Bayesian Diagnostic Checking of the Capital Asset Pricing Model

Jun Li[1], Shaun S. Wulff[2]

[1]Dun & Bradstreet, Short Hills, NJ, USA
[2]Department of Statistics, University of Wyoming, Laramie, WY, USA
Email: wulff@uwyo.edu

Abstract

The capital asset pricing model (CAPM) is a commonly used regression model in finance to model stock returns. Bayesian methods have been developed for the CAPM to account for market fluctuations within the industry. However, a Bayesian model checking procedure is needed to assess the CAPM in terms of the usual regression model assumptions of independence, homogeneity of variance, and normality. This paper develops Bayesian residuals and Bayesian p-values to check these model assumptions as well as to suggest model extensions to the CAPM.

Keywords

Finance Model, Model Expansion, Linear Regression, Normality, Outlier, Residual

1. Introduction

Asset pricing models are used to model the excess return of individual stocks which is defined as the difference between the stock return and that for the whole market. Many pricing models have been developed in the finance literature. One of the most popular models is the multifactor Capital Asset Pricing Model (CAPM) [1]. This model is based upon a linear regression of the excess return with three explanatory variables representing the market-wide factors: the market premium, the return of a portfolio of small stocks in excess of the return on a portfolio of large stocks, and the return of a portfolio of stocks with high ratios of book value to market value in excess of the return of a portfolio of stocks with low book-to-market ratios. Bayesian methods have been applied to the CAPM to incorporate both information concerning market fluctuations within the industry and the uncertainty of the decision maker about the accuracy of the

model [2]. Model accuracy is reflected in the regression coefficient for the intercept which represents the mispricing in the CAPM. The decision maker can incorporate uncertainty about model accuracy through the prior variance of this intercept term.

However, Bayesian inference is model dependent as it is also based upon the set of specific assumptions associated with the CAPM. A violation of the underlying assumptions can have special implications in financial applications. For example, if stock returns have serial correlation across time, the serial dependence is a violation of the random walk hypothesis [3]. Likewise, if the returns do not have constant variance, the change in pattern may require a conditional heteroscedastic model to model the stock volatility [4]. Returns also may not follow a normal distribution. Empirical studies have shown that the returns tend to be skewed to the right and there is a need to also model those "rare" events.

Thus, it is important to be able to check and evaluate regression models, such as the CAPM, within the Bayesian context. A Bayesian model checking technique based upon the posterior predictive distribution was used to show that the CAPM could be inconsistent with the long-horizon returns of initial public offerings [5]. Specifically, various percentiles computed from both the observed returns and replicated returns simulated from the posterior predictive distribution were compared, and substantial differences between the two sets of percentiles were used as evidence to question the validity of the chosen model for the long-horizon return data. However, knowing the validity of a proposed model is not the end of the analysis. Additional diagnostic checks can provide strategies for model expansion or modification if the current model is deemed inadequate. Thus, the purpose of this research is to develop a Bayesian diagnostic methodology suitable for the CAPM that can reveal violations of the model assumptions. This methodology consists of residual diagnostics and tail area probability calculations to quantity the violation. The diagnostic methods are performed on stock return data to illustrate how to assess the CAPM assumptions and how to identify suitable adjustments to the CAPM that accommodate violations of the assumptions. The proposed techniques would also be useful for assessment of the model assumptions for other regression models or even for the general linear model.

This paper is organized as follows. Section 2 introduces the CAPM and reviews the Bayesian methods that have been used to perform model fitting and posterior inference. Section 3 develops Bayesian diagnostic approaches and discusses computational strategies. Section 4 applies the methodology to modeling a series of monthly returns of a stock from a pharmaceutical company. Section 5 provides concluding remarks.

2. Bayesian Analysis of the CAPM

2.1. The CAPM

Financial studies typically use returns, instead of prices, of assets for two reasons.

First, for average investors, return of an asset is a complete and scale-free summary of the investment opportunity. Second, return series are easier to handle, and have more attractive statistical properties [6]. The monthly return at time t is calculated using the definition of one-period simple return, which is the percentage of change between the close prices of two neighboring months of t and $t-1$ [4]. That is,

$$r_t = 100 \times \left(\left(p_t - p_{t-1} \right) \big/ p_{t-1} \right) \tag{1}$$

where r and p denote the return and price of a stock.

The CAPM quantifies the insight that riskier assets should offer higher expected returns to the investors [1]. The model is given by

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta} + \varepsilon_t, \ t = 1, \cdots, T \ . \tag{2}$$

In Equation (2), y_t denotes the excess return of the stock, which is the difference between the stock return, as calculated in Equation (1), and that of the market portfolio. The vector of predictors is $\boldsymbol{x}_t' = \begin{bmatrix} 1 & r_{m,t} - r_{f,t} & \text{SMB}_t & \text{HML}_t \end{bmatrix}$ which is a collection of three market-wide macroeconomic risk factors defined as follows. The term $r_m - r_f$ stands for the difference between the market return and the risk-free return, and is also called the market premium. The predictor SMB is called "small minus big", which is the return of a portfolio of small stocks in excess of the return on a portfolio of large stocks. The predictor HML denotes "high minus low", and is the return of a portfolio of stocks with high ratios of book value to market value in excess of the return of a portfolio of stocks with low book-to-market ratios. The subscript t indexes the values at time t, and T is the total number of observations. The regression coefficient vector, denoted by $\boldsymbol{\beta}$, represents the impact of market premium, SMB, and HML on stock performance.

A matrix form of the CAPM is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \ \boldsymbol{\varepsilon} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_T \right), \tag{3}$$

where $\boldsymbol{y}$ is a $T \times 1$ vector of returns, $\boldsymbol{X}$ is a $T \times 4$ matrix containing all factor information, and $\boldsymbol{\varepsilon}$ is the vector of regression errors which is assumed to be independent and normally distributed with common variance σ^2.

2.2. A Bayesian Approach

A Bayesian approach can be used to incorporate market information through prior distributions [2]. These priors are based upon a normal-inverted-gamma distribution that is typical for regression parameters:

$$\boldsymbol{\beta} \mid \sigma^2 \sim N\left(\bar{\boldsymbol{\beta}}, \boldsymbol{V}_0(\sigma) \right), \ \sigma^2 \sim \text{IG}\left(\frac{\upsilon_0}{2}, \frac{2}{\upsilon_0 s_0^2} \right). \tag{4}$$

The hyperparameters $\bar{\boldsymbol{\beta}}$, υ_0 and s_0 are determined by setting the expectation and variance of the prior distribution equal to the first and second moments of the regression parameter estimates from all stocks in the same industry. The priors are the same as those in the conjugate case, except for the formation of

covariance matrix $V_0(\sigma)$, which is

$$V_0(\sigma) = \begin{bmatrix} \left(\dfrac{\sigma^2}{E(\sigma^2)}\right)V_{11} & \left(\dfrac{\sigma}{E(\sigma)}\right)[V_{12}\ V_{13}\ V_{14}] \\[12pt] \left(\dfrac{\sigma}{E(\sigma)}\right)[V_{12}\ V_{13}\ V_{14}]' & V^* \end{bmatrix}. \qquad (5)$$

Let V be the 4×4 covariance matrix obtained by computing the sample covariance matrix of β from all stocks in the industry. Then V_{12}, V_{13}, V_{14} stand for the corresponding elements in the matrix V, while V^* is the 3×3 sub-matrix of V corresponding to the covariance matrix of β associated with the three factors. The terms $E(\sigma^2)$ and $E(\sigma)$ are the averages of all estimates of σ^2 and σ in the industry. Hence, σ^2 is the only unknown parameter in $V_0(\sigma)$. This special construction is applied to represent an empirical finding in finance that the variability of all terms associated with the intercept is typically in line with the magnitude of the variance of the error.

2.3. Bayesian Computation

The prior specification in Equations (4) and (5) is not the conjugate specification for the regression model ([7], Section 3.2). Under the conjugate prior, σ^2 can be factored out of the covariance matrix, unlike in Equation (5) where only those terms related to the intercept are associated with σ^2. Thus, a brief discussion is provided of the Markov chain Monte Carlo (MCMC) technique used to obtain the posterior estimates of β and σ^2.

Two of the typical MCMC procedures, the Metropolis-Hastings (M-H) algorithm [8] and the Gibbs sampler [9] are used in this study. The purpose of the M-H algorithm is to generate a sequence of samples from a distribution $f(\cdot)$ that is difficult to directly sample. An alternative is to take candidate draws from a proposal density $q(\cdot)$ that is easy to sample, and these draws are then accepted with a suitable acceptance probability in such a way that the result is a Markov chain converging to the target distribution $f(\cdot)$. Let $\theta^{(j)}$ denote the jth element of the parameter vector. The Markov chain $\{\theta^{(j)}\}$ is formed as Algorithm 1.

Algorithm 1:

1) Given $\theta^{(j-1)}$, draw $\theta^{(j)}$ from a proposal distribution $q\left(\theta^{(j)} \mid \theta^{(j-1)}\right)$.

2) Accept the candidate $\theta^{(j)}$ with probability

$$\alpha\left(\theta^{(j-1)}, \theta^{(j)}\right) = \min\left\{1, \frac{f\left(\theta^{(j)}\right)q\left(\theta^{(j-1)} \mid \theta^{(j)}\right)}{f\left(\theta^{(j-1)}\right)q\left(\theta^{(j)} \mid \theta^{(j-1)}\right)}\right\}.$$

Otherwise, reject and let $\theta^{(j)} = \theta^{(j-1)}$.

Gibbs sampling is a special case of the M-H algorithm. It is applicable when the full conditional distributions are known for each model parameter. The Markov chain is updated one component at a time by drawing from the full conditional distributions given the values of the remaining components. Con-

sider the case where the parameter has p components. The updating is achieved as stated in Algorithm 2.

Algorithm 2:

Given $\boldsymbol{\theta}^{(j-1)} = \left(\theta_1^{(j-1)}, \cdots, \theta_p^{(j-1)} \right)$, obtain $\theta_1^{(j)}$ from $f_1 \left(\theta_1 \mid \theta_2^{(j-1)}, \cdots, \theta_p^{(j-1)} \right)$, obtain $\theta_2^{(j)}$ from $f_2 \left(\theta_2 \mid \theta_1^{(j)}, \theta_3^{(j-1)}, \cdots, \theta_p^{(j-1)} \right)$, $\cdots$, obtain $\theta_p^{(j)}$ from $f_p \left(\theta_p \mid \theta_1^{(j)}, \theta_2^{(j)}, \cdots, \theta_{p-1}^{(j)} \right)$.

If $\boldsymbol{\theta}$ denotes the vector of all model parameters under CAPM, then it equals $\begin{bmatrix} \boldsymbol{\beta} & \sigma^2 \end{bmatrix}'$. A hybrid algorithm is recommended to obtain the posterior draws [2]. Based on Equations (3) and (4), the joint posterior distribution can be derived as

$$
\begin{aligned}
f\left(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}\right) \propto \frac{1}{\sigma^{v_0+T+2}} \exp \Bigg\{ & -\frac{1}{2\sigma^2} \\
& \times \left(v_0 s_0^2 + T\hat{\sigma}^2 + \left(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}}\right)' \left(\frac{1}{\sigma^2} \boldsymbol{V}_0(\sigma)\right)^{-1} \left(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}}\right) + \left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)' \boldsymbol{X}'\boldsymbol{X} \left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right) \right) \Bigg\}
\end{aligned}
\tag{6}
$$

where $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{y}$ and $\hat{\sigma}^2 = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)' \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)$. Note that Equation (6) can also be treated as an expression for the conditional posterior distribution $f\left(\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2\right)$ or $f\left(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{\beta}\right)$ when σ^2 or $\boldsymbol{\beta}$ is known. Therefore, σ^2 is updated by the M-H procedure in Algorithm 1. The proposal distribution $q(\cdot)$ is set to be the conditional posterior distribution for σ^2 that arises when $\boldsymbol{\beta}$ and σ^2 are made independent in the normal-inverted-gamma prior. For a given $\boldsymbol{\beta}$, the target $f(\cdot)$ is the conditional distribution $f\left(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{\beta}\right)$, which is proportional to the right-hand side of Equation (6). Given σ^2 from the last step, $\boldsymbol{\beta}$ is updated using the conditional distribution $f\left(\boldsymbol{\beta} \mid \boldsymbol{y}, \sigma^2\right)$ via Gibbs sampling in Algorithm 2. This approach will be called the "hybrid" algorithm since the parameters are updated using the different MCMC procedures described in Algorithm 1 and Algorithm 2.

3. Bayesian Model Diagnostic Checking

In the frequentist setting, model diagnostic checking is an integral part of data analysis. Bayesian methods, however, have been criticized as being strong for inference under an assumed model, but weak for the development and assessment of models [10]. The purpose of model diagnostics is to assess the assumptions of a posited model and to identify troublesome features of the model. Bayesian analysis typically conditions on the whole probability model making it crucial to check whether or not the posited model fails to provide a reasonable summary of the data.

3.1. Bayesian Residuals

Many of the diagnostics in the frequentist setting utilize residuals to check model assumptions. Residuals are generally represented as the observed value of the response minus the fitted value of the response. From the Bayesian perspective, there is not an agreed upon candidate for a fitted value. As a result, there are many possible Bayesian residuals [11]. A few Bayesian residuals are defined below.

For Bayesians, the predicted value has a distribution called the posterior predictive distribution [12]. Let y_{rep} denote the value of the response that could have been observed under the values of the parameter θ and with the collection of explanatory variables X. The subscript arises from the fact that it is the data that "could appear if the experiment that produced y today were replicated tomorrow with the same model" [13]. The posterior predictive distribution, $f_{\text{post}}\left(y_{\text{rep}} \mid y\right)$, is given by

$$f_{\text{post}}\left(y_{\text{rep}} \mid y\right) = \int f\left(y_{\text{rep}} \mid \theta\right) f\left(\theta \mid y\right) \mathrm{d}\theta, \tag{7}$$

where $f\left(\theta \mid y\right)$ is the density function of the posterior distribution and $f\left(y_{\text{rep}} \mid \theta\right)$ is the density function of the sampling distribution evaluated at y_{rep}. The Bayesian residual vector mentioned by [14] is defined as

$$e_{\text{Bayes}} = y - E\left(y_{\text{rep}} \mid y\right). \tag{8}$$

Equation (8) reflects the difference between the observed value and the predicted value, but here the predicted value corresponds to the expectation of the posterior predictive distribution in Equation (7). The residuals in Equation (8) can be standardized by dividing each value by the square root of the variance of the posterior predictive distribution for that corresponding observation.

Numerically, the residuals can be calculated using a value of y_{rep} that is generated from Equation (3) where values of β and σ^2 are taken from a draw of the M-H and/or Gibbs algorithms ([12], p. 76). These values are generated for all draws and the value of $E\left(y_{\text{rep}} \mid y\right)$ is taken to be the mean of all y_{rep}. The Bayesian residuals can then be obtained according to Equation (8).

Other residuals can be calculated using draws from the posterior distribution. Let $\beta^{(j)}$ denote the jth draw for β. The observed residuals can be computed as

$$e_{\text{obs}}^{(j)} = y - X\beta^{(j)}. \tag{9}$$

Alternatively, the realized residuals can be computed from the jth draw from the posterior predictive distribution, say $y_{\text{rep}}^{(j)}$. In this case,

$$e_{\text{rep}}^{(j)} = y_{\text{rep}}^{(j)} - X\beta^{(j)}. \tag{10}$$

The residuals in Equations (9) and (10) can be standardized by dividing by the square root of $\sigma^{2(j)}$ where $\sigma^{2(j)}$ denotes the jth draw for σ^2. The standardized values of from Equations (9) and (10) represent a form of observed and realized measures of discrepancy [13]. Averages of Equations (9) and (10) across the posterior draws produce a single set of residuals, denoted e_{obs} and e_{rep}, respectively. These averages approximate expectations of Equations (9) and (10) with respect to the posterior distribution. The values of e_{obs} correspond to those in Equation (8) since y_{rep} is sampled from a normal distribution with mean $X\beta^{(j)}$ at the jth draw. The values of $e_{\text{rep}}^{(j)}$ at the jth draw can also be obtained by sampling independently from a normal distribution with mean 0 and variance $\sigma^{2(j)}$.

3.2. Bayesian P-Values

Another standard frequentist approach for model checking is the computation

of a tail-area probability to quantify inconsistency between the data and the proposed model. A test statistic $D(y)$ based upon the classical residuals is chosen to investigate a certain discrepancy from the model assumption. Without loss of generality, assume large values of D indicates incompatibility with the model. The classical p-value is

$$p_{\text{classical}} = \Pr\left(D\left(y_{\text{rep}}\right) \geq D\left(y\right) | \boldsymbol{\theta}\right). \tag{11}$$

The unknown parameter $\boldsymbol{\theta}$ is handled by substituting some estimate $\hat{\boldsymbol{\theta}}$. A p-value close to zero indicates a lack of fit in the direction of $D(y)$ under the current model. The probability function, denoted by Pr in Equation (11), is calculated with respect to the sampling distribution of y that generates y_{rep}. However, the classical p-value ignores the uncertainty in the estimation of $\boldsymbol{\theta}$, and is not suitable in the Bayesian setting where $\boldsymbol{\theta}$ is not assumed to be fixed. Several Bayesian p-values have been developed and their difference lies in the reference distribution used for the tail area computation.

The posterior predictive p-value (ppp) accounts for the distribution of $\boldsymbol{\theta}$ using the posterior predictive distribution in Equation (7) [13] [15]. The ppp is defined as

$$p_{\text{post}} = \int \Pr\left(D\left(y_{\text{rep}}\right) \geq D\left(y\right) | \boldsymbol{\theta}\right) f\left(\boldsymbol{\theta} | y\right) \mathrm{d}\boldsymbol{\theta} . \tag{12}$$

It can be generalized by replacing $D(y)$ with a parameter-dependent test statistic $D(y, \boldsymbol{\theta})$ [13] [15]. Equation (12) shows that p_{post} is the expected value of the classical p-value in Equation (11) with respect to the posterior distribution. This p-value is calculated from the MCMC algorithm using the following steps:

Algorithm 3:

1) For given draw $\boldsymbol{\theta}^j$ from the posterior distribution, obtain y_{rep}^j from $f\left(y | \boldsymbol{\theta}^j\right)$.

2) Compute $D\left(y_{\text{rep}}^j, \boldsymbol{\theta}^j\right)$ and $D\left(y, \boldsymbol{\theta}^j\right)$. The former quantity is called the "realized discrepancy", and the latter is called the "observed discrepancy".

3) Repeat step 1 and step 2 a large number of times. The posterior predictive p-value is estimated by the proportion of times in which $D\left(y_{\text{rep}}^j, \boldsymbol{\theta}^j\right) > D\left(y, \boldsymbol{\theta}^j\right)$.

A criticism of the posterior predictive p-value is the double use of data since observations are first used to produce the posterior distribution, and then used again to compute the tail area probability. This results in a conservative p-value where the distribution of the ppp is more concentrated around ½ as has been demonstrated for both small samples [16] and large samples [17].

The partial posterior predictive p-value (partial-ppp) was created to combine the advantages of the prior and posterior predictive p-values [16] [18]. The partial-ppp is

$$p_{\text{partial}} = \int \Pr\left(D\left(y_{\text{rep}}, \boldsymbol{\theta}\right) \geq D\left(y, \boldsymbol{\theta}\right)\right) f^*\left(\boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{\theta}, \tag{13}$$

$$f^*\left(\boldsymbol{\theta}\right) \propto f\left(y | D, \boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}\right) \propto \frac{f\left(y | \boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}\right)}{f\left(D | \boldsymbol{\theta}\right)}. \tag{14}$$

Equation (13) is similar to Equation (12) except that the posterior predictive probability density function is replaced by $f^*(\boldsymbol{\theta})$. The reference distribution for the probability in Equation (14) is $f^*(D) = \int f(\boldsymbol{y}|\boldsymbol{\theta}) f^*(\boldsymbol{\theta}) d\boldsymbol{\theta}$. As indicated by the numerator of the right-hand side of Equation (14), the partial-ppp is still based upon the posterior distribution of $\boldsymbol{\theta}$. However, it avoids the double use of data by conditioning on the test statistic, so the contribution of D to the posterior is removed before $\boldsymbol{\theta}$ is eliminated by integration [18]. Intuitively, the act of conditioning on D has the effect of splitting the information contained in the data set into two parts where the first part computes D and the remaining information in the second part forms the distribution $\pi^*(\boldsymbol{\theta})$. A simulation study found that the partial-ppp is asymptotically uniform [17].

The partial-ppp can be calculated using Algorithm 3 with a modification of step 1. When drawing $\boldsymbol{\theta}^j$, it should be taken from $f^*(D)$ which is the posterior distribution of parameters conditioning on the test statistic D. Generating $f^*(D)$ can be done in two ways [16] [18]. The first method is to construct a Metropolis chain based upon draws from the full posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$. Using it as the proposal density, the simulation moves from a current draw $\boldsymbol{\theta}^{(j-1)}$ to a new draw $\boldsymbol{\theta}^{(j)}$ with probability corresponding to the minimum of $\left\{1, f\left(D(\boldsymbol{y})|\boldsymbol{\theta}^{(j-1)}\right) \big/ f\left(D(\boldsymbol{y})|\boldsymbol{\theta}^{(j)}\right)\right\}$. When $f(D(\boldsymbol{y})|\boldsymbol{\theta})$ is highly variable in $\boldsymbol{\theta}$, the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ may not be a good probing distribution for $f^*(\boldsymbol{\theta})$ [16]. In that case, the second method is preferred. This procedure is as follows:

Algorithm 4:

1) Given $\boldsymbol{\theta}^{(j-1)}$, draw $\boldsymbol{\theta}^{(j)}$ from the posterior distribution. Draw u from the Uniform(0,1) distribution. The candidate draw is $\boldsymbol{\theta}^{*(j)} = \boldsymbol{\theta}^{(j)} + u\left(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} - \hat{\boldsymbol{\theta}}_{\mathrm{cMLE}}\right)$.

2) Replace the current $\boldsymbol{\theta}^{*(j)} = \boldsymbol{\theta}^{(j)} + u\left(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} - \hat{\boldsymbol{\theta}}_{\mathrm{cMLE}}\right)$ with $\boldsymbol{\theta}^{*(j)}$ with probability

$$\min\left\{1, \frac{f\left(D(\boldsymbol{y})|\boldsymbol{\theta}^{*(j-1)}\right)}{f\left(D(\boldsymbol{y})|\boldsymbol{\theta}^{*(j)}\right)} \frac{\pi\left(\boldsymbol{\theta}^{*(j)}\right)}{\pi\left(\boldsymbol{\theta}^{*(j-1)}\right)} \frac{\pi\left(\boldsymbol{\theta}^{(j-1)}\right)}{\pi\left(\boldsymbol{\theta}^{(j)}\right)} \frac{f\left(\boldsymbol{y}|\boldsymbol{\theta}^{*(j)}\right)}{f\left(\boldsymbol{y}|\boldsymbol{\theta}^{*(j-1)}\right)} \frac{f\left(\boldsymbol{y}|\boldsymbol{\theta}^{(j-1)}\right)}{f\left(\boldsymbol{y}|\boldsymbol{\theta}^{(j)}\right)}\right\}.$$

The quantities $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ and $\hat{\boldsymbol{\theta}}_{\mathrm{cMLE}}$ are the maximum likelihood estimator (MLE) and the conditional MLE, respectively. The MLE is defined as $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \arg\max f(\boldsymbol{y}|\boldsymbol{\theta})$ and the conditional MLE is defined as $\hat{\boldsymbol{\theta}}_{\mathrm{cMLE}} = \arg\max f(\boldsymbol{y}|\boldsymbol{\theta})\big/f(\boldsymbol{t}|\boldsymbol{\theta})$ [16]. The modification in step 1 of Algorithm 4 moves the draws from the posterior distribution by adding a portion of the difference between the two estimates, $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} - \hat{\boldsymbol{\theta}}_{\mathrm{cMLE}}$. This quantity is meant to push simulations lying around MLEs, which are regions of low probability if the model is incompatible with the data, to regions with high probability according to the distribution of $\pi(\boldsymbol{\theta}^*)$ [19].

The evaluation of both acceptance probabilities demands that $f\left(D(\boldsymbol{y})|\boldsymbol{\theta}^{(j)}\right)$ is available in closed form at each draw. A shortcut to this approach is to choose

a test statistic that is a pivotal quantity where the distribution of $D(y, \theta)$ does not depend on the parameter θ. If $D(y, \theta)$ is a pivotal quantity, then $D(y, \theta = \theta_0)$ and $D(y, \theta = \theta^{(j)})$ are identically distributed where θ_0 is the true data-generating value of the parameter and $\theta^{(j)}$ denotes a parameter vector drawn from the posterior [20]. With this property, the trouble of evaluating $f(D(y)|\theta^{(j)})$ at each posterior draw is circumvented as $D(y, \theta^{(j)})$ has the same known distribution as $D(y, \theta)$ under the proposed model.

As for the choice of the discrepancy statistic D, any test statistic can be formed as the check function. The general guideline is that it "formalizes questions that any competent statistician would raise having been presented with the supposed form of the model and the data" [21]. Specific choices for the CAPM will be presented in Section 4.2. The discrepancy statistic calculated at the jth draw is a function computed using the observed residuals, $D(e_{obs}^{(j)}, \theta)$, and the realized residuals, $D(e_{rep}^{(j)}, \theta)$.

4. Example from the Pharmaceutical Industry

In this section, the CAPM is fit to the series of excess returns for stock Serono, S.A. (ticker symbol SRA) using the Bayesian approach described in Section 2.2. Diagnostic checks are performed to check for potential problems with the model assumptions. The series consist of 52 monthly observations between September 2000 and December 2004. The firm is a pharmaceutical company specializing in drug production and medical research. Thus, information from the pharmaceutical industry, represented by all the 30 pharmaceutical companies listed in the New York Stock Exchange, is used to form the prior distributions.

4.1. Fit of the CAPM

The hybrid algorithm described in Section 2.3 is used to generate simulations from the posterior distributions. After discarding the initial 500 draws for the burn-in, 10,000 draws are used for posterior inference. The posterior estimates (mean and standard deviation) of the model parameters are given in Table 1.

The Deviance Information Criteria (DIC) is commonly used for Bayesian model selection. The DIC is formed as the posterior mean of the deviance, which is -2 times the loglikelihood, along with a penalty for the effective number of parameters ([12], Section 6.7). A second measure is the expected squared error loss (ESSE) given by

$$E\left[\sum_{t=1}^{T} (y_t - y_{t,obs})^2 \right], \tag{15}$$

Table 1. Posterior estimates of the CAPM for the SRA data.

parameter	beta [1]	beta [2]	beta [3]	beta [4]	sigma 2
mean	0.1938	1.286	0.1823	0.5904	67.3
std deviation	0.6059	0.2388	0.3285	0.3096	13.23

where the expectation is respect to the posterior predictive distribution [22]. The quantity in Equation (15) corresponds to the posterior mean of the error sum of squares and does not include a penalty for model complexity. For the CAPM model fit result of this particular stock, the DIC is 366.6 and the ESSE is 3102.73.

4.2. Evaluation of Fit of the CAPM

Figure 1 shows a plot of standardized Bayesian residuals in Equation (8) against the posterior mean values for each of the 52 monthly time points. This figure does not suggest any obvious departures from the homogeneity of variances assumption. However, it does show a few residuals with extreme negative values. **Figure 2** shows a plot of the residuals versus time as well as a plot of the autocorrelation function (ACF). In particular, the ACF plot in **Figure 2** suggests the model errors may contain autocorrelation of order 1 since the ACF exceeds this threshold at lag 1. **Figure 3** contains a histogram and normal quantile-quantile (QQ) plot of the standardized residuals. Both plots also highlight the extreme negative residual values. These small residual values are the only ones that tend to substantially deviate from the requisite line.

The three p-values in Section 3.2 can be used to quantify the degree to which model fit deviates from the required model assumptions. The classical p-value is computed by plugging in the posterior mean of the parameters as an estimate of θ in Equation (11). The ppp is calculated using Algorithm 3. The partial-ppp is calculated based upon Algorithm 4 as all test statistics are expected to be highly variable across the posterior draws.

A number of pivotal quantities could be used as measures of discrepancy $D(e, \theta)$ from the model assumptions. To quantify violations of independent errors, the Portmanteau test on the residuals is used. The associated Ljung-Box test statistic follows a Chi-squared distribution ([23], p. 541). To assess possible violation of the homogeneity of variance assumption, the Breusch-Pagan La grange multiplier test statistic is applied to the residuals. The test statistic is known to follow a Chi-squared distribution ([23], p. 510). To assess the assumption of normality, the Wilk-Shapiro test is applied to the residuals. With a
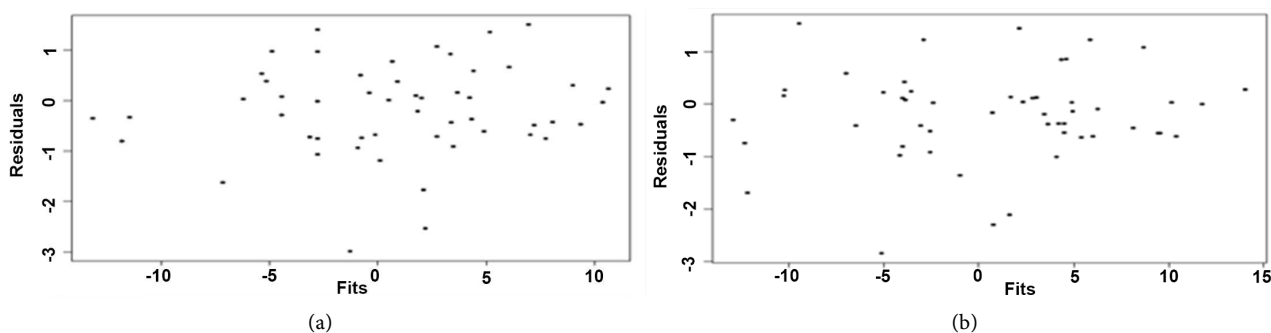

Figure 1. Plot of standardized Bayesian residuals versus fitted values for the CAPM with usual errors (a) and autocorrelated normal errors (b).
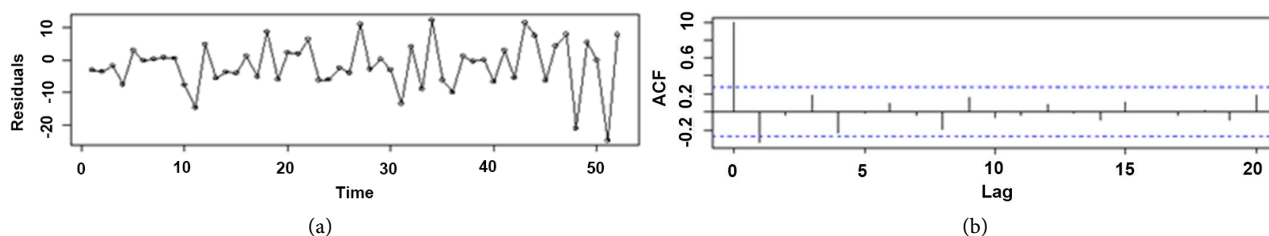
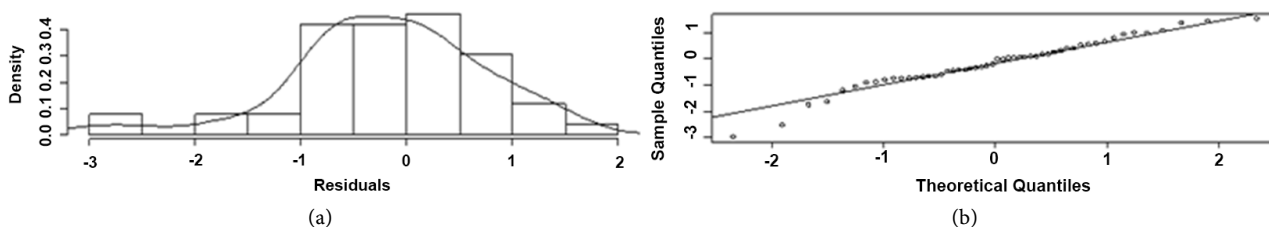**Figure 2.** Time series plot and ACF plot of Bayesian residuals for the CAPM.

Figure 3. Histogram and normal Q-Q plot of Bayesian residuals for the CAPM.

normalizing transformation, the test statistic follows a standard normal distribution [24].

The p-values presented in Table 2 confirm that there is no evidence against the homogeneity of variance assumption. There is some evidence against the normality assumption, though the ppp and partial-ppp show the evidence is not strong. The lack of normality is likely attributable to the two smallest residual values. The classical p-values and the partial-ppp show evidence against the assumption that the model errors are independent. The ppp appears to be a bit more conservative for this test. The Portmanteau test is used to detect that there is evidence that the model errors have lag 1 autocorrelation.

4.3. Fit of the CAPM with Autocorrelated Errors

The evaluation of the model fit in Section 4.2 revealed some evidence of a discrepancy from the assumption of independent model errors. The approach of [25] can be used to incorporate autocorrelation into the CAPM. This approach will be used in this subsection to model first-order autocorrelation (AR(1)). Thus, the model in Equation (2) is now represented as

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta} + \varepsilon_t, \text{ with } \varepsilon_t = \phi \varepsilon_{t-1} + a_t, \text{ and } a_t \sim N\left(0, \sigma^2\right). \tag{16}$$

It is convenient to reparameterize the model in Equation (16) using the Prais-Winsten transformation [26]. Under this reparameterization,

$$y_t^* = \boldsymbol{x}_t^{*'} \boldsymbol{\beta} + a_t, \text{ with } y_t^* = y_t - \phi y_{t-1}, \text{ and } \boldsymbol{x}_t^* = \boldsymbol{x}_t - \phi \boldsymbol{x}_{t-1}. \tag{17}$$

The model errors a_t for this transformed model should satisfy the usual assumptions listed in Equation (16) for $t = 2, \cdots, T$. This model can be written in the same form as Equation (3), but with $\boldsymbol{y}^*$ in place of $\boldsymbol{y}$ and $\boldsymbol{X}^*$ in place of $\boldsymbol{X}$.

The exact same prior distributions on $\boldsymbol{\beta}$ and σ^2 are used. The additional autoregressive coefficient ϕ has a non-informative Jeffrey prior in which

Table 2. P-Values for the fit of the CAPM to the SRA data.

check	classical	ppp	partial-ppp
Lag 1 Corr.	0.0364	0.0791	0.0460
Equal Var	0.8915	0.8216	0.8791
Normality	0.0623	0.1049	0.0953

$f(\phi) \sim \left(1 - \phi^2\right)^{-1/2}$. Thus, the parameter vector $\boldsymbol{\theta}$ is now $\begin{bmatrix} \boldsymbol{\beta} & \sigma^2 & \phi \end{bmatrix}'$. The posterior simulation is performed using the hybrid algorithm involving both the M-H procedure and Gibbs sampling as follows. The updating of $\boldsymbol{\beta}$ and ϕ is conducted via the Gibbs sampler in Algorithm 2 using the full conditionals from [25]. The updating of σ^2 is conducted via the M-H procedure described in Algorithm 1.

After discarding the initial 500 draws for burn-in, 10,000 draws are used to obtain the posterior estimates. These estimates are given in Table 3. The new model has a DIC of 364.6 and squared error loss of 3075.09. Based upon these goodness-of-fit measures alone, the Prais-Winstern reparameterization of the CAPM would be preferred. The next section provides diagnostic checks on this proposed model.

4.4. Evaluation of Fit of the CAPM with Autocorrelated Errors

The residuals in Equations (9) and (10) can also be used for the CAPM with autocorrelated errors given in Equation (16). These residuals are based upon y^* and X^* defined in Section 4.3. Figure 1 shows the plot of the corresponding standardized Bayesian residuals against the posterior mean values for each of the 52 monthly time points based upon the CAPM with autocorrelated errors. This figure does not suggest any obvious departures from the homogeneity of variances assumption. However, there are again a few residuals with extreme negative values. Figure 4 shows a plot of the residuals versus time as well as a plot of the autocorrelation function (ACF). The plot of the residuals against time in Figure 4 looks similar to Figure 2. However, the ACF plot in Figure 4 no longer suggests the presence of autocorrelation of any order. Figure 5 contains a histogram and a normal quantile-quantile plot of the residuals. Both plots show the extreme negative residual values. The distribution of the model errors may have tails that are too heavy for the normal assumption to hold.
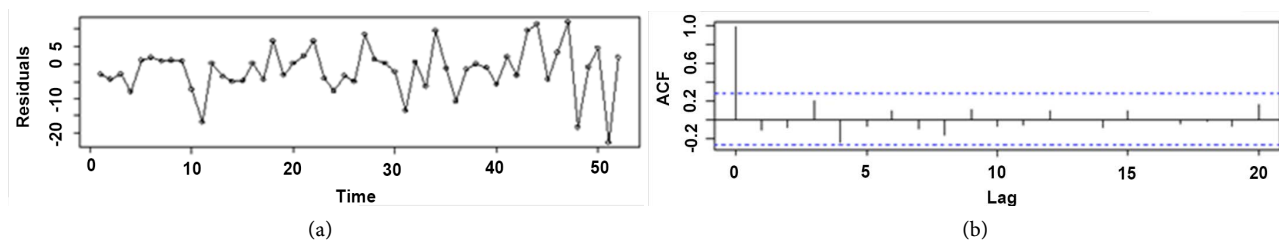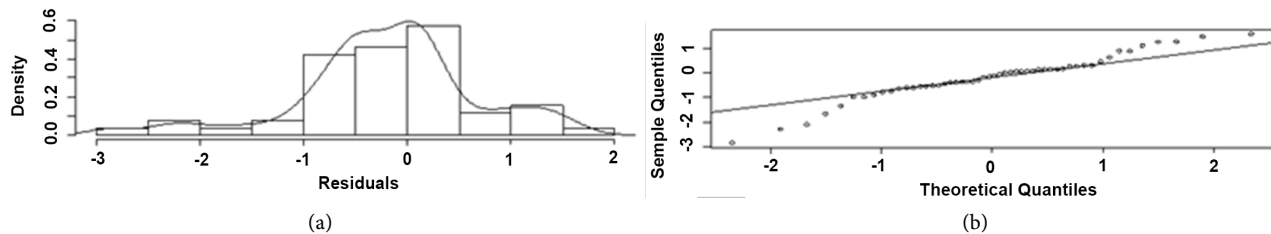
The p-values for the fit of the CAPM with autocorrelated errors are presented in Table 4. The p-values again confirm that there is no evidence against the homogeneity of variance assumption. The p-values show no evidence of autocorrelation. Thus, the modification in Equation (16) to account for the autocorrelation suitably addresses previously identified violations of the independence assumption. However, there is evidence against the normality assumption. This evidence is not as strong using the more conservative ppp.

Table 3. Posterior estimates of the CAPM with autocorrelated normal errors for the SRA data.

parameter	beta [1]	beta [2]	beta [3]	beta [4]	phi	sigma 2
mean	0.07877	1.328	0.1990	0.4630	−0.3232	63.21
std deviation	0.5587	0.2212	0.3312	0.2874	0.1553	12.47

Table 4. P-Values for the fit of the CAPM with autocorrelated normal errors to the SRA data.

check	classical	ppp	partial-ppp
Lag 1 Corr.	0.6397	0.5080	0.4073
Equal Var	0.4337	0.4352	0.2108
Normality	0.0142	0.0750	0.0471

(a)   (b)

Figure 4. Time series plot and ACF plot of Bayesian residuals for the CAPM with corrrelated normal errors.

(a)   (b)

Figure 5. Histogram and normal Q-Q plot of Bayesian residuals for the CAPM with corrrelated normal errors.

4.5. Fit of the CAPM with Autocorrelated *t* Errors

In order to address the problem of heavy tails in the fit of the CAPM with auto-regressive normal errors, one approach is to use the *t*-distribution to model the likelihood. A modification of Equation (17) to accommodate the *t*-distribution is given by

$$y_t^* = x_t^{*\prime} \beta + a_t, \ \text{with} \ a_t \sim T_{df}\left(0, \sigma^2\right). \tag{18}$$

The proper degrees of freedom (*df*) for the *t*-distribution can be determined using an exploratory method assuming the non-informative prior density $1/df \sim \text{Uniform}(0,1)$. This form of prior is recommended over the more obvious choice of $df \sim \text{Uniform}(1,\infty)$ because the latter essentially has all its mass near $df = \infty$ ([27], p. 193). For the SRA data, the posterior distribution of df centers around the value of 10 which is taken to be the value for the degrees of freedom.

As mentioned in ([12], p. 303), the *t* likelihood with degrees of freedom df

for each a_t can be represented as

$$a_t \mid h_t \sim N(0, h_t) \quad \text{and} \quad h_t \sim \text{IG}(df, \sigma^2), \tag{19}$$

where the variables h_t are auxiliary variables that cannot be directly observed. There is no direct way to compute h_t, but it is straightforward to obtain the conditional posterior distribution ([12], p 303) as

$$h_t \mid \boldsymbol{\beta}, \sigma^2, \phi, \boldsymbol{y} \sim \text{IG}\left(\frac{df+1}{2}, \frac{2}{df \cdot \sigma^2 + \left(y_t^* - \boldsymbol{x}_t^* \boldsymbol{\beta}\right)^2}\right). \tag{20}$$

The posterior simulation can be performed using a hybrid algorithm involving both the M-H procedure and Gibbs sampling. The updating of $\boldsymbol{\beta}$, ϕ, h_t is done via the Gibbs sampler in Algorithm 2. The updating of σ^2 is achieved via the M-H procedure described in Algorithm 1. All prior distributions remain the same for this model. The software program WinBUGS is used to obtain 10000 posterior draws after the 500 burn-in draws. Estimates under this model are given in Table 5. For the goodness-of-fit measures, DIC is 364.7 and the squared error loss is 3087.24. According to these criteria, the performance of this model is better than the CAPM in Section 4.1 and similar to the CAPM with autocorrelated normal errors in Section 4.3.

4.6. Evaluation of Fit of the CAPM with Autocorrelated *t* Errors

According to Equation (18), the model errors should have a *t*-distribution. Note that this corresponds to the marginal distribution of a_t across the values of h_t in Equation (19). Given values of h_t, the conditional distribution of $a_t \mid h_t$ is normal. Thus, for each draw from the posterior distribution, observed residuals are calculated using Equation (9). Values of h_t are also drawn according to Equation (20). Then the observed residuals are standardized by dividing by $\sqrt{h_t}$ as opposed to a function of σ based upon the *t*-distribution in Equation (18). The histogram and normal quantile-quantile plot of these standardized residuals are shown in Figure 6. The normal quantile-quantile plot shows some of the same deviations from the requisite line in the tails as that observed in Figure 5. However, the scaling from the *t*-distribution has substantially reduced the magnitude of these deviations at the tails. As a result, these residuals are closer to the reference line required for the normality assumption.

The p-values are provided in Table 6. The p-values for the test of normality show no evidence against the normality assumption. The p-values also show no evidence against the homogeneity of variance assumption or the independence assumption.

5. Conclusions

This paper reviews Bayesian methodology for diagnostic checking. These methods have been developed here for specifically assessing the assumptions of the CAPM as well as for suggesting meaningful model expansions. These approaches

Table 5. Posterior estimates of the CAPM with autocorrelated t errors for the SRA data.

parameter	beta [1]	beta [2]	beta [3]	beta [4]	phi	sigma 2
mean	0.1708	1.336	0.1875	0.5036	−0.2830	54.55
std deviation	0.5320	0.2243	0.3507	0.2944	0.1577	11.68

Table 6. P-Values for the fit of the CAPM with autocorrelated t errors to the SRA data.

check	classical	ppp	partial-ppp
Lag 1 Corr.	0.5698	0.4908	0.1288
Equal Var	0.5523	0.4992	0.3982
Normality	0.5161	0.4373	0.3820

(a)　　　　　　　　　　　　　　　　(b)
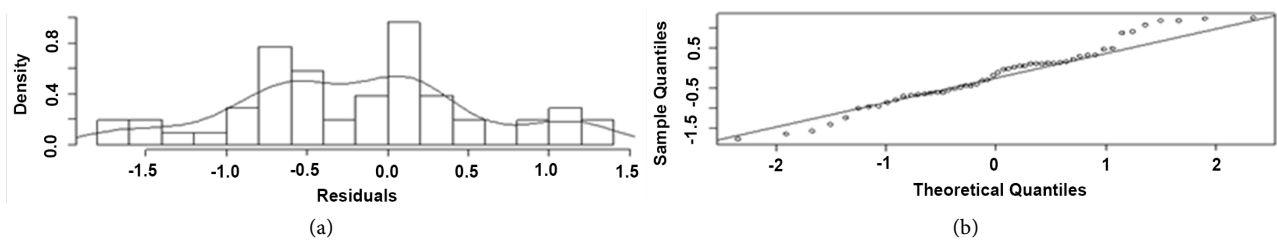
Figure 6. Histogram and normal Q-Q plot of Bayesian residuals for the CAPM with corrrelated t errors.

include the use of (standardized) observed residuals and (standardized) realized residuals as discrepancy statistics. The Bayesian residuals can be plotted to visualize possible violations of the model assumptions. Posterior probabilities can be used to quantify the amount of discrepancy. Specific assessments in this study included checks of autocorrelated errors, heterogeneous variances, and non-normality. Such Bayesian diagnostic approaches would be suitable for other linear regression models and for general linear models that have correlated observations.

These approaches for evaluating the model fit are demonstrated in Section 4 for a series of return data from a particular stock in the pharmaceutical industry. Serial correlation is detected in the fit of the CAPM. From Table 3, the incorporation of the first-order autocorrelation has a strong effect on the posterior estimate of the intercept and reduces the posterior variances associated with the estimates. The posterior estimate of −0.32 indicates a relatively strong autocorrelation. According to DIC and squared error loss, this model seems to provide a good fit. However, the Bayesian residuals indicate that for many observations there is large discrepancy between the data and the model fit. There are enough of these large discrepancy values for both residual plots and p-values to indicate problems with the normality assumption. If the objective is to have a suitable model for all stock returns, then the t likelihood can be used to model the extreme observations by increasing the model variability. The posterior estimates in Table 5 are roughly in between those in Table 1 and those in Table 3.

Acknowledgements

References

[1] Fama, E.F. and French K.R. (1993) Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, **33**, 3-56. https://doi.org/10.1016/0304-405X(93)90023-5

[2] Pastor, L. and Stambaugh, R.F. (1999) Cost of Equity Capital and Model Mispricing. *Journal of Finance*, **54**, 67-121. https://doi.org/10.1111/0022-1082.00099

[3] Lo, A.W. (2000) Finance: A Selective Survey. *Journal of American Statistical Association*, **95**, 629-635. https://doi.org/10.1080/01621459.2000.10474239

[4] Tsay, R.S. (2002) Analysis of Financial Time Series. John Wiley & Sons, New York. https://doi.org/10.1002/0471264105

[5] Brav, A. (2000) Inference in Long-Horizon Event Studies: A Bayesian Approach with Application to Initial Public Offerings. *The Journal of Finance*, **55**, 1979-2016. https://doi.org/10.1111/0022-1082.00279

[6] Campbell, J.Y., Lo, A.W., and MacKinlay, A.C. (1997) The Econometrics of Financial Markets. Princeton University Press, New Jersey.

[7] Zellner, A. (1971) An Introduction Bayesian Inference in Econometrics. John Wiley & Sons, New York.

[8] Hastings, W.K. (1970) Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**, 97-109. https://doi.org/10.1093/biomet/57.1.97

[9] Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398-409. https://doi.org/10.1080/01621459.1990.10476213

[10] Little, R. (2006) Calibrated Bayes: A Bayes/Frequentist Roadmap. *American Statistician*, **60**, 213-223. https://doi.org/10.1198/000313006X117837

[11] Pettit, L.I. (1986) Diagnostics in Bayesian Model Choice. *The Statistician*, **35**, 183-190. https://doi.org/10.2307/2987522

[12] Gelman, A., Carlin, J.B., Stern, H. and Rubin, D.B. (2003) Bayesian Data Analysis. 2nd Edition, Chapman and Hall, London.

[13] Gelman, A., Meng, X.L. and Stern, H. (1996) Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, **6**, 733-807.

[14] Carlin, B.P. and Louis, T.A. (2000) Bayes and Empirical Bayes Methods for Data Analysis. 2nd Edition. Chapman and Hall/CRC, Boca Raton. https://doi.org/10.1201/9781420057669

[15] Meng, X.L. (1994) Posterior Predictive P-values. *The Annals of Statistics*, **22**, 1142-1160. https://doi.org/10.1214/aos/1176325622

[16] Bayarri, M.J. and Berger, J.O. (2000) P Values for Composite Null Models. *Journal of the American Statistical Association*, **95**, 1127-1142. https://doi.org/10.2307/2669749

[17] Robins, J.M., van der Vaart, A. and Ventura, V. (2000) Asymptotic Distribution of P Values in Composite Null Models. *Journal of the American Statistical Association*,

95, 1143-1156.

[18] Bayarri, M.J. and Berger, J.O. (1999) Quantifying Surprise in the Data and Model Verification. In: Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., Eds., *Bayesian Statistics* 6, Oxford University Press, Oxford, 53-82.

[19] Bayarri, M.J., Castellanos, M.E. and Morales, J. (2006) MCMC Methods to Approximate Conditional Predictive Distributions. *Computational Statistics & Data Analysis*, **51**, 621-640. https://doi.org/10.1016/j.csda.2006.01.018

[20] Johnson, V.E. (2007) Bayesian Model Assessment using Pivotal Quantities. *Bayesian Analysis*, **2**, 719-734. https://doi.org/10.1214/07-BA229

[21] Box, G.E.P. (1980) Sampling and Bayes' Inference in Scientific Modeling and Robustness. *Journal of the Royal Statistical Society, Series A*, **143**, 383-430. https://doi.org/10.2307/2982063

[22] Gelfand, A.E. and Ghosh, S.K. (1998) Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika*, **85**, 1-11. https://doi.org/10.1093/biomet/85.1.1

[23] Greene, H.W. (2000) Econometric Analysis. 4th Edition, Prentice Hall, Upper Saddle River.

[24] Royston, J.P. (1982) An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Applied Statistics*, **31**, 115-124. https://doi.org/10.2307/2347973

[25] Chib, S. (1993) Bayesian Regression with Autoregressive Errors—A Gibbs Sampling Approach. *Journal of Econometrics*, **58**, 275-294. https://doi.org/10.1016/0304-4076(93)90046-8

[26] Prais, S.J. and Winsten, C.B. (1954) Trend Estimators and Serial Correlation. Cowles Commission Discussion Paper, No. 383.

[27] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) Markov Chain Monte Carlo in Practice. Chapman and Hall/CRC, London.