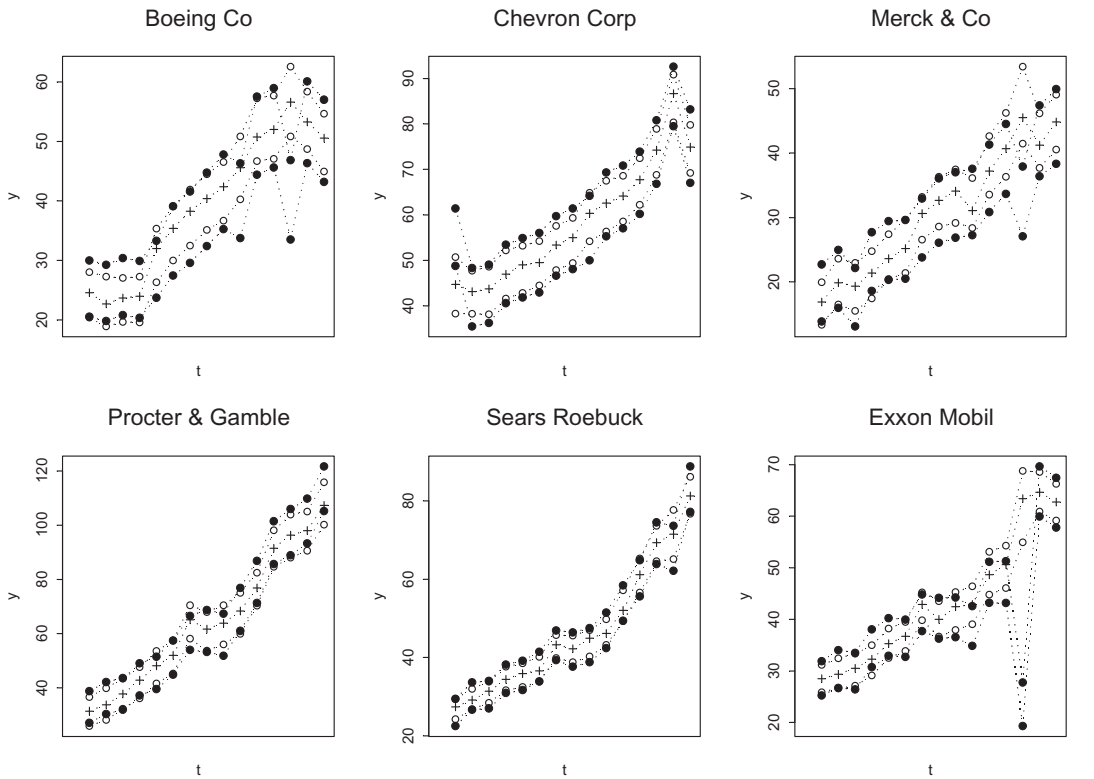

Figure 9. Forecasting. 95% predictive intervals based upon one-step-ahead forecasting for six companies in 15 quarters from the third quarter 1994 to the first quarter 1998 are plotted. The solid circles are from the classical method, the unfilled circles are from the HB method with the single prior, and the cross signs indicate the real stock prices

Consequently, the HB method yields improved prediction as compared to the classical method; that is the 95% predictive intervals from the HB method are tighter on average while at the same time containing more real stock prices than the classical method. On the other hand, the HB method with the single prior tends to over-shrink the regression coefficients and causes predictive credible intervals too tight to contain the real stock prices. Hence, it loses the predictive power to some extent as compared to the HB method with the mixture prior. The HB method with the mixture prior seems to play a compromise between the classical and the HB single prior methods. We are hopeful that this HB methodology with the mixture distribution structure on the regression coefficients will provide an applicable example for financial researchers to pool information from many firms to obtain more useful estimates.

REFERENCES

Albert J, Chib S. 1993. Bayesian analysis via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* **11**: 1–15.

Bernard V. 1995. The Feltham–Ohlson framework: implications for empiricists. *Contemporary Accounting Research* **11**: 733–747.
Best NG, Cowles MK, Vines SK. 1995. *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*, Version 0.3. MRC Biostatistics Unit: Cambridge.
Chen M-H, Shao Q-M, Ibrahim JG. 2000. *Monte Carlo Methods in Bayesian Computation*. Springer: New York.
Chib S. 1993. Bayes regression with autoregressive errors. *Journal of Econometrics* **58**: 275–294.
Chib S, Greenberg E. 1995. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics* **68**: 339–360.
Cowles MK, Carlin BP. 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* **91**: 883–904.
Dey D, Kuo L, Sahu S. 1995. A Bayesian predictive approach to determining the number of components in a mixture distribution. *Statistics and Computing* **5**: 297–305.
Feltham G, Ohlson J. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* **11**: 689–731.
Gelfand AE, Hills S, Racine-Poon A, Smith AFM. 1990. Illustration of 26 Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**: 972–985.
Gelfand AE, Sfiridis J. 1996. Bayesian analysis of financial event study data. *Advances in Econometrics* **11**: 25–62.
Gelfand AE, Smith AFM. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.
Gilks R, Wild P. 1992. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**: 337–348.
Kasim R, Raudenbush S. 1998. Application of Gibbs sampling to nested variance components models with heterogeneous within-group variance. *Journal of Educational and Behavioral Statistics* **23**: 93–116.
Lang M, Lundholm R. 1996. The relation between security returns, firm earnings, and industry earnings. *Contemporary Accounting Research* **13**: 607–629.
Lindley D, Smith AFM. 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B* **34**: 1–41.
Müller P, Rosner G. 1997. Semiparametric Bayesian population model with hierarchical mixture priors. *Journal of the American Statistical Association* **92**: 1279–1292.
Ohlson J. 1991. The theory of value and earnings, and an introduction to the Bell–Brown analysis. *Contemporary Accounting Research* **8**: 1–19.
Ohlson J. 1995. Earning, book value and dividends in security valuation. *Contemporary Accounting Research* **11**: 661–687.
Rosenberg M, Andrews R, Lenk P. 1999. A hierarchical Bayesian model for predicting the rate of non-acceptable in-patient hospital utilization. *Journal of Business and Economic Statistics* **17**: 1–8.
Stein C. 1955. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press: Berkeley, CA.
Wakefield J, Smith AFM, Racine-Poon A, Gelfand AE. 1994. Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistician* **43**: 201–221.
Ying J, Kuo L, Seow G. 2001. Forecasting stock prices using a hierarchical Bayesian approach. Technical Report TR01-11, University of Connecticut, Department of Statistics.

Authors' biographies:

Jun Ying is an assistant professor in the Department of Medicine, Division of Biostatistics at Indiana University School of Medicine. His current research interests include Bayesian computation, Bayesian and profile likelihood change point models, and stochastic volatility models.

Lynn Kuo is a professor in the Department of Statistics at the University of Connecticut. She received her PhD from the University of California at Los Angeles. Her current research interests include Bayesian computation, discrete choice models, event history analysis, software reliability, stochastic volatility models and bioinformatics.

Gim Seow is an associate professor in the Department of Accounting, School of Business, University of Connecticut. He received his PhD from the University of Oregon. His research interests include accounting standard-

setting and securities regulation, accounting and valuation of financial derivatives, and the effects of corporate risk management policies on firm value.

Authors' addresses:
Jun Ying, Department of Medicine, Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202-5181, USA.

Lynn Kuo, Department of Statistics, University of Connecticut, Storrs, CT 06269-4120, USA.

Gim S. Seow, Department of Accounting, University of Connecticut, Storrs, CT 06269-2041, USA.

