8-2012

Robust Deviance Information Criterion for Latent Variable Models

Yong LI
Renmin University of China

Tao ZENG
Wuhan University

Jun YU
Singapore Management University, yujun@smu.edu.sg

Citation

LI, Yong; ZENG, Tao; and YU, Jun. Robust Deviance Information Criterion for Latent Variable Models. (2012). 1-43. Research Collection School Of Economics.
Available at: https://ink.library.smu.edu.sg/soe_research/1403

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Robust Deviance Information Criterion for Latent Variable Models

Yong Li, Tao Zeng and Jun Yu

August 2012

Paper No. 30 – 2012

Robust Deviance Information Criterion for Latent Variable Models[*]

Yong Li
Renmin University of China

Tao Zeng
Singapore Management University

Jun Yu
Singapore Management University

Abstract

It is shown in this paper that the data augmentation technique undermines the theoretical underpinnings of the deviance information criterion (DIC), a widely used information criterion for Bayesian model comparison, although it facilitates parameter estimation for latent variable models via Markov chain Monte Carlo (MCMC) simulation. Data augmentation makes the likelihood function non-regular and hence invalidates the standard asymptotic arguments. A new information criterion, robust DIC (RDIC), is proposed for Bayesian comparison of latent variable models. RDIC is shown to be a good approximation to DIC without data augmentation. While the later quantity is difficult to compute, the expectation – maximization (EM) algorithm facilitates the computation of RDIC when the MCMC output is available. Moreover, RDIC is robust to nonlinear transformations of latent variables and distributional representations of model specification. The proposed approach is illustrated using several popular models in economics and finance.

JEL classification: C11, C12, G12
Keywords: AIC; DIC; EM Algorithm; Latent variable models; Markov Chain Monte Carlo.

1 Introduction

One of the most important developments in the Bayesian literature in recent years is the deviance information criterion (DIC) of Spiegelhalter et al. (2002). DIC is a Bayesian version of the well known Akaike Information Criterion (AIC) (Akaike (1973)). Like AIC, it trades off a measure of model adequacy against a measure of complexity and is concerned about how replicate data predict the observed data. DIC is constructed based on the posterior

[*] Yong Li, Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing, 1000872, P.R. China. He is currently a postdoc at Sim Kee Boon Institute for Financial Economics at Singapore Management University. Tao Zeng, School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Jun Yu, Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business. Email for Jun Yu: yujun@smu.edu.sg. URL: http://www.mysmu.edu/faculty/yujun/.

distribution of the log-likelihood or the deviance, and has several desirable features. First, DIC is simple to calculate when the likelihood function is available in closed-form and the posterior distributions of the models are obtained by Markov chain Monte Carlo (MCMC) simulation. Second, it is applicable to a wide range of statistical models. Third, unlike Bayes factors (BFs), it can be implemented under noninformative priors.

An important class of models in economics and finance involves latent variables. Latent variables have figured prominently in stories about consumption decision, investment decision, labor force participation, conducts of monetary policy, indices of economic activity, inflation dynamics and other economic, business and financial activities and decisions. For example, one important class of latent variable models, the state space model, in which the state variable is latent, provides a unified methodology for treating a wide range of problems in time series analysis. Another example can be found in the values of stocks, bonds, options, futures, and derivatives which are often determined by a small number of factors. Sometimes these factors, such as the level, the slope and the curvature in the term structure of interest rates, are not observed. In microeconometrics, discrete choices can depend on unobserved variables or there may be unobserved individual heterogeneity across economic entities.

For latent variable models, Bayesian methods via MCMC simulation have proven to be a powerful alternative to frequentist methods for estimating model parameters. In particular, the *data augmentation* strategy proposed by Tanner and Wong (1987), which expands the parameter space by treating the latent variables as additional model parameters, has been found very useful for simplifying the MCMC computation of posterior distributions. This simplification is achieved because data augmentation leads to a closed-form expression for the likelihood function.

Comparing alternative latent variable models in the Bayesian paradigm is a daunting and yet important task. The gold standard to carry out Bayesian model comparison is to compute BFs, which basically compare marginal likelihood of alternative models (Kass and Raftery (1995)). Several interesting developments have been made in recent years for computing marginal likelihood from the MCMC output; see for example, Chib (1995), Chib and Jeliazkov (2001). While these methods are very general and widely applicable, for latent variable models, they are difficult to use because the marginal likelihood may be very hard to calculate. In addition, BFs cannot be used under improper priors and are subject to the Jeffrey-Lindley paradox. Given that DIC is simple to calculate from the MCMC output with the data augmentation technique and also that data augmentation is often used for Bayesian parameter estimation, DIC has been used widely for comparing alternative latent variable models; see for example, Berg et al. (2004), Huang and Yu (2010).

The first contribution of this paper is that we argue DIC has to be used with care in the context of latent variable models. In particular, we believe DIC, as it is commonly implemented in practice, has some conceptual and practical problems. Firstly, DIC requires a concrete "focus" which is often not easily identified in practice. If the "focus" cannot be identified, using DIC violates the likelihood principle; see Gelfand and Trevisani (2002). Secondly, DIC is not robust to apparently innocuous transformations and distributional representations. This problem is made worse by the data augmentation technique for latent variable models. Data augmentation greatly inflates the number of parameters and hence the "effective" number of parameter used in DIC is very sensitive to transformations and distributional representations. The detail will be explained in Section 3. Finally, DIC requires that the likelihood function has a closed form expression for it to be computationally operational. For latent variable models, this requires the use of data augmentation and, as a consequence, DIC opens up to possible variations. It is unclear which variation should be used in practice; see Celeux et al. (2006) for further discussion of this problem. In this paper we argue that although data augmentation leads to a likelihood function in closed-form and greatly facilitates parameter estimation, DIC should NOT be used in connection to data augmentation. The reason is that data augmentation makes the likelihood function non-regular and hence invalidates the standard asymptotic arguments. Consequently, it undermines the theoretical underpinnings of DIC.

The source of the problem is data augmentation. With data augmentation, a closed-form expression for likelihood is ensured and it is easy to compute DIC, but the asymptotic justification of DIC is invalidated. Without data augmentation, the likelihood function does not have a closed form expression and hence DIC is not operational for latent variable models. However, it is asymptotically justified.

The second contribution of this paper is that we propose a new information criterion, robust DIC (RDIC), to make Bayesian comparison of latent variable models. It is shown that RDIC is a good approximation to DIC without data augmentation and hence is theoretically justified. We then show that the expectation – maximization (EM) algorithm facilitates the computation of RDIC for latent variable models when the MCMC output is available. Moreover, RDIC is robust to nonlinear transformations of latent variables and to distributional representations of model specification. The advantages of the proposed approach are illustrated using several popular models in economics and finance.

The paper is organized as follows. In Section 2, the latent variable models are introduced. The Bayesian estimation method with data augmentation and the EM algorithm are also reviewed. Section 3 reviews DIC, proposes and justifies RDIC for latent variable models, and discusses how to compute RDIC from the MCMC output. Section 4 illustrates the method using models from economics and finance. Section 5 concludes the paper. The Appendix collects the proof of the theoretical results in the paper.

2 Latent Variable Models, EM Algorithm and MCMC

Let $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ denote observed variables and $\mathbf{z} = (z_1, z_2, \cdots, z_n)'$ the latent variables. The latent variable model is indexed by the a set of P parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)'$. Let $p(\mathbf{y}|\boldsymbol{\theta})$ be the likelihood function of the observed data (denoted the observed-data likelihood), and $p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta})$ be the complete-data likelihood function. The relationship between the two functions is:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}, \mathbf{z}|\boldsymbol{\theta}) d\mathbf{z}. \tag{1}$$

In many cases, the integral does not have a closed-form solution. Consequently, statistical inferences, such as estimation and model comparison, are difficult to make. In the literature, maximum likelihood (ML) analysis using the EM algorithm and Bayesian analysis using MCMC are two popular approaches for carrying out statistical inference of the latent variable models.

2.1 Maximum likelihood via the EM algorithm

The EM algorithm is an iterative numerical method for finding the ML estimates of $\boldsymbol{\theta}$ in the latent variable models. It has been widely used in applications since Dempster et al. (1977) gave its name and did the convergence analysis. In this subsection, we briefly review the main idea of the EM algorithm. For more details, one can refer to McLachlan and Krishnan (2008).

Let $\mathbf{x} = (\mathbf{y}, \mathbf{z})$ be the complete data with a density $p(\mathbf{x}|\boldsymbol{\theta})$ parameterized by a P-dimension parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq R^P$. The observed-data log-likelihood $\mathcal{L}_o(\mathbf{y}|\boldsymbol{\theta}) = \ln p(\mathbf{y}|\boldsymbol{\theta})$ often involves some intractable integral, preventing researchers from directly optimizing $\mathcal{L}_o(\mathbf{y}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In many cases, however, the complete-data log-likelihood $\mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta}) = \ln p(\mathbf{x}|\boldsymbol{\theta})$ has a closed-form expression. Instead of maximizing $\mathcal{L}_o(\mathbf{y}|\boldsymbol{\theta})$ directly, the EM algorithm maximizes $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$, the conditional expectation of the complete-data log-likelihood function $\mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta})$ given the observed data $\mathbf{y}$ and a current fit $\boldsymbol{\theta}^{(r)}$ of the parameter.

Generally, a standard EM algorithm has two steps: the *expectation* (E) step and the *maximization* (M) step. The E-step evaluates

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = E_{\mathbf{z}}\{\mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta})|\mathbf{y}, \theta^{(r)}\}, \tag{2}$$

where the expectation is taken with respect to the conditional distribution $p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}^{(r)})$. The M-step determines a $\boldsymbol{\theta}^{(r+1)}$ that maximizes $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$. Under some mild regularity conditions, the sequence $\{\boldsymbol{\theta}^{(r)}\}$ obtained from the EM iterations converges to the ML estimate $\widehat{\boldsymbol{\theta}}$; see Dempster et al. (1977) and Wu (1983) for details about the convergence properties of $\{\boldsymbol{\theta}^{(r)}\}$.

2.2 Bayesian analysis using MCMC

Although the EM algorithm is a reasonable statistical approach for analyzing latent variable models, the numerical optimization in the M-step is often unstable. This numerical problem worsens as the dimension of $\boldsymbol{\theta}$ increases. It is well recognized that Bayesian methods using MCMC provide a powerful tool to analyze the latent variables models. However, if the posterior analysis is conducted from the observed-data likelihood, $p(\mathbf{y}|\boldsymbol{\theta})$, one would end up with the same problem as in the ML method as $p(\mathbf{y}|\boldsymbol{\theta})$ does not have a closed-form expression.

The novelty in the Bayesian methods is to treat the latent variable model as a hierarchical structure of conditional distributions, namely, $p(\mathbf{y}|\mathbf{z},\boldsymbol{\theta})$, $p(\mathbf{z}|\boldsymbol{\theta})$, and $p(\boldsymbol{\theta})$. In other words, one can use the data augmentation strategy of Tanner and Wong (1987) to expand the parameter space from $\boldsymbol{\theta}$ to $(\boldsymbol{\theta}$,$\mathbf{z})$. The advantage of data augmentation is that the Bayesian analysis is now based on the new likelihood function, $p(\mathbf{y}|\boldsymbol{\theta}$,$\mathbf{z})$ which often has a closed-form expression. Then the Gibbs sampler and other MCMC samplers can be used to generate random samples from the joint posterior distribution $p(\boldsymbol{\theta},\mathbf{z}|\mathbf{y})$. After a sufficiently long period for a burning-in phase, the simulated random samples can be regarded as random observations from the joint distribution. The statistical analysis can be established on the basis of these simulated posterior random observations. As a by-product to the Bayesian analysis, one also obtains Markov chains for the latent variables $\mathbf{z}$ and hence statistical inference can be made about $\mathbf{z}$. For further details about Bayesian analysis of latent variable models via MCMC, including algorithms, examples and references, see Geweke et al. (2011). From the above discussion, it can be seen that data augmentation is the key technique for Bayesian estimation of latent variable models.

Two observations are in order. First, with data augmentation, the parameter space is much bigger. More than often, the dimension of the space increases as the number of observations increases and is larger than the number of observations. In the latter case, the new likelihood function becomes non-regular. Second, it is difficult to argue that the latent variables can be always treated as the model parameters. Models parameters are typically fixed but the latent variables are often time varying. Consequently, the same treatment of these two types of variables does not seem to be justifiable from the perspective of model selection.

3 Bayesian Comparison of Latent Variable Models

3.1 DIC

Spiegelhalter et al. (2002) proposed DIC for Bayesian model comparison. The criterion is based on the deviance given by:

$$D(\boldsymbol{\theta}) = -2\ln p(\mathbf{y}|\boldsymbol{\theta}) + 2\ln f(\mathbf{y}),$$

where $f(\mathbf{y})$ is some fully specified standardizing term that is a function of the data alone. Based on the deviance, DIC takes the form of:

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + P_D. \tag{3}$$

The first term, used as a Bayesian measure of model fit, is defined as the posterior expectation of the deviance, that is,

$$\overline{D(\boldsymbol{\theta})} = E_{\theta|\mathbf{y}}[D(\boldsymbol{\theta})] = E_{\theta|\mathbf{y}}[-2\ln p(\mathbf{y}|\boldsymbol{\theta})].$$

The better the model fits the data, the larger the log-likelihood value and hence the smaller the value for $\overline{D(\boldsymbol{\theta})}$. The second term, used to measure the model complexity and also known as "effective number of parameters", is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$P_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) = -2\int [\ln p(\mathbf{y}|\boldsymbol{\theta}) - \ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})] p(\boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}, \tag{4}$$

where $\bar{\boldsymbol{\theta}}$ is the Bayesian estimator, and more precisely the posterior mean, of the parameter $\boldsymbol{\theta}$. Here, P_D can be explained as the expected excess of the true over the estimated residual information conditional on data $\mathbf{y}$. In other words, P_D can be interpreted as the expected reduction in uncertainty due to estimation.

Note that DIC can be rewritten by two equivalent forms:

$$\text{DIC} = D(\bar{\boldsymbol{\theta}}) + 2P_D, \tag{5}$$

and

$$\text{DIC} = 2\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) = -4E_{\theta|\mathbf{y}}[\ln p(\mathbf{y}|\boldsymbol{\theta})] + 2\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}). \tag{6}$$

DIC defined in Equation (5) bears similarity to AIC of Akaike (1973) and can be interpreted as a classical "plug-in" measure of fit plus a measure of complexity. In Equation (3) the Bayesian measure, $\overline{D(\boldsymbol{\theta})}$, is the same as $D(\bar{\boldsymbol{\theta}}) + P_D$ which already includes a penalty term for model complexity and thus could be better thought of as a measure of model adequacy rather than pure goodness of fit.

Remark 3.1 *The asymptotic justification of DIC requires that the candidate models nest the true model and that the posterior distribution is approximately normal. These two requirements parallel to those in AIC where the candidate models nest the true model and the ML estimator is asymptotically normally distributed. To see the importance of the asymptotic normality, Spiegelhalter et al. (2002) show that, when the prior is noninformative, P_D is approximately the same as P. In this case DIC is explained as Bayesian version of AIC. However, if the asymptotic normality does not hold true, P_D cannot be approximated by P and DIC is not the Bayesian version of AIC. Furthermore, the decision-theoretical explanation of DIC requires the asymptotic normality of the Bayesian posterior be held true.*

Remark 3.2 *If $p(\mathbf{y}|\boldsymbol{\theta})$ has a closed-form expression, DIC is trivially computable from the MCMC output. This is in sharp contrast to BFs and some other model selection criteria* *within the classical framework. The computational tractability, together with the versatility of MCMC and the fact that DIC is incorporated into a Bayesian software, WinBUGS, allows DIC to enjoy a very wide range of applications.*[1] *However, if $p(\mathbf{y}|\boldsymbol{\theta})$ is not available in closed form, such as in random effects models and state space models, computing DIC may become infeasible, or at least, very time consuming.*

Remark 3.3 *When an information criterion is used for model selection, the degrees of freedom are typically used to measure the model complexity. In the Bayesian framework, the prior information almost always imposes additional restrictions on the parameter space and hence the degrees of freedom may be reduced by the prior information. A useful contribution of DIC is to provide a way to measure the model complexity when the prior information is incorporated; see Brooks (2002).*

Remark 3.4 *Unlike BFs that address how observed data are predicted by the priors, DIC "addresses how well the posterior might predict future data generated by the same mechanism that gave rise to the observed data" (Spiegelhalter et al. (2002)). This predictive perspective for selecting a good model is important in many practical business, economic, and financial decisions.*

Remark 3.5 *DIC has a number of drawbacks, however. For instance, as acknowledged in Spiegelhalter et al. (2002), DIC requires a concrete specification of a "focus". In practice, however, the choice of a "focus" is not always easy. Unfortunately, it is well known that Bayesian decisions may depend on the choice of the "focus". For example, in Section 8.2 of Spiegelhalter et al. (2002), where Models 4 and 5 are predictively identical but their DIC values are quite different. In this example, it is unclear what should be the right "focus". The same difficulty also shows up in Model 8 of Berg et al. (2004). If the "focus" is not identified, DIC suffers from an incoherent inference problem. That is, when one model is a distributional representation of another model and the same prior is used in the two models, they have different DIC values. For further illustrations of the problem, see Gelfand and Trevisani (2002) and Daniels and Hogan (2008).*

For latent variable models, there are alternative ways to define DIC, as discussed in Celeux et al. (2006) (see also, DeIorio and Robert (2002)), two of which are especially important. First, DIC is based on the observed-data likelihood and denoted by DIC_1 in Celeux et al. (2006) as,

$$\text{DIC}_1 = -4E_{\theta|\mathbf{y}}[\ln p(\mathbf{y}|\boldsymbol{\theta})] + 2\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}). \tag{7}$$

For certain mixture models, such as scale mixtures of normals of Andrews and Mallows (1974) , the observed-data likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ is available in closed form. In this case, DIC_1 is trivially

[1] As of July 8, 2012, Spiegelhalter et al. (2002) has been cited 3396 times according to Google Scholar and 1,984 time according to Science Citation Index.

obtained, although its value depends on the choice of the “focus”, namely, the hierarchical structure here.

However, for state-space models, including linear Gaussian state space models, the observed-data likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ is not available in closed form.[2] In this case, computing DIC_1 from the MCMC output is time consuming or even infeasible since $p(\mathbf{y}|\boldsymbol{\theta})$ has to be computed at each draw from the Markov chain.

Second, DIC is defined based on the data augmentation technique, treating $\mathbf{z}$ as the additional parameters, and denoted by DIC_7 in Celeux et al. (2006) as,

$$\text{DIC}_7 = -4E_{\theta,\mathbf{z}|\mathbf{y}}[\ln p(\mathbf{y}|\mathbf{z},\boldsymbol{\theta})] + 2\ln p(\mathbf{y}|\bar{\mathbf{z}},\bar{\boldsymbol{\theta}})]. \tag{8}$$

The corresponding P_D is

$$P_D = -2\int [\ln p(\mathbf{y}|\mathbf{z},\boldsymbol{\theta}) - \ln p(\mathbf{y}|\bar{\mathbf{z}},\bar{\boldsymbol{\theta}})]p(\mathbf{z},\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\mathbf{z}\mathrm{d}\boldsymbol{\theta}. \tag{9}$$

For most state space models, including the nonlinear non-Gaussian state space models, $p(\mathbf{y}|\mathbf{z},\boldsymbol{\theta})$ is available in closed form and hence computing DIC_7 is trivial.

Remark 3.6 *For all random effects models and state space models, applied researchers always calculate DIC based on DIC_7 in (8) which is also implemented in WinBUGS. Examples that use DIC_7 in applications include Berg et al. (2004) and Wang et al. (2011). Clearly this choice of defining DIC is simply for computational convenience.*

Remark 3.7 *From a theoretical viewpoint, DIC_7 has a couple of serious problems. First, due to the data augmentation, the number of the latent variables often increases with the sample size in latent variable models, causing the problem of a non-regular likelihood-based statistical inference; see Gelman (2004). This invalidates the asymptotic justification of DIC because the standard asymptotic theory derived from regular likelihood is not applicable to non-regular likelihood. Second, due to the data augmentation, the dimension of the parameter space becomes larger and hence we expect that DIC_7 is more sensitive to transformations of latent variables than DIC_1.*

To illustrate the second problem, we consider a simple transformation of latent variables in the well known Clark model (Clark (1973)) which is given by,

$$\text{Model 1} : y_t \sim N(\mu, \exp(h_t)), h_t \sim N(0,\sigma^2), t = 1,\cdots,n. \tag{10}$$

An equivalent representation of the model is

$$\text{Model 2} : y_t \sim N(\mu, \sigma_t^2), \sigma_t^2 \sim LN(0,\sigma^2), t = 1,\cdots,n, \tag{11}$$

[2] For linear Gaussian state space models, to do ML, the Kalman filter can be used to obtain the likelihood function numerically.

where LN denotes the log-normal distribution. In Model 2 the latent variable is the volatility σ_t^2, while the latent variable is the logarithmic volatility $h_t = \ln \sigma_t^2$ in Model 1. Suppose the parameters of interest are μ and σ^2. With the same "focus", the two models are identical and hence are expected to have the same DIC and P_D. To calculate the P_D component in DIC$_7$, we simulate 1000 observations from the model with $\mu = 0, \sigma^2 = 0.5$. Vague priors are selected for the two parameters, namely, $\mu \sim N(0, 100)$, $\sigma^{-2} \sim \Gamma(0.001, 0.001)$. We run Gibbs sampler to make 240,000 simulated draws from the posterior distributions. The first 40,000 are discarded as burn-in samples. The remaining observations with every 10th observation are collected as effective observations for statistical inference. With the data augmentation, the latent variables, h_t and σ_t^2 are regarded as parameters, and we find that $P_D = 89.806$ for Model 1 but $P_D = 59.366$ for Model 2. The difference is very significant. Given that we have the identical models and priors, and use the same dataset, the vast difference suggests that DIC$_7$ and the corresponding P_D are very sensitive to transformations of latent variables.

For latent variable models, DIC$_1$ does not suffer from the same theoretical problem as DIC$_7$. However, computing DIC$_1$ from the MCMC output is not feasible since $p(\mathbf{y}|\boldsymbol{\theta})$ is not available in closed-form and computing $E_{\boldsymbol{\theta}|\mathbf{y}}[\ln p(\mathbf{y}|\boldsymbol{\theta})]$ necessitates numerical calculation of $p(\mathbf{y}|\boldsymbol{\theta})$ at each draw from the Markov chain.

To summarize the problems with DIC in the context of latent variable models, while DIC$_7$ is trivial to calculate but cannot be theoretically justified, DIC$_1$ is theoretically justified but infeasible to compute.

3.2 RDIC

In this section we introduce a new information criterion which is theoretically justified and easy to calculate. To do so, we define a robust deviance information criterion (RDIC):

$$\text{RDIC} = D(\bar{\boldsymbol{\theta}}) + 2\mathbf{tr}\left\{\mathbf{I}(\bar{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\right\} = D(\bar{\boldsymbol{\theta}}) + 2P_D^*, \tag{12}$$

where

$$P_D^* = \mathbf{tr}\left\{\mathbf{I}(\bar{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\right\}, \tag{13}$$

with **tr** denoting the trace of a matrix,

$$\mathbf{I}(\boldsymbol{\theta}) = -\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, V(\bar{\boldsymbol{\theta}}) = E\left[\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)' |\mathbf{y}\right].$$

To justify the choice of RDIC, we will show that RDIC well approximates DIC$_1$ and P_D^* well approximates P_D that corresponds to DIC$_1$. We then show that how the EM algorithm facilitates the computation of RDIC from the MCMC output for latent variable models.

Let $L_n(\boldsymbol{\theta}) = \ln p(\boldsymbol{\theta}|\mathbf{y})$, $L_n^{(1)}(\boldsymbol{\theta}) = \partial \ln p(\boldsymbol{\theta}|\mathbf{y})/\partial\boldsymbol{\theta}$, $L_n^{(2)}(\boldsymbol{\theta}) = \partial^2 \ln p(\boldsymbol{\theta}|\mathbf{y})/\partial\boldsymbol{\theta}\boldsymbol{\theta}'$. In this paper, we impose the following regularity conditions.

Assumption 1: There exists a finite sample size n^*, for $n > n^*$, there is a local maximum at $\hat{\boldsymbol{\theta}}_m$ so that $L_n^{(1)}\left(\hat{\boldsymbol{\theta}}_m\right) = 0$ and $L_n^{(2)}\left(\hat{\boldsymbol{\theta}}_m\right)$ is a negative definite matrix. Obviously, $\hat{\boldsymbol{\theta}}_m$ is the posterior mode.

Assumption 2: The largest eigenvalue of $\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}}_m)\right]^{-1}$, σ_n^2, goes to zero when $n \to \infty$.

Assumption 3: For any $\epsilon > 0$, there exists an integer n^{**} and some $\delta > 0$ such that for any $n > \max\{n^*, n^{**}\}$ and $\boldsymbol{\theta} \in H\left(\hat{\boldsymbol{\theta}}_m, \delta\right) = \left\{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_m|| \leq \delta\right\}$, $L_n^{(2)}(\boldsymbol{\theta})$ exists and satisfies

$$-A(\epsilon) \leq L_n^{(2)}(\boldsymbol{\theta}) L_n^{-(2)}\left(\hat{\boldsymbol{\theta}}_m\right) - \mathbf{I}_P \leq A(\epsilon),$$

where $\mathbf{I}_P$ is a $P \times P$ identity matrix, $A(\epsilon)$ a $P \times P$ semi-definite symmetric matrix whose largest eigenvalue goes to zero as $\epsilon \to 0$.

Assumption 4: For any $\delta > 0$, as $n \to \infty$,

$$\int_{\boldsymbol{\Theta} - H(\hat{\boldsymbol{\theta}}_m, \delta)} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \to 0,$$

where $\boldsymbol{\Theta}$ is the support of $\boldsymbol{\theta}$.

Assumption 5: For any $\delta > 0$, as $n \to \infty$, when $\boldsymbol{\theta} \in H\left(\hat{\boldsymbol{\theta}}_m, \delta\right)$, conditional on the observed data $\mathbf{y}$, $L_n^{(2)}(\boldsymbol{\theta})/n = O(1)$.

Assumption 6: The likelihood information dominates the prior information, that is, when the sample size goes to infinity, the prior information can be ignored.

Assumption 7: Assume $\mathbf{y}_{rep}$ is a dataset that is derived from the same data generating process as gave rise to the observed data $\mathbf{y}$. For any $\delta > 0$, as $n \to \infty$, when $\boldsymbol{\theta} \in H\left(\hat{\boldsymbol{\theta}}_m, \delta\right)$, we assume

$$\frac{1}{n}\left[\frac{\partial^2 \ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}\right] = \frac{1}{n}\left[\frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}\right] + o_p(1).$$

Lemma 3.1 *Under Assumptions 1-5, conditional on the observed data* $\mathbf{y}$*, we have*

$$\begin{aligned}
&\bar{\boldsymbol{\theta}} = E\left[\boldsymbol{\theta}|\mathbf{y}\right] = \hat{\boldsymbol{\theta}}_{\boldsymbol{m}} + o(n^{-1/2}), \\
&V\left(\hat{\boldsymbol{\theta}}_m\right) = E\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_m\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_m\right)' |\mathbf{y}\right] = -L_n^{-(2)}\left(\hat{\boldsymbol{\theta}}_m\right) + o(n^{-1}).
\end{aligned}$$

Remark 3.8 *Lemma 3.1 establishes Bayesian large sample theory. The regularity conditions 1-4 have been used in the literature to develop Bayesian large sample theory for stationary and nonstationary dynamic models and nondynamic models; see, for example, Chen (1985), Kim (1994), Kim (1998), Geweke (2005). The Bayesian large sample theory was also developed from different sets of regularity conditions in different contexts. For example, Ghosh and Ramamoorthi (2003) developed the asymptotic posterior normality and Lemma 3.1 in the iid case.*

Theorem 3.1 *Under Assumptions 1-6, it can be shown that, conditional on the observed data* $\mathbf{y}$*,*

$$P_D = P_D^* + o(1),\ DIC_1 = RDIC + o(1),$$

where P_D is defined in (4).

Remark 3.9 *As DIC_1 is theoretically justified for the latent variable models, Theorem 3.1 justifies RDIC asymptotically since RDIC and DIC_1 are asymptotically equivalent.*

Suppose a loss function, when using the observed data $\mathbf{y}$ to predict $\mathbf{y}_{rep}$ in a model, is given by $\mathcal{L}(\mathbf{y}_{rep}, \mathbf{y})$. From the decision-theoretic viewpoint, a desirable model selection criterion should choose a model to minimize Bayesian risk, $E_{\mathbf{y}} E_{\mathbf{y}_{rep}|\mathbf{y}} \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y})$. The following theorem provides the justification of RDIC from the decision-theoretic viewpoint.

Theorem 3.2 *Based on the predictive distribution, $p(\mathbf{y}_{rep}|\mathbf{y}) = \int p(\mathbf{y}_{rep}|\mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$, the posterior mean of the predictive loss may be expressed as:*

$$\begin{aligned}
& E_{\mathbf{y}_{rep}|\mathbf{y}} \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) \\
= & \int \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) p(\mathbf{y}_{rep}|\mathbf{y}) \, d\mathbf{y}_{rep} \\
= & \int \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) \int p(\mathbf{y}_{rep}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) \, d\boldsymbol{\theta} \, d\mathbf{y}_{rep} \\
= & \int \int \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) p(\mathbf{y}_{rep}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) \, d\boldsymbol{\theta} \, d\mathbf{y}_{rep} \\
= & \int \int \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) p(\mathbf{y}_{rep}|\boldsymbol{\theta}) \, d\mathbf{y}_{rep} p(\boldsymbol{\theta}|\mathbf{y}) \, d\boldsymbol{\theta} \\
= & E_{\theta|\mathbf{y}} \left\{ \int \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) p(\mathbf{y}_{rep}|\boldsymbol{\theta}) \, d\mathbf{y}_{rep} \right\} \\
= & E_{\theta|\mathbf{y}} E_{\mathbf{y}_{rep}|\theta} \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}).
\end{aligned}$$

If $\mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) = -2 \ln p(\mathbf{y}_{rep}|\bar{\boldsymbol{\theta}}(\mathbf{y}))$, it can be shown that, conditional on the observed data $\mathbf{y}$ and under Assumptions 1-7,

$$E_{\mathbf{y}} E_{\mathbf{y}_{rep}|\mathbf{y}} \mathcal{L}(\mathbf{y}_{rep}, \mathbf{y}) = E_{\mathbf{y}}[DIC_1] + o(1) = E_{\mathbf{y}}[RDIC] + o(1).$$

Remark 3.10 *RDIC is an unbiased estimator of Bayesian risk asymptotically.*

Remark 3.11 *Like DIC_1, RDIC addresses how well the posterior may predict future data generated by the same mechanism that gives rise to the observed data. This posterior predictive feature could be appealing in many applications.*

Remark 3.12 *Like DIC_1, RDIC is justified by the standard Bayesian large sample theory. When the Bayesian large sample theory is not available, RDIC is not justified. These include models in which the number of the parameters increases with the sample size, under-identified models, models with an unbounded likelihood, and models with improper posterior distributions. For more details about the standard Bayesian large sample theory, see Gelman (2004) and Geweke (2005). For the latent variable models, since the number of the latent variables* *increases with sample size, the standard Bayesian large sample theory is not applicable if the data augmentation technique is used. As a result, when calculating RDIC, data augmentation should NOT be used.*

Remark 3.13 *It is easy to verify that Assumptions 1-7 hold true for nondynamic models or stationary dynamic models. Hence, Lemma 3.1 and Theorem 3.1 are applicable to these models. For unit root models, Kim (1994) and Kim (1998) showed that the asymptotic normality of posterior distribution can be established. Hence, Lemma 3.1 is applicable to models with a unit root. Unfortunately, Assumption 7, which is critical for developing Theorem 3.2, does not hold true for models with a unit or explosive root due to the initial condition. Consequently, Theorem 3.2 is not applicable to models with unit or explosive roots. This topic on comparing non-stationary statistical models will be pursued in future studies. Within the classical framework, Phillips and Ploberger (1996) and Phillips (1996) have proposed model selection criteria for models without latent variables.*

Remark 3.14 *RDIC maintains all the good features of DIC_1. For example, RDIC incorporates the prior information when measuring the model complexity. As shown in Spiegelhalter et al. (2002),*

$$I\left(\hat{\boldsymbol{\theta}}_m\right) = -\left\{\frac{\partial^2 \ln p(\boldsymbol{\theta}|\mathbf{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_m} = -L_n^{(2)}(\hat{\boldsymbol{\theta}}_m) - \left\{-\frac{\partial^2 \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_m}.$$

Under Assumption 1-5, following Lemma 3.1 and the proof of Theorem 3.1, we get

$$\begin{aligned}
P_D^* &= \mathbf{tr}\left\{I(\hat{\boldsymbol{\theta}}_m)V(\bar{\boldsymbol{\theta}})\right\} + o(1) \\
&= \mathbf{tr}\left\{\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}}_m) - \left\{-\frac{\partial^2 \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_m}\right]V(\bar{\boldsymbol{\theta}})\right\} + o(1) \\
&= \mathbf{tr}\left\{-L_n^{(2)}(\hat{\boldsymbol{\theta}}_m)V(\bar{\boldsymbol{\theta}})\right\} - \mathbf{tr}\left\{\left[-\frac{\partial^2 \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_m}\right]V(\bar{\boldsymbol{\theta}})\right\} + o(1) \\
&= P - \mathbf{tr}\left\{\left[-\frac{\partial^2 \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_m}\right]V(\bar{\boldsymbol{\theta}})\right\} + o(1). \qquad (14)
\end{aligned}$$

From (14), it can be seen clearly that the prior information can reduce the model complexity.

Remark 3.15 *Conditional on the observed data $\mathbf{y}$, when the likelihood information dominates the prior information (say, for example, if $-\partial^2 \ln p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\boldsymbol{\theta}'|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_m} = O(1)$), from (14) it can be shown that $P_D = P_D^* + o(1) = P + o(1)$. In addition, as $n \to \infty$ the posterior mode $\hat{\boldsymbol{\theta}}_m$ is reduced to the ML estimator $\hat{\boldsymbol{\theta}}$. Hence,*

$$\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) = \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})' I(\tilde{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}),$$

where $\tilde{\boldsymbol{\theta}}$ lies in the segment between $\bar{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$. Using Assumption 5 and Lemma 3.1, we can show that $\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) = \ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}) + o(1)$. Consequently,

$$DIC_1 = RDIC + o(1) = -2\ln p(\mathbf{y}|\hat{\boldsymbol{\theta}}) + 2P + o(1) = AIC + o(1).$$

Namely, both RDIC and DIC$_1$ can be regarded as the Bayesian version of AIC.

Remark 3.16 *Since RDIC is defined from the observed-data likelihood $p(\mathbf{y}|\boldsymbol{\theta})$, there is no need to specify a "focus", and hence, RDIC does not suffer from the incoherent inference problem.*

Remark 3.17 *For the latent variable models, while the number of the model parameters (P) is fixed and usually not so big, the number of the latent variables increases as the sample size increases. In the definition of RDIC, the latent variables are not regarded as the parameters. Consequently, the problem of parameter transformation is less serious. For example, in the Clark model, with the same setting as before, we get $P_D^* = 1.75$ for Model 1 and $P_D^* = 1.80$ for Model 2. There is no significant difference between them. Moreover, these two values are close to 2, that is the actual number of parameters. This is what we expected given that the vague priors are used and hence $P_D^* \approx P = 2$.*

Remark 3.18 *An obvious computational advantage in RDIC is that P_D^* does not involve inverting a matrix. This advantage is not so important when the latent variable model only has a small number of parameters. However, for high dimensional latent variable models where there are many parameters, this computational advantage may be important.*

Remark 3.19 *If the observed-data likelihood function, $p(\mathbf{y}|\boldsymbol{\theta})$, does not have a closed-from expression, its second derivative, $\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ and hence RDIC will be difficult to compute. In the following section, we show how the EM algorithm may be used to compute the second derivative and RDIC.*

3.3 Computing RDIC by the EM algorithm

The definition of RDIC clearly requires the evaluation of observed-data likelihood at the posterior mean, $p(\mathbf{y}|\bar{\boldsymbol{\theta}})$, as well as the information matrix and the second derivative of the observed-data likelihood function. For most latent variable models, the observed-data likelihood function does not have a closed-from expression. In this section we show how the EM algorithm may be used to evaluate $p(\mathbf{y}|\bar{\boldsymbol{\theta}})$, the second derivative of the observed-data likelihood function, and hence RDIC for the latent variable models. It is important to point out that we do not need to numerically optimize any function here as in the EM algorithm. Consequently, our method is not subject to the instability problem found in the M-step.

Lemma 3.2 *For any $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ in Θ, let $\mathcal{H}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \int \ln p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta}^*)d\mathbf{z}$, the so-called $\mathcal{H}$ function in the EM algorithm. It was shown in Dempster et al. (1977) that*

$$\mathcal{L}_o(\mathbf{y},\boldsymbol{\theta}) = \mathcal{Q}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^*\right) - \mathcal{H}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^*\right),$$

where the $\mathcal{Q}$ function is defined in Equation (2).

Following Lemma 3.2, the Bayesian plug-in model fit, $\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})$, may be obtained as

$$\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) = \mathcal{Q}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}}) - \mathcal{H}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}}). \tag{15}$$

It can be seen that even when $\mathcal{Q}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}})$ is not available in closed form, it is easy to evaluate from the MCMC output because

$$\mathcal{Q}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}}) = \int \ln p(\mathbf{y}, \mathbf{z}|\bar{\boldsymbol{\theta}}) p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\theta}}) \mathrm{d}\mathbf{z} \approx \frac{1}{M} \sum_{m=1}^{M} \ln p\left(\mathbf{y}, \mathbf{z}^{(m)}|\bar{\boldsymbol{\theta}}\right).$$

where $\{\mathbf{z}^{(m)}, m = 1, 2, \cdots, M\}$ are random observations drawn from the posterior distribution $p(\boldsymbol{z}|\mathbf{y}, \bar{\boldsymbol{\theta}})$.

For the second term in (15), if $p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\theta}})$ is a standard distribution, $\mathcal{H}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}})$ can be easily evaluated from the MCMC output as

$$\mathcal{H}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}}) = \int \ln p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\theta}}) p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\theta}}) d\mathbf{z} \approx \frac{1}{M} \sum_{m=1}^{M} \ln p\left(\mathbf{z}^{(m)}|\mathbf{y}, \bar{\boldsymbol{\theta}}\right).$$

However, if $p(\mathbf{z}|\mathbf{y}, \bar{\boldsymbol{\theta}})$ is not a standard distribution, an alternative approach has to be used, depending on the specific model in consideration. We now consider two situations.

First, if the complete-data $(\mathbf{y}_i, \mathbf{z}_i)$ are independent with $i \neq j$, and $\mathbf{z}_i$ is of low-dimension, say ≤ 5, then a nonparametric approach may be used to approximate the posterior distribution $p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta})$. Note that

$$\mathcal{H}(\boldsymbol{\theta}|\boldsymbol{\theta}) = \int \ln p(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}) \pi(\mathbf{z}|\mathbf{y}, \boldsymbol{\theta}) d\mathbf{z} = \sum_{i=1}^{n} \int \ln p(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\theta}) \pi(\mathbf{z}_i|\mathbf{y}, \boldsymbol{\theta}) d\mathbf{z}_i = \sum_{i=1}^{n} \mathcal{H}_i(\boldsymbol{\theta}|\boldsymbol{\theta}).$$

The computation of $\mathcal{H}_i(\boldsymbol{\theta}|\boldsymbol{\theta})$ requires an analytic approximation to $p(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\theta})$ which can be constructed using a nonparametric method. In particular, MCMC allows one to draw some effective samples from $p\left(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\theta}\right)$. Using these random samples, one can then use nonparametric techniques such as the kernel-based methods to approximate $p\left(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\theta}\right)$. In a recent study, Ibrahim et al. (2008) suggested using a truncated Hermite expansion to approximate $p(\mathbf{z}_i|\mathbf{y}_i, \boldsymbol{\theta})$.

As a simple illustration, we apply this method to the Clark model. When the Gaussian kernel method is used, we get $\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) = -1448.97$, RDIC= 2901.46 for Model 1 and $\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) = -1449.41$, RDIC= 2902.42 for Model 2. These two sets of numbers are nearly identical. However, if the latent variable models are regarded as parameters, we get DIC$_7$ = 2884.37 for Model 1 and DIC$_7$ = 2852.85 for Model 2. The highly distinctive difference between them suggests that DIC$_7$ is not a reliable model selection criterion for the model. Note that DIC$_1$ is not really feasible to compute in this case.

Second, for some latent variable models, the latent variables $\mathbf{z}$ follow a multivariate normal distribution and the observed variables $\mathbf{y}$ are independent conditional on $\mathbf{z}$. This class of models is referred to as the Gaussian latent variable models in the literature. In economics and finance, many latent variable models belong to this class of models, including dynamic linear models, dynamic factor models, various forms of stochastic volatility models and credit risk models. In these models, the observed-data likelihood is non-Gaussian but has a Gaussian flavor in the sense that the posterior distribution, $p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})$, may be expressed as,

$$p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\mathbf{z}'\boldsymbol{V}(\boldsymbol{\theta})\mathbf{z} + \sum_{i=1}^{n}\ln p(\mathbf{y}_i|\mathbf{z}_i,\boldsymbol{\theta})\right).$$

Rue et al. (2004) and Rue et al. (2009) showed that this type of posterior distribution can be well approximated by a Gaussian distribution that matches the mode and the curvature at the mode. The resulting approximation is known as the Laplace approximation and can be expressed as,

$$p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\mathbf{z}'(V(\boldsymbol{\theta}) + diag(\mathbf{c}))\mathbf{z}\right),$$

where $\mathbf{c}$ comes from the second order term in the Taylor expansion of $\sum_{i=1}^{n}\ln p(\mathbf{y}_i|\mathbf{z}_i)$ at the mode of $p(\mathbf{z}|\mathbf{y},\boldsymbol{\theta})$. The Laplace approximation may be employed to compute $\mathcal{H}(\bar{\boldsymbol{\theta}}|\bar{\boldsymbol{\theta}})$. After $p(\mathbf{y}|\bar{\boldsymbol{\theta}})$ is obtained, it is easy to obtain $D(\bar{\boldsymbol{\theta}})$. It is important to point out that the numerical evaluation of $p(\mathbf{y}|\bar{\boldsymbol{\theta}})$ is needed only once, i.e., at the posterior mean.

To compute P_D^*, we have to calculate the second derivative of the observed-data likelihood function in (14). The following two lemmas show how to compute the second derivatives.

Lemma 3.3 *Under the mild regularity conditions, the observed-data information matrix may be expressed as:*

$$\mathbf{I}(\boldsymbol{\theta}) = -\frac{\partial^2\mathcal{L}_o(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \left\{-\frac{\partial^2\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{*\prime}}\right\}_{\boldsymbol{\theta}^*=\boldsymbol{\theta}}. \quad (16)$$

Lemma 3.4 *Let $S(\mathbf{x}|\boldsymbol{\theta}) = \partial\mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Under the mild regularity condition, the observed-data information matrix has an equivalent form:*

$$\begin{aligned}\mathbf{I}(\boldsymbol{\theta}) = -\frac{\partial^2\mathcal{L}_o(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= E_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\left\{-\frac{\partial^2\mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\} - Var_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\left\{S(\mathbf{x}|\boldsymbol{\theta})\right\} \quad (17)\\ &= E_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\left\{-\frac{\partial^2\mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - S(\mathbf{x}|\boldsymbol{\theta})S(\mathbf{x}|\boldsymbol{\theta})'\right\} + E_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\{S(\mathbf{x}|\boldsymbol{\theta})\}E_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\{S(\mathbf{x}|\boldsymbol{\theta})\}',\end{aligned}$$

where all the expectations are taken with respect to the conditional distribution of $\mathbf{z}$ *given* $\mathbf{y}$ *and* $\boldsymbol{\theta}$*.*

Remark 3.20 *Lemma 3.3 and Lemma 3.4 were developed in Oakes (1999) and Louis (1982), respectively, for finding the standard error in the EM algorithm. If the $\mathcal{Q}$ function is available,* *we can use Lemma 3.3 to evaluate the second derivatives. If the $\mathcal{Q}$ function does not have an analytic form, we may use Lemma 3.4 to evaluate the second derivatives as follows,*

$$\begin{aligned} & E_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\left\{-\frac{\partial^2 \mathcal{L}_c(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} - S(\mathbf{x}|\boldsymbol{\theta})S(\mathbf{x}|\boldsymbol{\theta})^{'}\right\}, \\ \approx & -\frac{1}{M}\sum_{m=1}^{M}\left\{\frac{\partial^2 \mathcal{L}_c(\mathbf{y},\mathbf{z}^{(m)}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} + S(\mathbf{y},\mathbf{z}^{(m)}|\boldsymbol{\theta})S(\mathbf{y},\mathbf{z}^{(m)}|\boldsymbol{\theta})^{'}\right\}, \\ & E_{\mathbf{z}|\mathbf{y},\boldsymbol{\theta}}\{S(\mathbf{x}|\boldsymbol{\theta})\} \approx \frac{1}{M}\sum_{m=1}^{M} S(\mathbf{y},\mathbf{z}^{(m)}|\boldsymbol{\theta}), \end{aligned}$$

where $\{\mathbf{z}^{(m)}, m = 1, 2, \cdots, M\}$ are random observations drawn from the posterior distribution $p(\boldsymbol{z}|\mathbf{y},\boldsymbol{\theta})$.

4 Examples

We now illustrate the proposed method in three applications, covering some popular models in economics and finance. In the first example, both $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta})$ and $\mathcal{H}(\boldsymbol{\theta}|\boldsymbol{\theta})$ are available in closed-form and hence RDIC is trivial to compute. In this example, we pay attention to implications of different distributional representations. In the second example, while $p(\mathbf{y}|\bar{\boldsymbol{\theta}})$ is not available in closed-form, Kalman filter provides a recursive algorithm to evaluate it. Hence, $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta})$ and $\mathcal{H}(\boldsymbol{\theta}|\boldsymbol{\theta})$ can be calculated in the same manner, facilitating the computation of RDIC. In the third example, $p(\mathbf{y}|\bar{\boldsymbol{\theta}})$ is not available in closed-form and Kalman filter cannot be applied. To compute RDIC, we use the Laplace approximation and the technique suggested in Lemma 3.4.

4.1 Comparing asset pricing models

Asset pricing theory is fundamentally important in modern finance. A basic assumption required by much asset pricing theory is that the return distribution is normal. Unfortunately, there has been overwhelming empirical evidence against normality for asset returns, which have led researchers to investigate asset pricing models with heavy-tailed distributions, including the family of elliptical distributions discussed in Zhou (1993). Kan and Zhou (2003) suggested to use the multivariate t distribution to replace the multivariate normal distribution. In addition, under the mean-variance efficiency, the asset excess premium should not be statistically different from zero. In this section, we compare the following six asset pricing models:

$$
\begin{aligned}
&Model\ 1: R_t = \boldsymbol{\beta}'\boldsymbol{F}_t + \epsilon_t, \epsilon_t \sim N[\mathbf{0}, \boldsymbol{\Sigma}];\\
&Model\ 2: R_t = \alpha + \boldsymbol{\beta}'\boldsymbol{F}_t + \epsilon_t, \epsilon_t \sim N[\mathbf{0}, \boldsymbol{\Sigma}];\\
&Model\ 3: R_t = \boldsymbol{\beta}'\boldsymbol{F}_t + \epsilon_t, \epsilon_t \sim t[\mathbf{0}, \boldsymbol{\Sigma}, \nu];\\
&Model\ 4: R_t = \boldsymbol{\beta}'\boldsymbol{F}_t + \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}/\omega_t), \omega_t \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right);\\
&Model\ 5: R_t = \boldsymbol{\alpha} + \boldsymbol{\beta}'\boldsymbol{F}_t + \epsilon_t, \epsilon_t \sim t[\mathbf{0}, \boldsymbol{\Sigma}, \nu];\\
&Model\ 6: R_t = \boldsymbol{\alpha} + \boldsymbol{\beta}'\boldsymbol{F}_t + \epsilon_t, \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}/\omega_t), \omega_t \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right),
\end{aligned}
$$

where R_t is the excess return of portfolio at period t with $N \times 1$ dimension, $\boldsymbol{F}_t$ a $K \times 1$ vector of factor portfolio excess returns, $\boldsymbol{\alpha}$ a $N \times 1$ vector of intercepts, $\boldsymbol{\beta}$ a $N \times K$ vector of scaled covariances, ϵ_t the random error, $t = 1, 2, \cdots, n$. For convenience, we restrict $\boldsymbol{\Sigma}$ to be a diagonal matrix and ν to be a known constant. Note that Model 4 is the distributional representation of Model 3, and Model 5 is the distributional representation of Model 6. This is especially true if ω_t is not the quantity of interest.

Monthly returns of 25 portfolios, constructed at the end of each June, are the intersections of 5 portfolios formed on size (market equity, ME) and 5 portfolios formed on the ratio of book equity to market equity (BE/ME). The Fama/French's three factors, market excess return, SMB (Small Minus Big), HML (High Minus Low) are used as the explanatory factors (Fama and French (1993)). The sample period is from July 1926 to July 2011, so that $N = 25$, $n = 1021$. The data are freely available from the data library of Kenneth French.[3]

Bayesian analysis of the asset pricing models has attracted a considerable amount of attentions in the empirical asset pricing literature.[4] Here we apply DIC$_7$ and RDIC to compare Models 1-6. Based on the result of Li and Yu (2012), in the empirical study, we simply set $\nu = 3$. Some vague conjugate prior distributions are used to represent the prior ignorance, namely,

$$\alpha_i \sim N[0, 100], \beta_{ij} \sim N[0, 100], \phi_{ii}^{-1} \sim \Gamma[0.001, 0.001].$$

The use of uninformative priors implies that P_D^* should be close to the actual number of the parameters, P, if the posterior distribution is well approximated by the normal distribution.

Under these prior specifications, we use WinBUGS to implement Bayesian analysis and to calculate DIC$_7$. An introduction to WinBUGS can be found in Spiegelhalter et al. (2003). To calculate RDIC, we use R2WinBUGS, a R package that calls WinBUGS and exports the results into R (Sturtz et al. (2005)).[5] Since both $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta})$ and $\mathcal{H}(\boldsymbol{\theta}|\boldsymbol{\theta})$ are available in closed-form, RDIC is trivial to compute.

We sample 100,000 random observations from the posterior distributions in each model, the first 40,000 of which form the burn-in period. The convergence of the next 60,000 iterations

[3] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[4] Avramov and Zhou (2010) provided an excellent review of the literature on Bayesian portfolio analysis.

[5] R code may be requested from the authors of the present paper.

is checked using the Raftery-Lewis diagnostic test statistic (Raftery and Lewis (1992)) with every 3th observation collected. Hence, 20,000 effective observations are used for computing the information criteria. The value of DIC$_7$ is automatically calculated by WinBUGS. Based on the observed log-likelihood given in formula (18) in Appendix D, we can compute DIC and RDIC for Model 3 and 5. Table I reports DIC$_7$, RDIC, P_D, and P_D^* for all six models. Note that when there is no latent variable DIC$_7$ is reduced into DIC$_1$.

From Table I, we see that P_D is almost identical to P_D^* in each of Models 1, 2, 3 and 5. Not surprisingly, DIC$_7$ and RDIC are almost the same in each of these models. As expected, DIC$_7$ in Model 3 is quite different from that in Model 4 although these two models are the same. The main reason for this distinctive difference is that in Model 4, the scale-mixture specification is used and, hence, a sequence of latent variables, $\{\omega_t\}$, is introduced artificially. In DIC$_7$ the latent variables, $\{\omega_t\}$, are treated as parameters. There is no latent variable for Model 3, however. For the same reason, DIC$_7$ in Model 5 is quite different from that in Model 6. As argued earlier, this conceptual difficulty is due to the lack of the likelihood principle and is consistent with what has been documented in the literature (Spiegelhalter et al., 2002 and Berg et al., 2004). The most important finding from Table I is that RDIC does not suffer from the same difficulty as DIC$_7$. RDIC and P_D^* for Model 3 (and Model 5) are nearly identical to those for Model 4 (and Model 6). In terms of the computational cost, for Model 3, after the effective random observations are collected, RDIC takes about 3 minutes in a laptop with Inter Core i5-540M (2.53GHz). On the other hand, DIC$_1$ involves $\int \ln p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$ when computing P_D, which is approximated by $\frac{1}{J}\sum_{j=1}^{J}\ln p\left(\mathbf{y}|\boldsymbol{\theta}^{(j)}\right)$. This quantity is much more expensive to compute because it requires numerical evaluation of $\ln p\left(\mathbf{y}|\boldsymbol{\theta}^{(j)}\right)$ for J times. For Model 3, based on the 20,000 posterior random observations, one has to evaluate $\ln p\left(\mathbf{y}|\boldsymbol{\theta}^{(j)}\right)$ 20,000 times. It requires 11 hours and 4 minutes to compute DIC$_1$ using the same laptop. The computational relative efficiency of RDIC over DIC$_1$ is obvious and increases as the number of effective observations increases.

It is important to emphasize that, although our method is motivated from the case of objective priors, informative priors can be also used in our method. In a recent study, Tu and Zhou (2010) explored a general approach to forming informative priors based on economic objectives and found that the proposed informative priors outperform significantly the objective priors in terms of investment performance. RDIC can be used in conjunction with the informative prior specifications. In this case, P_D^* can be quite different from P.

4.2 Comparing high dimensional dynamic factor models

For many countries, there exists a rich array of macroeconomic time series and financial time series. To reduce the dimensionality and to extract the information from the large number of time series, factor analysis has been widely used in the empirical macroeconomic

Table 1: Model selection results for Fama-French three factor models

Model	M_1	M_2	M_3	M_4	M_5	M_6
Number of Parameters	100	125	100	100	125	125
P_D	100	125	100	1021	125	1046
DIC$_7$	-119842	-119880	-133088	-134777	-133202	-134897
P_D^*	100	125	100	100	126	126
RDIC	-119842	-119880	-133087	-133087	-133201	-133201

literature and in the empirical finance literature. For example, by extending the static factor models previously developed for cross-sectional data, Geweke (1977) proposed the dynamic factor model for time series data. Many empirical studies, such as Sargent and Sims (1977), Giannone et al. (2004), have reported evidence that a large fraction of the variance of many macroeconomic series can be explained by a small number of dynamic factors. Stock and Watson (1999) and Stock and Watson (2002) showed that dynamic factors extracted from a large number of predictors can be used to lead to improvement in predicting macroeconomic variables. Not surprisingly, high dimensional dynamic factor models have become a popular tool under a data rich environment for macroeconomists and policy makers. An excellent review on the dynamic factor models is given by Stock and Watson (2010).

Following Bernanke et al. (2005) (BBE hereafter), the present paper considers the following fundamental dynamic factor model:

$$\begin{aligned} Y_t &= F_t L' + \varepsilon_t', \\ F_t &= F_{t-1}\Phi' + \eta_t, \end{aligned}$$

where Y_t is a $1 \times N$ vector of time series variables, F_t a $1 \times K$ vector of unobserved latent factors which contains the information extracted from all the N time series variables, L an $N \times K$ factor loading matrix, Φ the $K \times K$ autoregressive parameter matrix of unobserved latent factors. It is assumed that $\varepsilon_t \sim N(0, \Sigma)$ and $\eta_t \sim N(0, Q)$. For the purpose of identification, Σ is assume to be diagonal and ε_t and η_t are assumed to be independent with each other. Following BBE (2005), we set the first $K \times K$ block in the loading matrix L to be the identity matrix.

In this dynamic factor model, the observed variable Y_t consists of a balanced panel of 120 monthly macroeconomic time series. These series are initially transformed to induce stationarity. The description of the series and the transformation is provided in BBE (2005). The sample period is from January 1959 to August 2001. Because the data are of high dimension, the analysis of the dynamic factor models via a frequentist method is not trivial; see the discussion in Stock and Watson (2011). In the literature, Bayesian inference via the MCMC techniques has been popular for analyzing the dynamic factor models; see Otrok and

Whiteman (1998), Kose et al. (2003), Kose et al. (2008), BBE (2005).

Following BBE (2005), we specify the following prior distribution:

$$\begin{aligned}\Sigma_{ii} &\sim Inverse-\Gamma\left(3,0.001\right), L_i \sim N\left(0,\Sigma_{ii}M_0^{-1}\right),\\ vec\left(\Phi\right)|Q &\sim N\left(0,Q\otimes\Omega_0\right), Q\sim Inverse-\Gamma\left(Q_0,K+2\right),\end{aligned}$$

where M_0 is a $K\times K$ identity matrix, L_i the ith ($i>K$) column of L. The diagonal elements of Q_0 are set to be the residual variances of the corresponding one lag univariate autoregressions, $\widehat{\sigma}_i^2$. The diagonal elements of Ω_0 are constructed so that the prior variance of parameter on the jth variable in the ith equation equals $\widehat{\sigma}_i^2/\widehat{\sigma}_j^2$.

In this example, we aim to determine the number of factors in the dynamic factor models using model selection criteria. In BBE (2005) model comparison is achieved by graphic methods. Our approach can be regarded as a formal statistical alternative to the graphic methods. It is well documented that the determination of number of factors in the setting of the dynamic factor models is important; see Stock and Watson (1999). As in the previous example, we use DIC$_7$ and RDIC to compare models with different numbers of factors, namely $K=1$, 2 and 3, which are denoted by M_1, M_2, M_3 respectively. Using the Gibbs sampler, we sample 22,000 random observations from the corresponding posterior distributions. We discard the first 2,000 observations and keep the following 20,000 as the effective samples from the posterior distribution of the parameters.

Based on the 20,000 samples, we compute DIC$_7$, RDIC, P_D, P_D^* for all three models. Table II reports the simple count of the number of parameters (including the latent variables), DIC$_7$, the P_D component of DIC$_7$, (i.e. when the data augmentation technique is used), the simple count of the number of parameters (excluding the latent variables), RDIC, and the P_D^* component of RDIC (i.e. when the data augmentation technique is not used). Several conclusions may be drawn from Table II. First, both DIC$_7$ and RDIC suggest that M_3 is the best model. Second, since some very informative priors have been used, neither P_D nor P_D^* is close to the actual number of parameters. While it is cheap to compute RDIC, it is much harder to compute DIC$_1$. This is because the observed-data likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ is not available in closed-form and Kalman filter is used to numerically calculate $p(\mathbf{y}|\boldsymbol{\theta})$ which involves the computation of $\frac{1}{J}\sum_{j=1}^{J}\ln p(\mathbf{y}|\theta^{(j)})$, for $J=20,000$. We have to run Kalman filter 20,000 times, which takes more than 4 hours to compute in Matlab. In a sharp contrast, it only took less than 80 seconds to compute RDIC. Obviously, the discrepancy in CPU time increases with J.

4.3 Comparing stochastic volatility models

Stochastic volatility (SV) models have been found very useful for pricing derivative securities. In the discrete time log-normal SV models, the logarithmic volatility is the state variable

Table 2: Model selection results for dynamic factor models

Model	M_1	M_2	M_3
Number of Parameters	752	1385	2019
P_D	350	965	1391
DIC$_7$	-135480	-149010	-155060
Number of Parameters	241	363	486
P_D^*	87	20	326
RDIC	-22452	-34868	-40420

which is often assumed to follow an AR(1) model. The basic log-normal SV model is of the form:

$$\begin{aligned} y_t &= \alpha + \exp(h_t/2)u_t,\ u_t \sim N(0,1), \\ h_t &= \mu + \phi(h_{t-1} - \mu) + v_t,\ v_t \sim N(0,\tau^2), \end{aligned}$$

where $t = 1, 2, \cdots, n$, y_t is the continuously compounded return, h_t the unobserved log-volatility, $h_0 = \mu$, and (u_t, v_t) independently normal variables for all t. In this paper, we denote this model by M_1.

To carry out Bayesian analysis of M_1, following Meyer and Yu (2000), the prior distributions are specified as follows:

$$\begin{aligned} \alpha &\sim N(0,100),\ \ \mu \sim N(0,100), \\ \phi &\sim Beta(1,1),\ \ 1/\tau^2 \sim \Gamma(0.001, 0.001). \end{aligned}$$

An alternative specification of M_1 is given by:

$$\begin{aligned} y_t &= \alpha + \sigma_t u_t,\ u_t \sim N(0,1), \\ \ln \sigma_t^2 &= \mu + \phi\left(\ln \sigma_{t-1}^2 - \mu\right) + \nu_t,\ v_t \sim N(0,\tau^2), \end{aligned}$$

which is denoted by M_2. Obviously, the only difference between M_2 and M_1 is that the latent variable in M_2 is the exponential transformation of that in M_1. If the same priors are used for the model parameters, $\boldsymbol{\theta} = (\alpha, \mu, \phi, \tau)$, the two models are identical to each other. Our goal here is to compare the two models using DIC$_7$ and RDIC. In both models, $p(\mathbf{y}|\boldsymbol{\theta})$ is not available in closed-form. Since the models are of a nonlinear non-Gaussian form, Kalman filter cannot be applied and DIC$_1$ is infeasible to compute.

The dataset consists of 1,822 daily returns of the Standard & Poor (S&P) 500 index, covering the period between January 3, 2005 and March 28, 2012. For M_1 and M_2, after a burn-in period of 10,000 iterations we save the next 20,000 iterations.

Table III reports DIC$_7$, RDIC, P_D, P_D^* for both models. To calculate RDIC, since the $\mathcal{Q}$ function does not have a closed-form expression, we employ the technique in Lemma 3.3 to

Table 3: Model selection results for stochastic volatility models

Model	M_1	M_2
P_D	102.94	89.67
DIC$_7$	5200.56	5183.12
P_D^*	3.62	3.78
RDIC	5296.20	5296.55

compute the second order derivative of the observed-data likelihood. To compute P_D^*, we use the Laplace approximation of Rue, Martino and Chopin (2009).

The following findings can be obtained from Table III. First, P_D in M_1 is 13 points more than that in M_2. Similarly, DIC$_7$ in M_1 is nearly 20 points more than that in M_2. These differences are very large and indicate that M_2 is a much better model than M_1 although the two modes are actually the same. Second, P_D^* in M_1 is nearly identical to that in M_2, which is about the same as $P = 4$, the actual number of parameters. Similarly, RDIC in M_1 is nearly identical to that in M_2. Given that M_1 and M_2 are two equivalent representations to each other, the empirical results from RDIC are more reasonable than those from DIC$_7$.

5 Conclusion

In this paper, we have proposed a robust deviance information criteria (RDIC) for comparing models with latent variables. Although latent variable models can be conveniently estimated in the Bayesian framework via MCMC if the data augmentation technique is used, we argue that data augmentation cannot be used in connection to DIC. This is because that the justification of DIC rests on the validity of the standard Bayesian asymptotic theory. With data augmentation, the number of parameters increases with the number of observations, making the likelihood nonregular. As a consequence, the standard Bayesian asymptotic theory does not hold. In addition, the use of the data augmentation makes DIC is very sensitive to transformations and distributional representations.

While in principle one can use the standard DIC (i.e. DIC$_1$) without resorting to the data augmentation technique, in practice this standard DIC is very difficult to use because the observed-data likelihood is not available in closed-form for most latent variable models and because the standard DIC$_1$ has to numerically evaluate the observed-data likelihood at each MCMC iteration. These two observations make the implementation of DIC$_1$ practically non-operational.

The problem is overcome by RDIC. RDIC is defined without augmenting the parameter space and hence can be justified by the standard Bayesian asymptotic theory. We then show that how the EM algorithm can facilitate the computation of RDIC in different contexts.

Since the latent variables are not counted as parameters in our approach, RDIC is robust to nonlinear transformations of the latent variables and distributional representations of the model specification. Asymptotic justification, computational tractability and robustness to transformation and specification are the three main advantages of the proposed approach. These advantages are illustrated using several popular models in economics and finance.

A Proof of Lemma 3.1

Using the Taylor-expansion on the log-posterior probability density function, we can show that

$$\begin{aligned}\ln p(\boldsymbol{\theta}|\mathbf{y}) &= \ln p(\hat{\boldsymbol{\theta}}_m|\mathbf{y}) + L_n^{(1)}(\hat{\boldsymbol{\theta}}_m)'(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m) + \frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)\\ &= \ln p(\hat{\boldsymbol{\theta}}_m|\mathbf{y}) + \frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m),\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ lies on the segment between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_m$. It follows that

$$p(\boldsymbol{\theta}|\mathbf{y}) = p(\hat{\boldsymbol{\theta}}_m|\mathbf{y}) \exp\left[\frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)\right].$$

Let $\boldsymbol{\omega} = \sqrt{n}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)$, $J(\boldsymbol{\theta}) = -\frac{1}{n}L_n^{(2)}(\boldsymbol{\theta})$, $c_n^* = \int \exp[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}]\mathrm{d}\boldsymbol{\omega}$, $c_n = \int \exp[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}]\mathrm{d}\boldsymbol{\omega}$. It can be shown that

$$p(\boldsymbol{\omega}|\mathbf{y}) \propto \exp\left[\frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)\right] = \exp\left\{-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right\}.$$

Then, we have

$$\begin{aligned}
P_n &= \int \left| p(\boldsymbol{\omega}|\mathbf{y}) - \frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\\
&= \int \left| \frac{1}{c_n^*}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] - \frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\\
&= \frac{1}{c_n}\int \left| \frac{c_n}{c_n^*}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] - \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\\
&= \frac{1}{c_n}\int \left| \frac{c_n - c_n^*}{c_n^*}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] + \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] - \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\\
&\leq \frac{1}{c_n}\left\{\int \left|\frac{c_n - c_n^*}{c_n^*}\right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega} + \int \left|\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] - \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\right\}\\
&\leq \frac{|c_n - c_n^*|}{c_n} + \frac{1}{c_n}\int \left|\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] - \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\\
&\leq \frac{2}{c_n}\int \left|\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right] - \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega}\\
&\leq \frac{2}{c_n}\int \left|\exp\left\{-\frac{1}{2}\boldsymbol{\omega}'\left[J(\tilde{\boldsymbol{\theta}}) - J(\hat{\boldsymbol{\theta}}_m)\right]\boldsymbol{\omega}\right\} - 1\right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega}.
\end{aligned}$$

By Assumption 3, for any $\epsilon > 0$, there exists some $\delta > 0$ such that when $\Omega = \{\boldsymbol{\omega} : ||\boldsymbol{\omega}|| < \sqrt{n}\delta\}$ we have $\boldsymbol{\theta} \in H(\hat{\boldsymbol{\theta}}_m, \delta)$ and $-A(\epsilon) \leq [J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m) - I_P] \leq A(\epsilon)$. By Hölder inequality, we have

$$
\begin{aligned}
&\lim_{n\to\infty} Q_n = \lim_{n\to\infty} \int \left| \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}}) - J(\hat{\boldsymbol{\theta}}_m) \right] \boldsymbol{\omega} \right\} - 1 \right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega} \\
= \;& \lim_{n\to\infty} \int_\Omega \left| \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}}) - J(\hat{\boldsymbol{\theta}}_m) \right] \boldsymbol{\omega} \right\} - 1 \right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega} \\
= \;& \lim_{n\to\infty} \int_\Omega \left| \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m) - I_P \right] J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right\} - 1 \right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega} \\
\leq \;& \left\{ \lim_{n\to\infty} \int_\Omega \left| \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m) - I_P \right] J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right\} - 1 \right|^2 \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega} \right\}^{1/2} \\
= \;& (D_1 - 2D_2 + D_3)^{1/2},
\end{aligned}
$$

where

$$
\begin{aligned}
D_1 &= \lim_{n\to\infty} \int_\Omega \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega}, \\
D_2 &= \lim_{n\to\infty} \int_\Omega \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m) - I_P \right] J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right\} \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega} \\
&= \lim_{n\to\infty} \int_\Omega \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega}, \\
D_3 &= \lim_{n\to\infty} \int_\Omega \exp\left\{ -\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m) - I_P \right] J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right\} \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega} \right] \mathrm{d}\boldsymbol{\omega}.
\end{aligned}
$$

It can be shown that $D_1 = (2\pi)^{P/2}|J(\hat{\boldsymbol{\theta}}_m)|^{-1/2}$. Following the proof of the posterior normality in Lemma 2.1 and Theorem 2.1 of Chen (1985), we have $D_2^- \leq D_2 \leq D_2^+, D_3^- \leq D_2 \leq D_3^+$ and

$$
\begin{aligned}
D_2^+ &= \left| J(\hat{\boldsymbol{\theta}}_m) \right|^{1/2} |I_P - A(\epsilon)|^{-1/2} \int_{||Z||<s_n} \exp\left[-\frac{1}{2}\boldsymbol{Z}'\boldsymbol{Z} \right] \mathrm{d}\boldsymbol{Z}, \\
D_2^- &= \left| J(\hat{\boldsymbol{\theta}}_m) \right|^{1/2} |I_P + A(\epsilon)|^{-1/2} \int_{||Z||<t_n} \exp\left[-\frac{1}{2}\boldsymbol{Z}'\boldsymbol{Z} \right] \mathrm{d}\boldsymbol{Z}, \\
D_3^+ &= \left| J(\hat{\boldsymbol{\theta}}_m) \right|^{1/2} |I_P - 2A(\epsilon)|^{-1/2} \int_{||Z||<s_n} \exp\left[-\frac{1}{2}\boldsymbol{Z}'\boldsymbol{Z} \right] \mathrm{d}\boldsymbol{Z}, \\
D_3^- &= \left| J(\hat{\boldsymbol{\theta}}_m) \right|^{1/2} |I_P + 2A(\epsilon)|^{-1/2} \int_{||Z||<t_n} \exp\left[-\frac{1}{2}\boldsymbol{Z}'\boldsymbol{Z} \right] \mathrm{d}\boldsymbol{Z},
\end{aligned}
$$

where $s_n = \delta(1 - e^*(\epsilon))^{1/2}/\sigma_n^*$ and $t_n = \delta(1 + e(\epsilon))^{1/2}/\sigma_n$, σ_n^2 and σ_n^{*2} is the largest and smallest eigenvalue of $\{nJ(\hat{\boldsymbol{\theta}}_m)\}^{-1}$, $e(\epsilon)$ and $e^*(\epsilon)$ is the largest and smallest eigenvalue of

$A(\epsilon)$. Under the regularity conditions, when $n \to \infty$, $s_n \to \infty$ and $t_n \to \infty$, if $\epsilon \to 0$, we get

$$\begin{aligned}
&\lim_{n\to\infty} |I_P \pm A(\epsilon)| = 1, \lim_{n\to\infty} |I_P \pm 2A(\epsilon)| = 1, \\
&\lim_{n\to\infty} \int_{||Z||<s_n} \exp\left[-\frac{1}{2}\boldsymbol{Z}'\boldsymbol{Z}\right] \mathrm{d}\boldsymbol{Z} = (2\pi)^{P/2}, \\
&\lim_{n\to\infty} \int_{||Z||<t_n} \exp\left[-\frac{1}{2}\boldsymbol{Z}'\boldsymbol{Z}\right] \mathrm{d}\boldsymbol{Z} = (2\pi)^{P/2}.
\end{aligned}$$

Then, we can show that $D_1 = D_2 = D_3 = (2\pi)^{P/2}|J(\hat{\boldsymbol{\theta}}_m)|^{-1/2}$ which impliesthat $\lim_{n\to\infty} Q_n = 0$ and that $\lim_{n\to\infty} P_n = 0$.

For $i, j = 1, 2, \cdots, P$, it can be shown that

$$\begin{aligned}
& \int \omega_i \left\{ p(\boldsymbol{\omega}|\mathbf{y}) - \frac{1}{c_n} \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right\} \mathrm{d}\boldsymbol{\omega} \\
\leq \; & \int |\omega_i| \left| p(\boldsymbol{\omega}|\mathbf{y}) - \frac{1}{c_n} \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega} \\
\leq \; & \frac{2}{c_n} \int |\omega_i| \left| \exp\left\{-\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}}) - J(\hat{\boldsymbol{\theta}}_m)\right] \boldsymbol{\omega}\right\} - 1 \right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right] \mathrm{d}\boldsymbol{\omega} \\
\leq \; & \frac{2}{c_n} \left\{ \lim_{n\to\infty} \int_\Omega |\omega_i|^2 \left| \exp\left\{ -\frac{\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m) - I_P\right] J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}}{2} \right\} - 1 \right|^2 \exp\left[-\frac{\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}}{2} \right] \mathrm{d}\boldsymbol{\omega} \right\}^{\frac{1}{2}} \\
= \; & \frac{2}{c_n}(ED_1 - 2ED_2 + ED_3)^{1/2},
\end{aligned}$$

$$\begin{aligned}
& \int \omega_i\omega_j \left\{ p(\boldsymbol{\omega}|\mathbf{y}) - \frac{1}{c_n} \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right\} \mathrm{d}\boldsymbol{\omega} \\
\leq \; & \int |\omega_i\omega_j| \left| p(\boldsymbol{\omega}|\mathbf{y}) - \frac{1}{c_n} \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right| \mathrm{d}\boldsymbol{\omega} \\
\leq \; & \frac{2}{c_n} \int |\omega_i\omega_j| \left| \exp\left\{-\frac{1}{2}\boldsymbol{\omega}' \left[J(\tilde{\boldsymbol{\theta}}) - J(\hat{\boldsymbol{\theta}}_m)\right] \boldsymbol{\omega}\right\} - 1 \right| \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right] \mathrm{d}\boldsymbol{\omega} \\
= \; & \frac{2}{c_n}(VD_1 - 2VD_2 + VD_3)^{1/2},
\end{aligned}$$

where

$$
\begin{aligned}
ED_1 &= \lim_{n\to\infty}\int_{\Omega}\omega_i^2 \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
ED_2 &= \lim_{n\to\infty}\int_{\Omega}\omega_i^2 \exp\left\{-\frac{1}{2}\boldsymbol{\omega}'\left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m)-I_P\right]J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right\}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
&= \lim_{n\to\infty}\int_{\Omega}\omega_i^2 \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
ED_3 &= \lim_{n\to\infty}\int_{\Omega}\omega_i^2 \exp\left\{-\boldsymbol{\omega}'\left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m)-I_P\right]J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right\}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
VD_1 &= \lim_{n\to\infty}\int_{\Omega}\omega_i^2\omega_j^2 \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
VD_2 &= \lim_{n\to\infty}\int_{\Omega}\omega_i^2\omega_j^2 \exp\left\{-\frac{1}{2}\boldsymbol{\omega}'\left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m)-I_P\right]J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right\}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
&= \lim_{n\to\infty}\int_{\Omega}\omega_i^2\omega_j^2 \exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\tilde{\boldsymbol{\theta}})\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega},\\
VD_3 &= \lim_{n\to\infty}\int_{\Omega}\omega_i^2\omega_j^2 \exp\left\{-\boldsymbol{\omega}'\left[J(\tilde{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}}_m)-I_P\right]J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right\}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\mathrm{d}\boldsymbol{\omega}.
\end{aligned}
$$

For the same argument, we can prove that $ED_1 = ED_2 = ED_3$ and $VD_1 = VD_2 = VD_3$. Hence, we have

$$
\begin{aligned}
&\int \omega_i\left\{p(\boldsymbol{\omega}|\mathbf{y})-\frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right\}\mathrm{d}\boldsymbol{\omega}\\
\leq &\int |\omega_i|\left|p(\boldsymbol{\omega}|\mathbf{y})-\frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right|\mathrm{d}\boldsymbol{\omega}\to 0,\\
&\int \omega_i\omega_j\left\{p(\boldsymbol{\omega}|\mathbf{y})-\frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right\}\mathrm{d}\boldsymbol{\omega}\\
\leq &\int |\omega_i\omega_j|\left|p(\boldsymbol{\omega}|\mathbf{y})-\frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right|\mathrm{d}\boldsymbol{\omega}\to 0.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\int \omega_i\left\{\frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right\}\mathrm{d}\boldsymbol{\omega}=0,\\
&\int \omega_i\omega_j\left\{\frac{1}{c_n}\exp\left[-\frac{1}{2}\boldsymbol{\omega}' J(\hat{\boldsymbol{\theta}}_m)\boldsymbol{\omega}\right]\right\}\mathrm{d}\boldsymbol{\omega}=J_{ij}^{-1}(\hat{\boldsymbol{\theta}}_m),
\end{aligned}
$$

where $J_{ij}^{-1}(\hat{\boldsymbol{\theta}}_m)$ is the $(i,j)^{th}$ element of $J^{-1}(\hat{\boldsymbol{\theta}}_m)$. Hence, we have $E(\boldsymbol{\omega}|\mathbf{y}) = 0 + o(1)$ and $E(\boldsymbol{\omega}\boldsymbol{\omega}'|\mathbf{y}) = J^{-1}(\hat{\boldsymbol{\theta}}_m) + o(1)$ which imply that

$$
E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)|\mathbf{y}] = o(n^{-1/2}), E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_m)'|\mathbf{y}] = -L_n^{-(2)}(\hat{\boldsymbol{\theta}}_m) + o(n^{-1}).
$$

B Proof of Theorem 3.1

Under Assumption 6, when $n\to\infty$, we have

$$
\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = L_n^{(1)}(\boldsymbol{\theta}), -I(\boldsymbol{\theta}) = \frac{\partial^2 \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\boldsymbol{\theta}'} = L_n^{(2)}(\boldsymbol{\theta}),
$$

and the ML estimator $\hat{\boldsymbol{\theta}}$ is asymptotically equivalent to the posterior mode $\hat{\boldsymbol{\theta}}_m$. According to Lemma 3.1, we can show that $\bar{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{y}) = \hat{\boldsymbol{\theta}}_m + o(n^{-1/2})$. Hence, there exists an integer N, when $n > N$, $\bar{\boldsymbol{\theta}} \in H(\hat{\boldsymbol{\theta}}, \delta)$. We can then find some δ_1 with $0 < \delta_1 < ||\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}||$ so that $H(\bar{\boldsymbol{\theta}}, \delta_1) \subset H(\hat{\boldsymbol{\theta}}, \delta)$.

Applying the Taylor expansion to the log-likelihood function, we get

$$\ln p(\mathbf{y}|\boldsymbol{\theta}) = \ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) + L_n^{(1)}(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}),$$

where $\tilde{\boldsymbol{\theta}}$ is some $\boldsymbol{\theta}$ lying on the segment between $\boldsymbol{\theta}$ and $\bar{\boldsymbol{\theta}}$. When $n \to \infty$, $H(\bar{\boldsymbol{\theta}}, \delta_1) \subset H(\hat{\boldsymbol{\theta}}, \delta)$ and $\tilde{\boldsymbol{\theta}} \in H(\bar{\boldsymbol{\theta}}, \delta_1) \subset H(\hat{\boldsymbol{\theta}}, \delta)$. Hence, for any $\epsilon > 0$, there exists an integer N such that for any $n > N$, $L_n^{(2)}(\tilde{\boldsymbol{\theta}})$ satisfies

$$[\boldsymbol{I_P} - A(\epsilon)]\,[-L_n^{(2)}(\hat{\boldsymbol{\theta}})] \leq -L_n^{(2)}(\tilde{\boldsymbol{\theta}}) = \left[L_n^{(2)}(\tilde{\boldsymbol{\theta}})L_n^{-(2)}(\hat{\boldsymbol{\theta}})\right]\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}})\right] \leq [\boldsymbol{I_P} + A(\epsilon)]\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}})\right].$$

That is,

$$[\boldsymbol{I_P} - A(\epsilon)]\, I(\hat{\boldsymbol{\theta}}) \leq I(\tilde{\boldsymbol{\theta}}) = \left[I(\tilde{\boldsymbol{\theta}})I^{-1}(\hat{\boldsymbol{\theta}})\right] I(\hat{\boldsymbol{\theta}}) \leq [\boldsymbol{I_P} + A(\epsilon)]\, I(\hat{\boldsymbol{\theta}}).$$

Hence, under the regularity conditions, when $n \to \infty$, we have

$$\begin{aligned} P_D = -2\int_{\Theta}&\left[\ln p(\mathbf{y}|\boldsymbol{\theta}) - \ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})\right] p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \\ = \quad & -2\int_{\Theta}\left[L_n^{(1)}(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})\right] p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \\ = \quad & \int_{\Theta} -(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' L_n^{(2)}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \\ = \quad & \int_{H(\hat{\theta},\delta)} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' I(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} \\ = \quad & \int_{H(\hat{\theta},\delta)} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' I(\tilde{\boldsymbol{\theta}})I^{-1}(\hat{\boldsymbol{\theta}})I(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}, \end{aligned}$$

which is bounded above by

$$P_D^+ = \int_{H(\hat{\theta},\delta)} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' \left[\boldsymbol{I_P} + A(\epsilon)\right] I(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} = \mathbf{tr}\left\{\left[\boldsymbol{I_P} + A(\epsilon)\right] I(\hat{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\right\},$$

and below by

$$P_D^- = \int_{H(\hat{\theta},\delta)} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' \left[\boldsymbol{I_P} - A(\epsilon)\right] I(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta} = \mathbf{tr}\left\{\left[\boldsymbol{I_P} - A(\epsilon)\right] I(\hat{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\right\}.$$

Under the regularity conditions, for $\epsilon \to 0$, we have $\lim_{n\to\infty} P_D = \mathbf{tr}\{-L_n^{(2)}(\hat{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\}$ or $P_D = \mathbf{tr}\{I(\hat{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\} + o(1)$.

Conditional on the observed data $\mathbf{y}$, note that $L_n^{(2)}(\bar{\boldsymbol{\theta}})/n = O(1)$, $L_n^{(2)}(\hat{\boldsymbol{\theta}})/n = O(1)$, we get $L_n^{(2)}(\bar{\boldsymbol{\theta}})/n = L_n^{(2)}(\hat{\boldsymbol{\theta}})/n + o(1)$. According to Lemma 3.1, we have $nV(\hat{\boldsymbol{\theta}}) = n[V(\bar{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}} -$

$\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})']=nV(\bar{\boldsymbol{\theta}})+no(n^{-1})=nV(\bar{\boldsymbol{\theta}})+o(1)$ and $nV(\hat{\boldsymbol{\theta}})=[-L_n^{(2)}(\hat{\boldsymbol{\theta}})/n]^{-1}+o(1)=O(1)$ so that $nV(\bar{\boldsymbol{\theta}})=O(1)$. Thus, we have

$$\begin{aligned}
P_D &= \mathbf{tr}\left\{I(\hat{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\right\}+o(1)=\mathbf{tr}\left\{[I(\hat{\boldsymbol{\theta}})/n][nV(\bar{\boldsymbol{\theta}})]\right\}+o(1)\\
&= \mathbf{tr}\left\{[I(\bar{\boldsymbol{\theta}})/n][nV(\bar{\boldsymbol{\theta}})]\right\}+o(1)O(1)+o(1)\\
&= \mathbf{tr}\left\{[I(\bar{\boldsymbol{\theta}})/n][nV(\bar{\boldsymbol{\theta}})]\right\}+o(1)=\mathbf{tr}\left\{I(\bar{\boldsymbol{\theta}})V(\bar{\boldsymbol{\theta}})\right\}+o(1)=P_D^*+o(1).
\end{aligned}$$

Similarly, DIC_1 =RDIC+$o(1)$ and the theorem is proved.

C Proof of Theorem 3.2

According to Spiegelhalter et al. (2002), we get

$$\begin{aligned}
& -2E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\left\{\ln p(\mathbf{y}_{rep}|\bar{\boldsymbol{\theta}})-\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})\right\}\\
= & -2E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\{\ln p(\mathbf{y}_{rep}|\bar{\boldsymbol{\theta}})-\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})+\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})-\ln p(\mathbf{y}|\boldsymbol{\theta})+\ln p(\mathbf{y}|\boldsymbol{\theta})-\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})\}\\
= & D_1+D_2+D_3,
\end{aligned}$$

where

$$\begin{aligned}
&D_1=-2E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\left\{\ln p(\mathbf{y}_{rep}|\bar{\boldsymbol{\theta}})-\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})\right\},\\
&D_2=-2E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\{\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})-\ln p(\mathbf{y}|\boldsymbol{\theta})\},\\
&D_3=-2E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\{\ln p(\mathbf{y}|\boldsymbol{\theta})-\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})\}=-2\{\ln p(\mathbf{y}|\boldsymbol{\theta})-\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})\}.
\end{aligned}$$

Note that $E_{\mathbf{y}}E_{\boldsymbol{\theta}|\mathbf{y}}D_2=0$. From Theorem 3.1, we have $E_{\boldsymbol{\theta}|\mathbf{y}}D_3=P_D=P_D^*+o(1)$. For D_1, applying the Taylor expansion, conditional on the observed data $\mathbf{y}$, we get

$$\begin{aligned}
E_{\boldsymbol{\theta}|\mathbf{y}}D_1 &= -2E_{\boldsymbol{\theta}|\mathbf{y}}E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\left\{(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\frac{\partial\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})}{\partial\theta}+\frac{1}{2}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta})'\frac{\partial^2\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta})\right\}\\
&= -E_{\boldsymbol{\theta}|\mathbf{y}}\left\{(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta})'E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\frac{\partial^2\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta})\right\},
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ lies on the segment between $\boldsymbol{\theta}$ and $\bar{\boldsymbol{\theta}}$. When $n\to\infty$, for any $\delta>0$, $\tilde{\boldsymbol{\theta}}\in H(\hat{\boldsymbol{\theta}},\delta)$. Under Assumption 7, we have

$$\begin{aligned}
&\frac{1}{n}E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\left[\frac{\partial^2\ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}\right]=\frac{1}{n}E_{\mathbf{y}_{rep}|\boldsymbol{\theta}}\left[\frac{\partial^2\ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}+o_p(n)\right]\\
= &\frac{1}{n}\left[\frac{\partial^2\ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta\partial\theta'}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}+o(n)\right]=-\frac{1}{n}I(\tilde{\boldsymbol{\theta}})+o(1).
\end{aligned}$$

Following Lemma 3.1 and Theorem 3.1, as $n \to \infty$, we have

$$
\begin{aligned}
E_{\boldsymbol{\theta}|\mathbf{y}} D_1 &= E_{\boldsymbol{\theta}|\mathbf{y}} \left\{ (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left[-E_{\mathbf{y}_{rep}|\boldsymbol{\theta}} \frac{\partial^2 \ln p(\mathbf{y}_{rep}|\boldsymbol{\theta})}{\partial\theta\partial\theta'} |_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} \right] (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\} \\
&= E_{\boldsymbol{\theta}|\mathbf{y}} \left\{ (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})' I(\tilde{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\} + E_{\boldsymbol{\theta}|\mathbf{y}} \{ (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})' o(n) \} \\
&= E_{\boldsymbol{\theta}|\mathbf{y}} \left\{ (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})' I(\tilde{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\} + o(1) \\
&= E_{\boldsymbol{\theta}|\mathbf{y}} \left\{ (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})' I(\hat{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right\} + o(1) \\
&= \mathbf{tr} \left\{ I(\hat{\boldsymbol{\theta}}) V(\bar{\boldsymbol{\theta}}) \right\} + o(1) = P_D^* + o(1).
\end{aligned}
$$

Hence, under Assumptions 1-6, based on Theorem 3.1, we can further obtain

$$
\begin{aligned}
& -2E_{\mathbf{y}} E_{\boldsymbol{\theta}|\mathbf{y}} E_{\mathbf{y}_{rep}|\boldsymbol{\theta}} \{ \ln p(\mathbf{y}_{rep}|\bar{\boldsymbol{\theta}}) \} \\
= & -2E_{\mathbf{y}} E_{\boldsymbol{\theta}|\mathbf{y}} E_{\mathbf{y}_{rep}|\boldsymbol{\theta}} \{ \ln p(\mathbf{y}_{rep}|\bar{\boldsymbol{\theta}}) - \ln p(\mathbf{y}|\bar{\boldsymbol{\theta}}) \} - 2E_y [\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})] \\
= & -2E_{\mathbf{y}} [\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})] + E_{\mathbf{y}} E_{\boldsymbol{\theta}|\mathbf{y}} [D_1 + D_2 + D_3] \\
= & -2E_{\mathbf{y}} [\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})] + E_{\mathbf{y}} [P_D + P_D^*] + o(1) \\
= & -2E_{\mathbf{y}} [\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})] + 2P_D + o(1) = E_{\mathbf{y}} [DIC_1] + o(1) \\
= & -2E_{\mathbf{y}} [\ln p(\mathbf{y}|\bar{\boldsymbol{\theta}})] + 2P_D^* + o(1) = E_{\mathbf{y}} [RDIC] + o(1).
\end{aligned}
$$

D The derivation of RDIC for the asset pricing models

It has been noted in Kan and Zhou (2003) that under the multivariate t specification, a direct numerical optimization of the observed data likelihood function is very difficult. By using normal-gamma scale-mixture distribution to replace the t distribution, the powerful EM algorithm can be used to obtain the $\mathcal{Q}$ function. Since Models 1-5 are nested by Model 6, we only need to derive the first and second derivatives for Model 6.

Let $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_n\}$, $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_n\}$, $\boldsymbol{\omega} = \{\omega_1, \omega_2, \cdots, \omega_n\}$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$. The density function of the multivariate t is given by

$$
f(\boldsymbol{\epsilon}_t) = \frac{\Gamma(\frac{\nu+N}{2})}{(\pi\nu)^{\frac{2}{N}} \Gamma(\frac{\nu}{2}) |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left\{ 1 + \frac{\boldsymbol{\epsilon}_t' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}_t}{\nu} \right\}^{-\frac{\nu+N}{2}}.
$$

Hence, the observed data log-likelihood function, $L_o(\mathbf{R}|\boldsymbol{\theta})$, is:

$$
L_o(\mathbf{R}|\boldsymbol{\theta}) = C(\nu) - \frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{\nu + N}{2} \sum_{t=1}^{n} \log \left[\nu + \varphi \left(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta} \right) \right], \tag{18}
$$

where

$$
\begin{aligned}
& C(\nu) = -\frac{nN}{2} \log(\pi\nu) + n \left[\ln \Gamma \left(\frac{\nu + N}{2} \right) - \ln \Gamma \left(\frac{\nu}{2} \right) \right] + \frac{n(\nu + N) \ln \nu}{2}, \\
& \varphi \left(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta} \right) = \left(R_t - \alpha - \boldsymbol{\beta} \boldsymbol{F}_t \right)' \boldsymbol{\Sigma}^{-1} \left(R_t - \alpha - \boldsymbol{\beta} \boldsymbol{F}_t \right).
\end{aligned}
$$

Based on the normal-gamma mixture representation for the multivariate t distribution, the complete log-likelihood, $L_c(\mathbf{R}, \boldsymbol{\omega}|\boldsymbol{\theta})$, can be expressed as

$$
\begin{aligned}
&-\frac{1}{2}nN\ln(2\pi) + \frac{N}{2}\sum_{t=1}^{n}\ln\omega_t - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{n}\omega_t\varphi\left(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta}\right) \\
&-n\ln\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\ln\left(\frac{\nu}{2}\right) + \frac{\nu}{2}\sum_{t=1}^{n}(\ln\omega_t - \omega_t) - \sum_{t=1}^{n}\ln\omega_t.
\end{aligned}
$$

Thus, the posterior expectation of ω_t is

$$
\omega_t|\mathbf{y} \sim \Gamma\left[\frac{\nu+N}{2}, \frac{\nu+\varphi\left(\mathbf{R}_t, \mathbf{F}_t\boldsymbol{\theta}\right)}{2}\right].
$$

According to McLachlan and Krishnan (2008), it can be shown that

$$
\begin{aligned}
E\left(\boldsymbol{\omega}_t|\boldsymbol{\theta}, \mathbf{R}_t\right) &= \frac{\nu+N}{\nu+\varphi\left(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta}\right)}, \\
E\left(\ln\boldsymbol{\omega}_t|\boldsymbol{\theta}, \mathbf{R}_t\right) &= \ln E\left(\boldsymbol{\omega}_t|\boldsymbol{\theta}, \mathbf{R}_t\right) + \psi\left(\frac{\nu+N}{2}\right) - \ln\left(\frac{\nu+N}{2}\right),
\end{aligned}
$$

where $\psi(x)$ is the Digamma function, $\partial\Gamma(x)/\partial x/\Gamma(x)$. Hence, we get

$$
\begin{aligned}
&\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = \int L_c(\mathbf{R}, \boldsymbol{\omega}|\boldsymbol{\theta})p(\boldsymbol{\omega}|\mathbf{R}, \boldsymbol{\theta}^*)\mathrm{d}\boldsymbol{\omega} \\
= \quad &-\frac{1}{2}nK\ln(2\pi) + \frac{N}{2}\sum_{t=1}^{n}E(\ln\omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{t=1}^{n}E(\omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*)\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta}) \\
&-n\ln\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\ln\left(\frac{\nu}{2}\right) + \frac{\nu}{2}\sum_{t=1}^{n}E(\ln\omega_t - \omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*) - \sum_{t=1}^{n}E(\ln\omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*).
\end{aligned}
$$

For the asset price models considered in this paper, we obtain the second derivatives:

$$
\begin{aligned}
\frac{\partial\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} &= \frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} - \frac{1}{2}\sum_{t=1}^{n}E(\omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*)\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{'}} \\
\frac{\partial\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{*'}} &= -\frac{1}{2}\sum_{t=1}^{n}\frac{\partial\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial E(\omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}^{*'}} \\
= \quad \frac{1}{2}\sum_{t=1}^{n}&\frac{1}{\nu+\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta}^*)}E(\omega_t|\mathbf{R}_t, \boldsymbol{\theta}^*)\frac{\partial\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}^{*'}}
\end{aligned}
$$

For $i, j = 1, 2, \cdots, N$, letting $\phi_i = \sigma_{ii}^{-1}$, we get

$$\begin{aligned}
&\frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} = 0, \frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}^{'}} = 0, \frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\alpha}\partial\phi_i} = 0,\\
&\frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{'}} = 0, \frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\boldsymbol{\beta}\partial\phi_i} = 0, \frac{\partial^2(-\frac{n}{2}\ln|\boldsymbol{\Sigma}|)}{\partial\phi_i^2} = -\frac{n}{2\phi_i^2},\\
&\frac{\partial\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i} = -2\phi_i(R_{it} - \alpha_i - \boldsymbol{\beta}_i^{'}\mathbf{F}_t),\\
&\frac{\partial\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i} = -2\phi_i(R_{it} - \alpha_i - \boldsymbol{\beta}_i^{'}\mathbf{F}_t)\mathbf{F}_t,\\
&\frac{\partial\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\phi_i} = (R_{it} - \alpha_i - \boldsymbol{\beta}_i^{'}\mathbf{F}_t)^2,\\
&\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i^2} = 2\phi_i, \frac{\partial\varphi^2(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i\partial\alpha_j} = 0, i \neq j,\\
&\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i\partial\boldsymbol{\beta}_i} = 2\phi_i\boldsymbol{F}_t, \frac{\partial\varphi^2(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i\partial\boldsymbol{\beta}_j} = 0, i \neq j,\\
&\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i\partial\phi_i} = -2(R_{it} - \alpha_i - \boldsymbol{\beta}_i^{'}\boldsymbol{F}_t), \frac{\partial\varphi^2(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\alpha_i\partial\phi_j} = 0, i \neq j,\\
&\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_i^{'}} = 2\phi_i\boldsymbol{F}_t\boldsymbol{F}_t^{'}, \frac{\partial\varphi^2(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i\partial\boldsymbol{\beta}_j} = 0, i \neq j,\\
&\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i\partial\phi_i} = -2(R_{it} - \alpha_i - \boldsymbol{\beta}_i^{'}\boldsymbol{F}_t)\boldsymbol{F}_t, \frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i\partial\phi_j} = 0, i \neq j,\\
&\frac{\partial^2\varphi(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\phi_i^2} = 0, \frac{\partial\varphi^2(\boldsymbol{R}_t, \boldsymbol{F}_t, \boldsymbol{\theta})}{\partial\phi_i\partial\phi_j} = 0, i \neq j.
\end{aligned}$$

E The derivation of RDIC for the dynamic factor models

The complete-data log-likelihood function is:

$$\begin{aligned}
\ln f(Y, F|L, \Sigma, \Phi, Q) \quad = \quad & -\frac{(K+N)T - K}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\mathbf{tr}\left[\Sigma^{-1}\left(Y - FL'\right)^{'}\left(Y - FL'\right)\right]\\
& -\frac{T-1}{2}\ln|Q| - \frac{1}{2}\mathbf{tr}\left[Q^{-1}\left(F_{+1} - F_{-1}\Phi'\right)^{'}\left(F_{+1} - F_{-1}\Phi'\right)\right],
\end{aligned}$$

where $Y = [Y_1', Y_2', ..., Y_T']'$, $F = [F_1', F_2', ..., F_T']'$, $F_{+1} = [F_2', F_3', ..., F_T']'$, $F_{-1} = \left[F_1', F_2', ..., F_{T-1}'\right]'$. Denote this function by $\varphi(L, \Sigma, \Phi, Q)$, In this appendix, we derive the first and second derivative of the complete-data log-likelihood function. The matrix differentiation used here follows the rules discussed in Magnus and Neudecker (1999).

The first order derivatives of $\varphi(L, \Sigma, \Phi, Q)$:

Whenever there is no confusion, we denote $\varphi(L, \Sigma, \Phi, Q)$ simply by φ. The differential of

$\varphi(L, \Sigma, \Phi, Q)$ with respect to L is

$$\begin{aligned}
d_L(\varphi) &= d\left(-\frac{1}{2}\mathbf{tr}\left[\Sigma^{-1}\left(Y-FL'\right)^{'}\left(Y-FL'\right)\right]\right) \\
&= -\frac{1}{2}\mathbf{tr}\left\{-\Sigma^{-1}\left(dL\right)F^{'}\left(Y-FL'\right)+\Sigma^{-1}\left(Y-FL\right)^{'}\left(-F\left(dL\right)'\right)\right\} \\
&= \frac{1}{2}\mathbf{tr}\left\{\Sigma^{-1}dLF^{'}\left(Y-FL'\right)+\Sigma^{-1}\left(Y-FL'\right)^{'}F\left(dL\right)'\right\} \\
&= \frac{1}{2}\mathbf{tr}\left\{F^{'}\left(Y-FL'\right)\Sigma^{-1}dL+dLF'\Sigma^{-1}\left(Y-FL'\right)\left(\Sigma^{-1}\right)'\right\} \\
&= \frac{1}{2}\mathbf{tr}\left\{F^{'}\left(Y-FL'\right)\left(\left(\Sigma^{-1}\right)^{'}+\Sigma^{-1}\right)dL\right\} \\
&= \mathbf{tr}\left(\widetilde{c}dL\right),
\end{aligned}$$

where

$$\widetilde{c} = \frac{1}{2}F^{'}\left(Y-FL'\right)\left(\left(\Sigma^{-1}\right)^{'}+\Sigma^{-1}\right).$$

Taking *vec* both sides, we get

$$d\left(vec\left(-\frac{1}{2}\mathbf{tr}\left[\Sigma^{-1}(Y-FL')'(Y-FL)\right]\right)\right) = d(vec(\varphi)) = \left(vec(\widetilde{c})'\right)'d(vec(L)).$$

The first derivative of $\varphi\left(L, \Sigma, \Phi, Q\right)$ is

$$D_L\left(\varphi\right) = \left(vec\left(\left[\frac{1}{2}F^{'}\left(Y-FL'\right)\left(\left(\Sigma^{-1}\right)^{'}+\Sigma^{-1}\right)\right]^{'}\right)\right)^{'}.$$

Similarly, we have

$$D_\Sigma\left(\varphi\right) = \left(vec\left(-\frac{T}{2}\Sigma^{-1}+\frac{1}{2}\Sigma^{-1}\left(Y-FL'\right)^{'}\left(Y-FL'\right)\Sigma^{-1}\right)^{'}\right)^{'},$$

$$D_\Phi\left(\varphi\right) = \left(vec\left(\left[\frac{1}{2}F_{-1}'\left(F_{+1}-F_{-1}\Phi'\right)\left(\left(Q^{-1}\right)^{'}+Q^{-1}\right)\right]^{'}\right)\right)^{'},$$

$$D_Q\left(\varphi\right) = \left(vec\left(-\frac{T-1}{2}Q^{-1}+\frac{1}{2}Q^{-1}\left(F_{+1}-F_{-1}\Phi'\right)^{'}\left(F_{+1}-F_{-1}\Phi'\right)Q^{-1}\right)^{'}\right)^{'}.$$

The second order derivatives of $\varphi\left(L, \Sigma, \Phi, Q\right)$:

The first order derivative of $\widetilde{c}$ is

$$d\widetilde{c} = d\left(\frac{1}{2}F^{'}\left(Y-FL'\right)\left(\left(\Sigma^{-1}\right)^{'}+\Sigma^{-1}\right)\right) = -\frac{1}{2}F^{'}F\left(dL\right)^{'}\left(\left(\Sigma^{-1}\right)^{'}+\Sigma^{-1}\right).$$

And the second order derivative is

$$\begin{aligned}
d_L^2\varphi &= \mathbf{tr}\left(d\widetilde{c} * dL\right) \\
&= \mathbf{tr}\left(-\frac{1}{2}F^{'}F\left(dL\right)^{'}\left(\left(\Sigma^{-1}\right)^{'}+\Sigma^{-1}\right)dL\right).
\end{aligned}$$

Then, we have,

$$D_{L,L}(\varphi) = -\frac{1}{2}\left(F^{'}F \otimes \left(\left(\Sigma^{-1}\right)^{'} + \Sigma^{-1}\right)\right),$$

$$\begin{aligned}
H &= G(T) = T^{'}, \quad T = S(\Sigma) = \frac{1}{2}F^{'}\left(Y - FL'\right)\left(\left(\Sigma^{-1}\right)^{'} + \Sigma^{-1}\right), \\
D(G(T)) &= K_{KN}, \\
D(S(\Sigma)) &= I_N \otimes \left(F^{'}\left(Y - FL'\right)\right) \cdot \left(-\frac{1}{2}\left(K_{NN} + I_{NN}\right)\right) \cdot \left(\left(\Sigma^{-1}\right)^{'} \otimes \Sigma^{-1}\right), \\
DH(\Sigma) &= (DG(T))(DS(\Sigma)),
\end{aligned}$$

where K_{KN} is the commutation matrix for a matrix with K rows and N columns. Thus, we have

$$\begin{aligned}
D_{L,\Sigma}(\varphi) &= \frac{\partial D_L(\varphi)}{\left(\partial vec\Sigma\right)^{'}} = (DG(T))(DS(\Sigma)) \\
&= K_{KN} \cdot I_N \otimes \left(F^{'}\left(Y - FL'\right)\right) \cdot \left(-\frac{1}{2}\left(K_{NN} + I_{NN}\right)\right) \cdot \left(\left(\Sigma^{-1}\right)^{'} \otimes \Sigma^{-1}\right),
\end{aligned}$$

$$\begin{aligned}
D_{L,\Phi}(\varphi) &= 0, \\
D_{L,Q}(\varphi) &= 0,
\end{aligned}$$

$$D_{\Sigma,\Sigma}(\varphi) = K_{NN} \cdot \left(\begin{array}{c} \frac{T}{2} \cdot \frac{1}{2}\left(\left(\Sigma^{-1}\right)^{'} \otimes \Sigma^{-1} + \left(\Sigma^{-1}\right)^{'} \otimes \Sigma^{-1}\right) \\ -\frac{1}{2}\left(\begin{array}{c} \left(\Sigma^{-1}\left(Y - FL'\right)^{'}\left(Y - FL'\right)\Sigma^{-1}\right)^{'} \otimes \Sigma^{-1} \\ + \left(\Sigma^{-1}\right)^{'} \otimes \left(\Sigma^{-1}\left(Y - FL'\right)^{'}\left(Y - FL'\right)\Sigma^{-1}\right) \end{array}\right) \end{array}\right),$$

$$\begin{aligned}
D_{\Sigma,\Phi}(\varphi) &= 0, \\
D_{\Sigma,Q}(\varphi) &= 0,
\end{aligned}$$

$$\begin{aligned}
& D_{\Phi,Q}(\varphi) \\
= \; & K_{KK} \cdot \left(I_K \otimes F_{-1}^{\prime}\left(F_{+1} - F_{-1}\Phi'\right)\right) \cdot \left(-\frac{1}{2}\left(K_{KK} + I_{KK}\right)\right) \cdot \left(\left(Q^{-1}\right)^{'} \otimes Q^{-1}\right),
\end{aligned}$$

$$D_{\Phi,\Phi}(\varphi) = -\frac{1}{2}\left(F_{-1}^{'}F_{-1} \otimes \left(\left(Q^{-1}\right)^{'} + Q^{-1}\right)\right),$$

$$D_{Q,Q}(\varphi) = K_{KK} \cdot \left(\begin{array}{c} \frac{T-1}{2} \cdot \frac{1}{2}\left(\left(Q^{-1}\right)^{'} \otimes Q^{-1} + \left(Q^{-1}\right)^{'} \otimes Q^{-1}\right) \\ -\frac{1}{2}\left(\begin{array}{c} \left(Q^{-1}\left(F_{+1} - F_{-1}\Phi'\right)^{'}\left(F_{+1} - F_{-1}\Phi'\right)Q^{-1}\right)^{'} \otimes Q^{-1} \\ + \left(Q^{-1}\right)^{'} \otimes \left(\Sigma^{-1}\left(F_{+1} - F_{-1}\Phi'\right)^{'}\left(F_{+1} - F_{-1}\Phi'\right)Q^{-1}\right) \end{array}\right) \end{array}\right).$$

The special structure of parameter matrix:

Let L, Σ, Φ, Q have some special structures. In particular, let

$$L^* = vec\left(\overline{L}\right),$$

where $\overline{L}$ is the last $(N-K)\times K$ block of L, and

$$\Sigma^* = diag\,(\Sigma)\,,\ \Phi^* = vec\,(\Phi)\,,\ Q^* = vech\,(Q)\,.$$

The first order derivatives are as follows:

$$\begin{aligned}
D_{L^*}(\varphi) &= D_L(\varphi)\cdot D_{L^*}(L(L^*)) = D_L(\varphi)\cdot \dot{I}_{L^*},\\
D_{\Sigma^*}(\varphi) &= D_\Sigma(\varphi)\cdot D_{\Sigma^*}(\Sigma(\Sigma^*)) = D_\Sigma(\varphi)\cdot \dot{I}_{\Sigma^*},\\
D_{\Phi^*}(\varphi) &= D_\Phi(\varphi)\cdot \dot{I}_{\Phi^*},\\
D_{Q^*}(\varphi) &= D_Q(\varphi)\cdot \dot{I}_{Q^*}.
\end{aligned}$$

The second order derivatives are as follows:

$$\begin{aligned}
D_{L^*,L^*}(\varphi) &= D_{L^*}(D_{L^*}(\varphi)) = D_{L^*}\left(D_L(\varphi)\cdot \dot{I}_{L^*}\right)\\
&= \left(\dot{I}'_{L^*}\otimes I_1\right)\cdot D_{L^*}(D_L(\varphi))\\
&= \left(\dot{I}'_{L^*}\otimes I_1\right)\cdot D_{L,L}(\varphi)\cdot \dot{I}_{L^*},\\
D_{L^*,\Sigma^*}(\varphi) &= D_{\Sigma^*}(D_{L^*}(\varphi)) = D_{\Sigma^*}\left(D_L(\varphi)\cdot \dot{I}_{L^*}\right)\\
&= \left(\dot{I}'_{L^*}\otimes I_1\right)\cdot D_{\Sigma^*}(D_L(\varphi))\\
&= \left(\dot{I}'_{L^*}\otimes I_1\right)\cdot D_\Sigma(D_L(\varphi))\cdot D_{\Sigma^*}(\Sigma(\Sigma^*))\\
&= \dot{I}'_{L^*}\cdot D_{L,\Sigma}(\varphi)\cdot \dot{I}_{\Sigma^*},\\
D_{L^*,\Phi^*}(\varphi) &= 0,\\
D_{L^*,Q^*}(\varphi) &= 0,
\end{aligned}$$

$$\begin{aligned}
D_{\Sigma^*,\Sigma^*}(\varphi) &= D_{\Sigma^*}(D_{\Sigma^*}(\varphi)) = D_{\Sigma^*}\left(D_\Sigma(\varphi)\cdot \dot{I}_{\Sigma^*}\right)\\
&= \dot{I}'_{\Sigma^*}\otimes I_1\cdot D_{\Sigma^*}(D_\Sigma(\varphi))\\
&= \dot{I}'_{\Sigma^*}\cdot D_\Sigma(D_\Sigma(\varphi))\cdot \dot{I}_{\Sigma^*},\\
D_{\Sigma^*,\Phi^*}(\varphi) &= 0,\\
D_{\Sigma^*,Q^*}(\varphi) &= 0.
\end{aligned}$$

$$\begin{aligned}
D_{\Phi^*,\Phi^*}(\varphi) &= \dot{I}'_{\Phi^*}\cdot(D_{\Phi,\Phi}(\varphi))\cdot \dot{I}_{\Phi^*},\\
D_{\Phi^*,Q^*}(\varphi) &= \dot{I}'_{\Phi^*}\cdot(D_{\Phi,Q}(\varphi))\cdot \dot{I}_{Q^*},\\
D_{Q^*,Q^*}(\varphi) &= \dot{I}'_{Q^*}\cdot D_{Q,Q}(\varphi)\cdot \dot{I}_{Q^*},
\end{aligned}$$

where $D_{L^*}(L(L^*)) = \dot{I}_{L^*}$, $D_{\Sigma^*}(\Sigma(\Sigma^*)) = \dot{I}_{\Sigma^*}$.

For $\dot{I}_{L^*}$ which is a block diagonal matrix, we have

$$\dot{I}_{L^*} = diag\,(P_1, P_2, ..., P_K)\,,$$

where

$$P_i = \begin{bmatrix} 0_{K\times(N-K)} \\ I_{N-K} \end{bmatrix}.$$

And for $\dot{I}_{\Sigma^*}$, which is an $N^2 \times N$ matrix whose n^{th} column has 1 in the $((n-1)\times N + n)^{th}$ row and other elements are all zeros. For $\dot{I}_{\Phi^*}$, we have

$$\dot{I}_{\Phi^*} = I_{K*K}.$$

For $\dot{I}_{Q^*}$, we have

$$\dot{I}_{Q^*} = diag\left(R_1, R_2, ...R_k, ..., R_K\right).$$

where

$$R_k = \begin{bmatrix} 0_{(k-1)\times(K-k+1)} \\ I_{K-k+1} \end{bmatrix}_{K\times(K-k+1)},$$

since Q is a symmetric matrix.

The first order derivative matrix of the complete-data likelihood with respect to $L^*, \Sigma^*, \Phi^*, Q^*$ is:

$$vec\left(\begin{bmatrix} D_{L^*}(\varphi) & D_{\Sigma^*}(\varphi) & D_{\Phi^*}(\varphi) & D_{Q^*}(\varphi) \end{bmatrix}\right).$$

The second order derivative matrix of the complete-data likelihood with respect to $L^*, \Sigma^*, \Phi^*, Q^*$ should be:

$$\begin{bmatrix} D_{L^*,L^*}(\varphi) & D_{L^*,\Sigma^*}(\varphi) & 0 & 0 \\ D_{\Sigma^*,L^*}(\varphi) & D_{\Sigma^*,\Sigma^*}(\varphi) & 0 & 0 \\ 0 & 0 & D_{\Phi^*,\Phi^*}(\varphi) & D_{\Phi^*,Q^*}(\varphi) \\ 0 & 0 & D_{Q^*,\Phi^*}(\varphi) & D_{Q^*,Q^*}(\varphi) \end{bmatrix}.$$

F The derivation of RDIC for the stochastic volatility model

F.1 The derivatives of the complete-data log-likelihood for M_1

The complete-data log-likelihood function

$$\begin{aligned} \ln p(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) &= -n\ln 2\pi + \frac{n}{2}\ln\nu - \frac{1}{2}\sum_{t=1}^{n} h_t - \frac{1}{2}\sum_{t=1}^{n}\frac{(y_t-\alpha)^2}{\exp(h_t)} \\ & -\frac{1}{2}\nu\left[\sum_{t=1}^{n}\left(h_t - \mu - \phi\left(h_{t-1}-\mu\right)\right)^2\right], \end{aligned}$$

where $\mathbf{y} = (y_1, y_2, ...y_n)'$, $\mathbf{h} = (h_1, h_2, ...h_n)'$, $\nu = 1/\tau^2$.

The first order derivatives

$$\frac{\partial \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \alpha}=\sum_{t=1}^{n}\frac{\left(y_t-\alpha\right)}{\exp\left(h_t\right)},$$

$$\begin{aligned}\frac{\partial \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \mu} &= -\frac{1}{2}\nu\left[-2\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)\left(1-\phi\right)\right]\\ &= \nu\left[\left(1-\phi\right)\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \phi} &= -\frac{1}{2}\nu\left[-2\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)\left(h_{t-1}-\mu\right)\right]\\ &= \nu\left[\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)\left(h_{t-1}-\mu\right)\right],\end{aligned}$$

$$\frac{\partial \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \nu}=\frac{n}{2}\frac{1}{\nu}-\frac{1}{2}\left[\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)^2\right].$$

The second order derivatives

$$\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \alpha\partial \alpha}=-\sum_{t=1}^{n}\frac{1}{\exp\left(h_t\right)}=-\sum_{t=1}^{n}\exp\left(-h_t\right),$$

$$\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \alpha\partial \mu}=\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \alpha\partial \phi}=\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \alpha\partial \nu}=0,$$

$$\begin{aligned}\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \mu\partial \mu} &= \nu\left[-\left(1-\phi^2\right)-\left(1-\phi\right)\sum_{t=1}^{n}\left(1-\phi\right)\right]\\ &= -\nu\left[n\left(1-\phi\right)^2\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \mu\partial \phi} &= \nu\left[-\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)-\left(1-\phi\right)\sum_{t=1}^{n}\left(h_{t-1}-\mu\right)\right]\\ &= -\nu\left[\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)+\left(1-\phi\right)\sum_{t=1}^{n}\left(h_{t-1}-\mu\right)\right],\end{aligned}$$

$$\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \mu\partial \nu}=\left(1-\phi\right)\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right),$$

$$\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \phi\partial \phi}=\nu\left[-\sum_{t=1}^{n}\left(h_{t-1}-\mu\right)^2\right],$$

$$\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \phi\partial \nu}=\sum_{t=1}^{n}\left(h_t-\mu-\phi\left(h_{t-1}-\mu\right)\right)\left(h_{t-1}-\mu\right),$$

$$\frac{\partial^2 \ln p\left(\mathbf{y},\mathbf{h}|\boldsymbol{\theta}\right)}{\partial \nu\partial \nu}=-\frac{n}{2\nu^2}.$$

F.2 The derivatives of the complete-data log-likelihood for M_2

The complete-data log-likelihood function

$$\begin{aligned}\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right) &= \sum_{t=1}^{n}\ln\sigma_t^2-\frac{n}{2}\ln 2\pi+\frac{n}{2}\ln\nu-\frac{1}{2}\nu\left[\sum_{t=1}^{n}\left(\sigma_t^2-\mu-\phi\left(\ln\sigma_{t-1}^2-\mu\right)\right)^2\right] \\ & -\frac{1}{2}\sum_{t=1}^{n}\frac{\left(y_t-\alpha\right)^2}{\sigma_t^2}-\frac{n}{2}\ln 2\pi-\frac{1}{2}\sum_{t=1}^{n}\ln\sigma_t^2,\end{aligned}$$

where $\boldsymbol{\sigma^2}=\left(\sigma_1^2,\sigma_2^2,...\sigma_n^2\right)'$.

The first order derivatives

$$\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\alpha}=\sum_{t=1}^{n}\frac{y_t-\alpha}{\sigma_t^2},$$

$$\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\mu} = \nu\left[(1-\phi)\sum_{t=1}^{n}\left(\sigma_t^2-\mu-\phi\left(\ln\sigma_{t-1}^2-\mu\right)\right)\right],$$

$$\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\phi} = \nu\left[\sum_{t=1}^{n}\left(\sigma_t^2-\mu-\phi\left(\ln\sigma_{t-1}^2-\mu\right)\right)\left(\ln\sigma_{t-1}^2-\mu\right)\right],$$

$$\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\nu}=\frac{n}{2\nu}-\frac{1}{2}\left[\sum_{t=1}^{n}\left(\sigma_t^2-\mu-\phi\left(\ln\sigma_{t-1}^2-\mu\right)\right)^2\right].$$

The second order derivatives

$$\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\alpha\partial\alpha}=-\sum_{t=1}^{n}\frac{1}{\sigma_t^2},$$

$$\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\alpha\partial\mu}=\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\alpha\partial\phi}=\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\alpha\partial\nu}=0,$$

$$\begin{aligned}\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\mu\partial\mu} &= \nu\left[-\left(1-\phi^2\right)-(1-\phi)\sum_{t=1}^{n}(1-\phi)\right] \\ &= -\nu\left[n\left(1-\phi\right)^2\right],\end{aligned}$$

$$\begin{aligned}\frac{\partial\ln p\left(\mathbf{y},\boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\mu\partial\phi} &= \nu\left[-\sum_{t=1}^{n}\left(\sigma_t^2-\mu-\phi\left(\ln\sigma_{t-1}^2-\mu\right)\right)-(1-\phi)\sum_{t=1}^{n}\left(\ln\sigma_{t-1}^2-\mu\right)\right] \\ &= -\nu\left[\sum_{t=1}^{n}\left(\sigma_t^2-\mu-\phi\left(\ln\sigma_{t-1}^2-\mu\right)\right)+(1-\phi)\sum_{t=1}^{n}\left(\ln\sigma_{t-1}^2-\mu\right)\right],\end{aligned}$$

$$\frac{\partial \ln p\left(\mathbf{y}, \boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\mu\partial\nu} = (1-\phi)\sum_{t=1}^{n}\left(\sigma_t^2 - \mu - \phi\left(\ln\sigma_{t-1}^2 - \mu\right)\right),$$

$$\frac{\partial \ln p\left(\mathbf{y}, \boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\phi\partial\phi} = \nu\left[-\sum_{t=1}^{n}\left(\ln\sigma_{t-1}^2 - \mu\right)^2\right],$$

$$\frac{\partial \ln p\left(\mathbf{y}, \boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\phi\partial\nu} = \sum_{t=1}^{n}\left(\sigma_t^2 - \mu - \phi\left(\ln\sigma_{t-1}^2 - \mu\right)\right)\left(\ln\sigma_{t-1}^2 - \mu\right),$$

$$\frac{\partial^2 \ln p\left(\mathbf{y}, \boldsymbol{\sigma^2}|\boldsymbol{\theta}\right)}{\partial\nu\partial\nu} = -\frac{n}{2\nu^2}.$$

F.3 Gaussian Approximation

The complete-data log-likelihood function of M_1 can be also expressed as follows:

$$\begin{aligned}\ln\left(p\left(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}\right)\right) &= -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\tau^2\right) - \frac{1}{2}\left(\mathbf{h}-\boldsymbol{\mu}\right)'\boldsymbol{Q}\left(\mathbf{h}-\boldsymbol{\mu}\right) \\ &\quad -\frac{n}{2}\ln\left(2\pi\right) - \frac{1}{2}\sum_{t=1}^{n}h_t - \sum_{t=1}^{n}\frac{\left(y_t-\alpha\right)^2}{2}\exp\left(-h_t\right),\end{aligned}$$

where $\mathbf{h} = (h_1, h_2, ..., h_n)$, $\boldsymbol{\mu} = \mu\mathbf{e}$, $\mathbf{e}' = (1, \ldots, 1)_n$, $\boldsymbol{Q}$ is a tri-diagonal precision matrix, $\boldsymbol{Q} = \boldsymbol{Q^*}/\tau^2$, $\boldsymbol{Q^*}$ is defined as follows:

$$\boldsymbol{Q^*} = \begin{pmatrix} \phi^2 & -\phi & & & & \\ -\phi & 1+\phi^2 & -\phi & & & \\ & & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot & \\ & & & -\phi & 1+\phi^2 & -\phi \\ & & & & -\phi & 1 \end{pmatrix}.$$

The posterior density of $\mathbf{h}$ is

$$\begin{aligned}p\left(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}\right) &\propto \exp\left[-\frac{1}{2}\left(\mathbf{h}-\boldsymbol{\mu}\right)'\boldsymbol{Q}\left(\mathbf{h}-\boldsymbol{\mu}\right) - \sum_{t=1}^{n}\left(\frac{1}{2}h_t + \frac{\left(y_t-\alpha\right)^2}{2}\exp\left(-h_t\right)\right)\right] \\ &= \exp\left(f\left(\mathbf{h}\right)\right) \approx \exp\left(-\frac{1}{2}\mathbf{h}'\boldsymbol{c}\mathbf{h} + \boldsymbol{b}\mathbf{h} + cons\right).\end{aligned}$$

In order to obtain the parameters c and b of the canonical form, we use the first and second order derivatives:

$$\begin{aligned}\dot{f}\left(\mathbf{h}\right) &= -\mathbf{h}'\boldsymbol{Q} + \boldsymbol{\mu}'\boldsymbol{Q} - \frac{1}{2}\mathbf{e}' + \frac{1}{2}\left(\mathbf{y}^{*2}\right)'\odot\exp\left(-\mathbf{h}\right)' \\ \ddot{f}\left(\mathbf{h}\right) &= -\boldsymbol{Q} - diag\left(\frac{1}{2}\left(\mathbf{y}^*\right)^2\odot\exp\left(-\mathbf{h}\right)\right),\end{aligned}$$

where $\mathbf{y}^* = \mathbf{y} - \boldsymbol{\alpha}$ and $\boldsymbol{\alpha} = \alpha\mathbf{e}$, $\mathbf{e}' = (1, \ldots, 1)_n$, $\mathbf{y}^{*2} = (y_1^{*2}, \ldots, y_n^{*2})'$ and $\exp(-\mathbf{h}) = (\exp(-h_1), \ldots, \exp(-h_n))'$.

Denoting the mode of f by $\mathbf{m}$, we apply the Taylor expansion to $f(x)$:

$$\begin{aligned} f(\mathbf{h}) &\approx (\mathbf{h}-\mathbf{m})' \frac{\ddot{f}(\mathbf{m})}{2} (\mathbf{h}-\mathbf{m}) + \dot{f}(\mathbf{m})(\mathbf{h}-\mathbf{m}) + \mathit{cons} \\ &= -\frac{1}{2}\mathbf{h}' \left(-\ddot{f}(\mathbf{m})\right) \mathbf{h} - \mathbf{m}'\ddot{f}(\mathbf{m})\,\mathbf{h} + \dot{f}(\mathbf{m})\,\mathbf{h} + \mathit{cons} \\ &= -\frac{1}{2}\mathbf{h}'\mathbf{c}\mathbf{h} + \mathbf{b}\mathbf{h} + \mathit{cons}. \end{aligned}$$

Now, we obtain $\mathbf{c}$ and $\mathbf{b}$ as

$$\mathbf{c} = -\ddot{f}(\mathbf{m}) = \mathbf{Q} + \mathit{diag}\left(\frac{1}{2}\mathbf{y}^{*2} \odot \exp(-\mathbf{m})\right),$$

$$\begin{aligned} \mathbf{b} &= -\mathbf{m}'\ddot{f}(\mathbf{m}) + \dot{f}(\mathbf{m}) \\ &= \mathbf{m}'\mathbf{Q} + \mathbf{m}'\mathit{diag}\left(\frac{1}{2}\mathbf{y}^{*2} \odot \exp(-\mathbf{m})\right) \\ &\quad -\mathbf{m}'\mathbf{Q} + \boldsymbol{\mu}'\mathbf{Q} - \frac{1}{2}\mathbf{e}' + \frac{1}{2}\left(\mathbf{y}^{*2}\right)' \odot \exp(-\mathbf{m})' \\ &= \mathbf{m}'\mathit{diag}\left(\frac{1}{2}\mathbf{y}^{*2} \odot \exp(-\mathbf{m})\right) + \frac{1}{2}\left(\mathbf{y}^{*2}\right)' \odot \exp(-\mathbf{m})' + \boldsymbol{\mu}'\mathbf{Q} - \frac{1}{2}\mathbf{e}'. \end{aligned}$$

Using

$$-\frac{1}{2}\mathbf{h}'\mathbf{c}\mathbf{h} + \mathbf{b}\mathbf{h} + \mathit{cons} = -\frac{1}{2}(\mathbf{h}-\mathbf{m}^*)'\,\mathbf{Q}^*(\mathbf{h}-\mathbf{m}^*),$$

we obtain

$$\begin{aligned} \mathbf{Q}^* &= \mathbf{c} = \mathbf{Q} + \mathit{diag}\left(\frac{1}{2}\mathbf{y}^{*2} \odot \exp(-\mathbf{m})\right), \\ \mathbf{m}^* &= \mathbf{Q}^{*-1}\mathbf{b}'. \end{aligned}$$

In order to obtain the optimal mode of $\mathbf{Q}^*$ and $\mathbf{m}^*$, we run the above procedure recursively until convergence.

References

Akaike, H. (1973): "Information theory and an extension of the maximum likelihood principle," in *Second international symposium on information theory*, Springer Verlag, vol. 1, 267–281.

Andrews, D. and C. Mallows (1974): "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society Series B*, 36, 99–102.

Berg, A., R. Meyer, and J. Yu (2004): "Deviance information criterion for comparing stochastic volatility models," *Journal of Business and Economic Statistics*, 22, 107–120.

BERNANKE, B., J. BOIVIN, AND P. ELIASZ (2005): "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly Journal of Economics*, 120, 387–422.

BROOKS, S. (2002): "Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002)," *Journal of the Royal Statistical Society Series B*, 64, 616–618.

CELEUX, G., F. FORBES, C. ROBERT, AND D. TITTERINGTON (2006): "Deviance Information Criteria for Missing Data Models," *Bayesian Analysis*, 1, 651–674.

CHEN, C. (1985): "On asymptotic normality of limiting density functions with Bayesian implications," *Journal of the Royal Statistical Society Series B*, 540–546.

CHIB, S. (1995): "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

CHIB, S. AND I. JELIAZKOV (2001): "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270–281.

CLARK, P. (1973): "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, 41, 135–155.

DANIELS, M. AND J. HOGAN (2008): *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, vol. 109, Chapman & Hall.

DEIORIO, M. AND C. ROBERT (2002): "Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002)," *Journal of the Royal Statistical Society Series B*, 64, 629–630.

DEMPSTER, A., N. LAIRD, AND D. RUBIN (1977): "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B*, 39, 1–38.

FAMA, E. AND K. FRENCH (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

GELFAND, A. AND M. TREVISANI (2002): "Discussion on the paper by Spiegelhalter, Best, Carlin, and van de Linde (2002)," *Journal of the Royal Statistical Society Series B*, 64, 629–630.

GELMAN, A. (2004): *Bayesian Data Analysis*, CRC Press.

GEWEKE, J. (1977): "The dynamic factor analysis of economic time-Series models," in *Latent variables in socio-economic models*, ed. by A. Aigner and A. Goldberger, North-Holland, 365–395.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*, vol. 537, Wiley-Interscience.

Geweke, J., G. Koop, and H. Van Dyk (2011): *Oxford Handbook of Bayesian Econometrics*, Oxford Univ Press.

Ghosh, J. and R. Ramamoorthi (2003): *Bayesian nonparametrics*, Springer Verlag.

Giannone, D., L. Reichlin, and L. Sala (2004): "Monetary policy in real time," *NBER Macroeconomics Annual*, 161–200.

Huang, S. and J. Yu (2010): "Bayesian analysis of structural credit risk models with microstructure noises," *Journal of Economic Dynamics and Control*, 34, 2259–2272.

Ibrahim, J., H. Zhu, and N. Tang (2008): "Model selection criteria for missing-data problems using the EM algorithm," *Journal of the American Statistical Association*, 103, 1648–1658.

Kan, R. and G. Zhou (2003): "Modeling non-normality using multivariate t: Implications for asset pricing," Tech. rep., Working Paper, University of Toronto.

Kass, R. and A. Raftery (1995): "Bayes factors," *Journal of the American Statistical Association*, 90, 773–795.

Kim, J. (1994): "Bayesian asymptotic theory in a time series model with a possible nonstationary process," *Econometric Theory*, 10, 764–773.

——— (1998): "Large sample properties of posterior densities, bayesian information criterion and the likelihood principle in nonstationary time series models," *Econometrica*, 66, 359–380.

Kose, A. M., C. Otrok, and C. Whiteman (2003): "International business cycles: World, region, and country-specific factors," *American Economic Review*, 93, 1216–1239.

——— (2008): "Understanding the evolution of world business cycles," *Journal of International Economics*, 75, 110–130.

Li, Y. and J. Yu (2012): "Bayesian hypothesis testing in latent variable models," *Journal of Econometrics*, 166, 237–246.

Louis, T. (1982): "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society Series B*, 44, 226–233.

Magnus, J. and H. Neudecker (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester.

McLachlan, G. and T. Krishnan (2008): *The EM Algorithm and Extensions*, vol. 382, John Wiley and Sons.

Meyer, R. and J. Yu (2000): "BUGS for a Bayesian analysis of stochastic volatility models," *The Econometrics Journal*, 3, 198–215.

Oakes, D. (1999): "Direct calculation of the information matrix via the EM," *Journal of the Royal Statistical Society Series B*, 61, 479–482.

Otrok, C. and C. Whiteman (1998): "Bayesian leading indicators: measuring and predicting economic conditions in Iowa," *International Economic Review*, 39, 997–1014.

Phillips, P. (1996): "Econometric model determination," *Econometrica*, 64, 763–812.

Phillips, P. and W. Ploberger (1996): "An asymptotic theory of bayesian inference for time series," *Econometrica*, 64, 381–412.

Raftery, A. and S. Lewis (1992): "How many iterations in the Gibbs sampler," *Bayesian statistics*, 4, 763–773.

Rue, H., S. Martino, and N. Chopin (2009): "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society Series B*, 71, 319–392.

Rue, H., I. Steinsland, and S. Erland (2004): "Approximating hidden Gaussian Markov random fields," *Journal of the Royal Statistical Society Series B*, 66, 877–892.

Sargent, T. and C. Sims (1977): "Business cycle modeling without pretending to have too much a priori economic theory," *New Methods in Business Research, Federal Reserve Bank of Minneapolis, Minneapolis.*

Spiegelhalter, D., N. Best, B. Carlin, and A. Van Der Linde (2002): "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society Series B*, 64, 583–639.

Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn (2003): "WinBUGS version 1.4 user manual," .

Stock, J. and M. Watson (1999): "Forecasting inflation," *Journal of Monetary Economics*, 44, 293–335.

——— (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–162.

——— (2010): "Dynamic factor models," *Prepared for the Oxford Handbook of Economic Forecasting.*

Sturtz, S., U. Ligges, and A. Gelman (2005): "R2WinBUGS: A Package for Running WinBUGS from R," *Journal of Statistical Software*, 12, 1–16.

Tanner, M. and W. Wong (1987): "The calculation of posterior distributions by data augmentation," *Journal of the American statistical Association*, 82, 528–540.

Tu, J. and G. Zhou (2010): "Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty," *Journal of Financial and Quantitative Analysis*, 45, 959–986.

Wang, J., J. Chan, and S. Choy (2011): "Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures," *Computational Statistics and Data Analysis*, 55, 852–862.

Wu, C. (1983): "On the convergence properties of the EM algorithm," *The Annals of Statistics*, 11, 95–103.

Zhou, G. (1993): "Asset-pricing tests under alternative distributions," *Journal of Finance*, 48, 1927–1942.