5-2015

A Bayesian Specification Test

Yong LI
Renmin University of China

Tao ZENG
Wuhan University

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

A Bayesian Specification Test*

Yong Li
Renmin University of China

Tao Zeng
Wuhan University

Jun Yu
Singapore Management University

May 26, 2015

Abstract

A Bayesian test statistic is proposed to assess the model specification after the model is estimated by Bayesian MCMC methods. The proposed approach does not require an alternative model to be specified and is applicable to a variety of models, including latent variable models, structural dynamic choice models, and dynamics stochastic general equilibrium (DSGE) models, for which frequentist methods are difficult to use. The properties of the test statistic are established and its implementation is discussed. The test is easy to use and the test statistic can be calculated from MCMC outputs even when there are latent variables. The method is illustrated using a dynamic factor model, a DSGE model and a stochastic volatility model.

JEL classification: C11, C12, G12
Keywords: EM algorithm; Specification test; Latent variable models; Markov chain Monte Carlo; Dynamic factor; DSGE; Stochastic volatility.

1 Introduction

Economic theory has long been used to justify a particular choice of econometric models. These so-called structural econometric models are often based on a set of economic assumptions used to develop the underlying economic theory. When some of the assumptions are invalid, the corresponding structural econometric models may be misspecified. In many cases, economic theory may not be available and the choice of econometric models

*Li gratefully acknowledges the financial support of the Chinese Natural Science fund (No.71271221) and the hospitality during his research visits to Singapore Management University. Yu would like to acknowledge the financial support from Singapore Ministry of Education Academic Research Fund Tier 2 under the grant number MOE2011-T2-2-096. Yong Li, Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing, 1000872, China. Tao Zeng, Economics and Management School, Wuhan University, Wuhan, 430072, China. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.edu.sg.

may be arbitrary. Consequently, models in a reduced form are used and reduced form models are vulnerable to specification errors.

In general misspecification of econometric models can potentially lead to inconsistent estimation, which in turn may have serious implications for statistical inferences such as hypothesis testing and out-of-sample forecasting and for economic decision makings such as policy recommendation and investment decision. Consequently and not surprisingly, a considerable amount of strenuous effort has been devoted in econometrics to detect model misspecification.

One strand of the literature on specification tests unifies under the m-test of Newey (1985), Tauchen (1985) and White (1987). These tests include as a special case of the Lagrangian multiplier (LM) test, the tests of Sargan (1958) and Hansen (1982), the tests of Cox (1961, 1962), the Hausman (1978) test, the information matrix test of White (1982), the conditional moment test of Newey (1985), the IOS test of Presnell and Boos (2004). These tests are in the frequentist paradigm, typically requiring parameters in the null hypothesis be estimated by the maximum likelihood (ML) method, or by generalized method of moments (GMM).

Another strand of the literature is based on tests that rely on the distances between nonparametric and parameter counterparts. The idea originated from the Kolmogorov-Smirnov test or the closely related family such as the Cramer-von Mises and Andersen-Darling tests. Examples in this case include Eubank and Spiegelman (1990), Wooldrige (1992), Fan and Li (1996), Gozalo (1993), Zheng (2000), Aït-Sahalia (1996), and Hong and Li (2005). All the tests in this category are also in the frequentist paradigm, but requiring either a nonparametric estimate of a function or a nonparametric estimate of a density (either a marginal density or a conditional density).

For many widely used models in economics, such as latent variable models, structural dynamic choice models (Imai, Jain and Ching, 2009) and dynamic stochastic general equilibrium models (DSGE), it is not easy to obtain the ML estimate or construct a nonparametric estimate. In some cases, even when a frequentist method is available, a Bayesian method is preferred as it can take into account of strong priors imposed by researchers to shrink the unrestricted model towards a parsimonious specification. Typical examples where researchers would like to impose strong priors include Bayesian VAR models (Karlsson, 2015), DSGE models (An and Schorfheide, 2006), the estimation of a large dimensional covariance matrix (Ledoit and Wolf, 2004). Not surprisingly, it is difficult to apply any of the specification tests mentioned above. On the other hand, there has been an increasing interest in using Bayesian methods to estimate econometric models. With the advancement of the Markov chain Monte Carlo (MCMC) algorithms and the rapid growth in computer capability, fitting models of increasing complexity has become easier and easier.

Given the increasing popularity of Bayesian MCMC methods in practical applications, it is therefore natural to introduce a Bayesian test to assess the goodness-of-fit of candidate models. Unfortunately, model specification test is a challenge task in the Bayesian paradigm. Perhaps the most obvious Bayesian way to assess the goodness-of-fit of the model is to compare the posterior model probability in consideration with that of a competing model. This can be achieved by using, for example, Bayes factors (BFs), although BFs are not free of problems. However, it is often not clear how to specify the alternative model and empirical researchers may simply wish to know if the model she employs is adequate or not after the model is estimated without worrying about any alternative model.

The question we ask in the present paper is, after the model is estimated by a Bayesian approach, how we can assess the validity of the model specification. The main purpose of this paper is to introduce a Bayesian approach to test model specification without specifying an alternative model. The proposed Bayesian test statistic is the Bayesian version of the IOS_a test of Presnell and Boos (2004). Properties of our test statistic are established. We show how to compute the test statistic from MCMC output when there are latent variables in the model for which the likelihood function does not have a closed-form expression. We also show how the Bayesian credible intervals can be used to implement our method.

The paper is organized as follows. Section 2 briefly reviews the literature on the specification tests. Section 3 proposes the new Bayesian test statistic and establishes the properties of the proposed test. In Section 4, we discuss how to compute the test statistic from the MCMC output for latent variable models. Section 5 illustrates the new method using three real examples. Section 6 concludes the paper. Appendix collects the proof of the theoretical results in the paper and derives the quantities that are needed to compute the test statistic.

2 Specification Tests: A Literature Review

To begin, let $\mathbf{y} = (y_1, \ldots, y_n)$ denote observed variables from a probability measure P_0 on the probability space (Ω, F, P_0). Let model P be a collection of candidate models indexed by parameters $\boldsymbol{\theta}$ whose dimension is p. Denote P indexed by $\boldsymbol{\theta}$ by $P_{\boldsymbol{\theta}}$. Following White (1987), if there exists $\boldsymbol{\theta}$, such that $P_0 \in P_{\boldsymbol{\theta}}$, we say the model P is correctly specified. However, if for all $\boldsymbol{\theta}$, $P_0 \notin P_{\boldsymbol{\theta}}$, we say the model P is misspecified. We would like to test the null hypothesis that the model in concern is correctly specified.

One of the earliest specification tests is based on the informative matrix equivalence due to White (1982). Let $p(\mathbf{y}|\boldsymbol{\theta})$ denote the likelihood function of model P and

$$\mathbf{s}(\mathbf{y},\boldsymbol{\theta}) := \partial \log p(\mathbf{y}|\boldsymbol{\theta})/\partial\boldsymbol{\theta},\ \mathbf{h}(\mathbf{y},\boldsymbol{\theta}) := \partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}',$$
$$\mathbf{H}(\boldsymbol{\theta}) := \int \mathbf{h}(\mathbf{y},\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\mathrm{d}\mathbf{y},\ \mathbf{J}(\boldsymbol{\theta}) := \int \mathbf{s}(\mathbf{y},\boldsymbol{\theta})\mathbf{s}'(\mathbf{y},\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})\mathrm{d}\mathbf{y}.$$

Under the null hypothesis that the model is correctly specified, it is well-known that $\mathbf{H}(\boldsymbol{\theta}) + \mathbf{J}(\boldsymbol{\theta}) = 0$. Define

$$d(\mathbf{y},\boldsymbol{\theta}) := vech\left[\mathbf{h}(\mathbf{y},\boldsymbol{\theta}) + \mathbf{s}(\mathbf{y},\boldsymbol{\theta})\mathbf{s}'(\mathbf{y},\boldsymbol{\theta})\right],$$

where *vech* is the column-wise vectorization with the upper portion excluded. Hence, $d(\mathbf{y},\boldsymbol{\theta}) = (d_k(\mathbf{y},\boldsymbol{\theta}))$ is a q $(= p(p+1)/2)$ dimensional vector. Let $\mathbf{y} = (y_1,\ldots,y_n)$ denote the iid observations and

$$\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) := \frac{1}{n}\sum_{t=1}^{n}\mathbf{h}(y_t,\hat{\boldsymbol{\theta}}),\ \hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) := \frac{1}{n}\sum_{t=1}^{n}\mathbf{s}(y_t,\hat{\boldsymbol{\theta}})\mathbf{s}'(y_t,\hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$. Let

$$D_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} d(y_t,\hat{\boldsymbol{\theta}}),$$
$$\dot{D}_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial d(y_t,\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}},$$

where $D_n(\hat{\boldsymbol{\theta}})$ is a q-dimensional vector and $\dot{D}_n(\hat{\boldsymbol{\theta}})$ is a $q\times p$ matrix. White (1982) proposed the following information matrix test

$$IMT = nD_n(\hat{\boldsymbol{\theta}})V_n^{-1}(\hat{\boldsymbol{\theta}})D_n(\hat{\boldsymbol{\theta}}), \tag{1}$$

where $V_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n}\left[d(y_t,\hat{\boldsymbol{\theta}}) - \dot{D}_n(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{s}(y_t,\hat{\boldsymbol{\theta}})\right]\left[d(y_t,\hat{\boldsymbol{\theta}}) - \dot{D}_n(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{s}(y_t,\hat{\boldsymbol{\theta}})\right]'$.

Under a set of regularity conditions, White (1982) showed that $IMT \xrightarrow{d} \chi^2(q)$ under the null hypothesis. White (1987) extended the method to cover dynamic models. Lancaster (1984) pointed out that the covariance matrix of the information matrix test can be estimated without computing the third derivatives of the density function analytically. Dhaene and Hoorelbeke (2004) suggested using the bootstrap method to estimate the covariance matrix. Moreover, it is well documented that the χ^2 distribution can be a poor approximation in finite sample so that the test statistic suffers from a serious size distortion; see Orme (1990), Chesher and Spady (1991), Davidson and Mackinnon (1992), Horowitz (1994). To improve the finite sample performance of IMT, Chesher and Spady (1991) used the high-order Edgeworth expansion to obtain better critical values while Horowitz (1994) advocated the use of bootstrap methods to obtain better critical values.

To deal with the difficulties associated with the information matrix test, Presnell and Boos (2004) proposed an "in-and-out" likelihood ratio (IOS) test for models with iid observations. Let $\hat{\boldsymbol{\theta}}_{(t)}$ be the MLE of $\boldsymbol{\theta}$ when the t-th observation, y_t, is deleted from the whole sample. From the predictive perspective, the single likelihood $p\left(y_t, \hat{\boldsymbol{\theta}}_{(t)}\right)$ can be regarded as the predictive likelihood by the other observations. Presnell and Boos (2004) defined the "in-and-out" likelihood ratio test as:

$$IOS = \log \frac{\prod_{t=1}^{n} p(y_t, \hat{\boldsymbol{\theta}})}{\prod_{t=1}^{n} p\left(y_t, \hat{\boldsymbol{\theta}}_{(t)}\right)} = \sum_{t=1}^{n} \left[\log p(y_t|\hat{\boldsymbol{\theta}}) - \log p\left(y_t, \hat{\boldsymbol{\theta}}_{(t)}\right)\right],$$

and showed that the asymptotic form of IOS is

$$IOS_a = \mathbf{tr}\left[-\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\right],$$

and $IOS - IOS_a = o_p(n^{-1/2})$. Under the null hypothesis, $IOS_a \xrightarrow{p} \mathbf{tr}\left[-\mathbf{H}^{-1}(\boldsymbol{\theta_0})\mathbf{J}(\boldsymbol{\theta_0})\right] = p$. Under a set of regularity conditions, Presnell and Boos (2004) further showed that both $n^{1/2}(IOS - p)$ and $n^{1/2}(IOS_a - p)$ converge to a normal distribution under the null hypothesis.

Note that IOS_a is the same as the penalty term of the well-known information criterion, TIC, proposed by Takeuchi (1976). When the dimension of $\boldsymbol{\theta}$ is high, to compute IOS_a, one has to calculate the inverse of $\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})$ which may be computationally demanding. Another important feature of the IOS test is that it only deals with iid models. It is not clear how to implement the IOS test when the iid assumption breaks down.

It is not necessary to base a specification test on maximum likelihood (ML) estimation. Newey (1985) developed a class of specification tests based on a finite set of moment conditions and the GMM estimator. Under some regularity conditions, the test statistic of Newey follows asymptotically a χ^2 distribution. It was shown that his test includes as a special case of the tests of Hausman (1978) and Hansen (1982).

Specification of a stationary dynamic model implicitly implies a distributional assumption for the marginal density and that for the conditional density. Not surprisingly, many specification tests check the validity of these distributional assumptions based on the Kolmogorov-Smirnov test or the closely related family such as the Cramer-von Mises and Andersen-Darling tests. Examples include Zheng (2000), Andrews (1997), Corradi and Swanson (2004), Aït-Sahalia (1996), and Hong and Li (2005). For example, Aït-Sahalia (1996) compares the parametric marginal density implied by the assumed continuous time model to the marginal density estimated nonparametrically. The nonparametric test of Hong and Li (2005) is based on the transition density.

The literature is much less extensive on Bayesian specification tests although Bayesian MCMC methods have been used more and more frequently for model estimation in practice. A notable exception is the Bayesian χ^2 test of Johnson (2004). Geweke and McCauland (2001) outlines some essentials of Baysian specification analysis. In this paper we propose a Bayesian specification test that is widely applicable and easy to implement.

3 A New Bayesian Approach for Specification Test

The problem concerned in this paper is to assess the specification of a candidate model given that the model is estimated by MCMC without worrying about any competing model. Before proposing the test, we need to introduce some notations. Let $\mathbf{y}^t := (y_1, \ldots, y_t)$, and

$$\begin{aligned}
&\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta}) := \frac{\partial \log p(\mathbf{y}^t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},\ \mathbf{h}(\mathbf{y}^t, \boldsymbol{\theta}) := \frac{\partial^2 \log p(\mathbf{y}^t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'},\\
&\mathbf{s}_t(\boldsymbol{\theta}) := \mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta}) - \mathbf{s}(\mathbf{y}^{t-1}, \boldsymbol{\theta}),\ \mathbf{h}_t(\boldsymbol{\theta}) := \mathbf{h}(\mathbf{y}^t, \boldsymbol{\theta}) - \mathbf{h}(\mathbf{y}^{t-1}, \boldsymbol{\theta}),\\
&\hat{\mathbf{J}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{s}_t(\boldsymbol{\theta})\mathbf{s}'_t(\boldsymbol{\theta}),\ \hat{\mathbf{H}}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{h}_t(\boldsymbol{\theta}),\\
&L_n(\boldsymbol{\theta}) := \log p(\boldsymbol{\theta}|\mathbf{y}),\ L_n^{(k)}(\boldsymbol{\theta}) := \partial^k \log p(\boldsymbol{\theta}|\mathbf{y})/\partial \boldsymbol{\theta}^k.
\end{aligned}$$

In this paper, we assume that the following mild regularity conditions are satisfied.

Assumption 1: Let $\hat{\boldsymbol{\theta}}$ is the posterior mode such that $L_n^{(1)}(\hat{\boldsymbol{\theta}}) = 0$. For any $\epsilon > 0$, there exists an integer N_1 and some $\delta > 0$ such that for when $n > N_1$ and $\boldsymbol{\theta} \in H(\hat{\boldsymbol{\theta}}, \delta) = \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|| \leq \delta\}$, $L_n^{(2)}(\boldsymbol{\theta})$ is negative definite.

Assumption 2: The largest eigenvalue of $[-L_n^{(2)}(\hat{\boldsymbol{\theta}})]^{-1}$ tends to zero as $n \to \infty$.

Assumption 3: For any $\epsilon > 0$, there exists an integer N_2 and some $\delta > 0$ such that for any $n > \max\{N_1, N_2\}$ and $\boldsymbol{\theta} \in H(\hat{\boldsymbol{\theta}}, \delta) = \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|| \leq \delta\}$, $L_n^{(2)}(\boldsymbol{\theta})$ satisfies the following inequality

$$-A(\epsilon) \leq L_n^{(2)}(\boldsymbol{\theta}) L_n^{-(2)}(\hat{\boldsymbol{\theta}}) - \mathbf{I}_p \leq A(\epsilon),$$

where $\mathbf{I}_p$ is a p-dimensional identity matrix, $A(\epsilon)$ is a positive semidefinite symmetric matrix whose largest eigenvalue goes to zero as $\epsilon \to 0$. $A \leq B$ means that $A_{ij} \leq B_{ij}$ for all i, j.

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

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

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

Assumption 5: Let $g(\mathbf{y})$ be the true data generating process (DGP), and denote $\boldsymbol{\theta}_0 \in \Theta \subset R^p$ the pseudo-true value that minimizes the Kullback-Leibler (KL) loss between the DGP and the parametric model,

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} \int \log \frac{g(\mathbf{y})}{p(\mathbf{y}|\boldsymbol{\theta})} g(\mathbf{y}) d\mathbf{y}.$$

For any sequence $k_n \to 0$,

$$\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0||<k_n} n^{-1}\sum_{i=1}^{n} ||h_t(\boldsymbol{\theta}) - h_t(\boldsymbol{\theta}_0)|| \xrightarrow{p} 0.$$

Furthermore, it is assumed that

$$\sup_{||\theta-\theta_0||<k_n}\left[\sup_{t\leq n}||h_t(\theta)||\right] = o_p(n),$$

$\frac{1}{n}\sum||h_t(\boldsymbol{\theta}_0)|| = O_p(1)$ and $\sup_{t\leq n} n^{-1/2}||s_t(\theta_0)|| \xrightarrow{p} 0$

Assumption 6: The prior $p(\boldsymbol{\theta})$ is $O_p(1)$.

Remark 3.1 *The regularity conditions 1-4 have been used to develop the Bayesian large sample theory; see, for example, Chen (1985), Kim (1994, 1998), Geweke (2005). Based on these assumptions, Li et al. (2014) showed that,*

$$\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\mathbf{y}] = \int p(\boldsymbol{\theta}|\mathbf{y})\boldsymbol{\theta}\,d\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + o_p(n^{-1/2}),$$
$$V(\hat{\boldsymbol{\theta}}) = -L_n^{-(2)}(\hat{\boldsymbol{\theta}}) + o_p(n^{-1}),$$

where $V(\tilde{\boldsymbol{\theta}}) = E\left[(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'|\mathbf{y}\right] = \int(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'p(\boldsymbol{\theta}|\mathbf{y})\,d\boldsymbol{\theta}$ *for any estimator* $\tilde{\boldsymbol{\theta}}$*. Assumption 5 is fairly standard regularity conditions about the Hessians for misspecified models; see Müller (2013).*

The new Bayesian test statistic is defined as:

$$\text{BIMT} = n\int(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}. \quad (2)$$

Let $g(\boldsymbol{\theta}) := n(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})$ be the normalized distance measure between $\boldsymbol{\theta}$ and $\bar{\boldsymbol{\theta}}$ where the distance is in a quadratic form with $\hat{\mathbf{J}}$ being the weighting matrix. Clearly, the proposed test statistic is the posterior mean of $g(\boldsymbol{\theta})$. When the observed-data likelihood function has a close-form expression, its first order derivatives should be easy to compute and hence it is easy to compute BIMT from the MCMC output. When the observed-data likelihood function does not have a close-form expression, we will discuss how to compute BIMT from the MCMC output below. Let us first establish some properties of BIMT.

Theorem 3.1 *Under Assumptions 1-6, we have*

$$BIMT = IOS_a + o_p(1).$$

Corollary 3.2 *Under the null hypothesis that the model is correctly specified, as* $n \to \infty$*,* $BIMT \xrightarrow{p} p$ *where* p *is the dimension of parameter* $\boldsymbol{\theta}$*.*

Remark 3.2 *According to Theorem 3.1, BIMT may be regarded as the Bayesian version of IOS_a of Presnell and Boos (2004). However, there are two important advantage in our statistic over the IOS_a test. The first is that there is no need to invert $\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}})$. Inversion of $\hat{\mathbf{I}}(\hat{\boldsymbol{\theta}})$ may be difficult when the dimension of $\hat{\boldsymbol{\theta}}$ is high. The second is that our test is not only applicable to the iid case but also to the dependent case.*

The intuition why BIMT is asymptotically equivalent to IOS_a is because $E\left[n(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})|\mathbf{y}\right] = -\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})+o_p(1)$. When the model is correctly specified, $-\hat{\mathbf{H}}(\boldsymbol{\theta}_0)=\hat{\mathbf{J}}(\boldsymbol{\theta}_0)$ and BIMT $\xrightarrow{p} p$. When the model is misspecified, it is expectd that $-\hat{\mathbf{H}}(\boldsymbol{\theta}_0)\neq\hat{\mathbf{J}}(\boldsymbol{\theta}_0)$, and hence, BIMT should be different from p.

To implement the BIMT test, threshold values are needed. Unfortunately, since it is a challenge to obtain the finite sample distribution of BIMT in closed-form or the asymptotic distribution of BIMT, obtaining critical values is difficult. A brute force method is to get the threshold values based on Monte Carlo simulations. The detailed steps can be summarized as follows:

Step 1: Set $\boldsymbol{\theta}_0=\bar{\boldsymbol{\theta}}$, based on the model considered, we generate n random observations from the candiate model, then run the MCMC simulations based on simulated data and the candidate model.

Step 2: Based on the MCMC output, compute $\hat{\mathbf{J}}\left(\bar{\boldsymbol{\theta}}^{(1)}\right)$ and $\text{BIMT}^{(1)} = n\int(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}^{(1)})'\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}}^{(1)})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}^{(1)})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}$, where $\bar{\boldsymbol{\theta}}^{(1)}$ is the posterior mean of $\boldsymbol{\theta}$ calculated from the MCMC output.

Step 3: Repeat Step 2 and Step 3 for another M simulated paths and obtain $\text{BIMT}^{(m)}$, $m=1,\ldots,M$.

Step 4: Based on $\left\{BIMT^{(m)}\right\}, m=1,\ldots,M$, we obtain the threshold values at certain probability levels.

This brute force method fits the same model to simulated data by MCMC for M times and hence is time-consuming. If the computing cost is not a concern, we recommend the use of this method for obtaining the critical values. However, if the computational cost is too high, we propose an alternative method to do the specification test based on BIMT. To do so, we first obtain the asymptotic distribution of $g(\boldsymbol{\theta})$.

Theorem 3.3 *Under Assumptions 1-6 and the null hypothesis, as $n\to\infty$, the posterior distribution of $g(\boldsymbol{\theta})$ converges to $\chi^2(p)$.*

Remark 3.3 *Since BIMT is the posterior mean of $g(\boldsymbol{\theta})$ that converges to $\chi^2(p)$ under the null hypothesis, at the significance level of α,*[1] *we can get a Bayesian credible interval from the asymptotic distribution $\chi^2(p)$ such as $[q_{\alpha/2}, q_{1-\alpha/2}]$ where $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are*

[1] The usual choice of significance level is 1%, 5% and 10%.

the $\alpha/2$ and $1-\alpha/2$ quantitles of $\chi^2(p)$.[2] *With the usual choice of significance level (say 1%, 5% or 10%), under the null hypothesis, $[q_{\alpha/2}, q_{1-\alpha/2}]$ includes p and, hence, BMIT asymptotically. Therefore, if BIMT takes a value outside of $[q_{\alpha/2}, q_{1-\alpha/2}]$, the model under the null hypothesis must be misspecified. However, BIMT takes a value inside of $[q_{\alpha/2}, q_{1-\alpha/2}]$, no conclusion can be made.*

4 Latent Variable Models

Given the wide range of applications of latent variable models, we now discuss how to compute BIMT for latent variable models after they are estimated by MCMC. To introduce a latent variable model, let $\mathbf{y} = (y_1, \ldots, y_n)$ denote observed variables and $\mathbf{z} = (z_1, \ldots, z_n)$ denote latent variables. The model is given by

$$\begin{cases} y_t = F(z_t, u_t, \boldsymbol{\theta}) \\ z_t = G(z_{t-1}, v_t, \boldsymbol{\theta}) \end{cases} . \tag{3}$$

The first equation that relates y_t to z_t is the observation equation where u_t is the error term whose distribution is given. The second equation that determines the dynamic of the latent variable is the state equation where v_t is the error term whose distribution is also given. When the distribution of u_t and v_t is Gaussian or the functional form of F and G is linear, the model is referred to as the linear Gaussian state space model. When the distribution of u_t or v_t is non-Gaussian or the functional form of F or G is nonlinear, the model is often referred to as the nonlinear non-Gaussian state space model in the literature.

Let $p(\mathbf{y}|\boldsymbol{\theta})$ be the observed-data likelihood function, and $p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\theta})$ the complete-data likelihood function. Obviously these two functions are related to each other by

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}. \tag{4}$$

The complete-data likelihood function $p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\theta})$ can be expressed as $p(\mathbf{y}|\boldsymbol{z}, \boldsymbol{\theta})p(\boldsymbol{z}|\boldsymbol{\theta})$. Usually analytical expressions for $p(\mathbf{y}|\boldsymbol{z}, \boldsymbol{\theta})$ and $p(\boldsymbol{z}|\boldsymbol{\theta})$ are given by the specification of the model. In particular, the observation equation gives the analytical expression for $p(\mathbf{y}|\boldsymbol{z}, \boldsymbol{\theta})$ while the state equation gives the analytical expression for $p(\boldsymbol{z}|\boldsymbol{\theta})$. However, in general the integral in (4) does not have an analytical expression. Consequently, the statistical inferences, such as estimation and hypothesis testing, are difficult to implement if they are based on the ML approach. For linear Gaussian state space models, $p(\mathbf{y}|\boldsymbol{\theta})$ and its derivatives with respect to $\boldsymbol{\theta}$ can be computed numerically by the Kalman filter. For nonlinear non-Gaussian state space models, other methods are needed to compute $p(\mathbf{y}|\boldsymbol{\theta})$ and the derivatives.

[2] That is $P(\chi^2(p) < q_{\alpha/2}) = \alpha/2$ and $P(\chi^2(p) > q_{1-\alpha/2}) = \alpha/2$.

The latent variables models can be efficiently and easily estimated in the Bayesian framework using MCMC techniques. Let $p(\boldsymbol{\theta})$ be the prior distribution of $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta}|\mathbf{y})$ the posterior distribution of $\boldsymbol{\theta}$. The goal of the Bayesian inference is to obtain $p(\boldsymbol{\theta}|\mathbf{y})$. The data augmentation strategy of Tanner and Wong (1987), that expands the parameter space with the latent variable $\boldsymbol{z}$, is a Bayesian method that uses a MCMC algorithm to generate random samples from the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{z}|\mathbf{y})$. Geweke et al. (2011) reviews algorithms, examples and references for Bayesian estimation of latent variable models.

To implement our test, we still need to calculate $p(\mathbf{y}|\boldsymbol{\theta})$ and its derivatives with respect to $\boldsymbol{\theta}$. It is important to point out that there is no need to optimize $p(\mathbf{y}|\boldsymbol{\theta})$ in our test. Since there is no analytical expression for the observed-data likelihood function for many latent variable models, in this section, we show how to use the EM algorithm, the Kalman filter, and the particle filters to calculate $p(\mathbf{y}|\boldsymbol{\theta})$ and its derivatives with respect to $\boldsymbol{\theta}$.

4.0.1 Computing BIMT by the EM algorithm

The EM algorithm is a powerful tool to deal with latent variable models. Instead of maximizing the observed-data likelihood function, the EM algorithm maximizes the so-called $\mathcal{Q}$ function given by

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = E_{\boldsymbol{\theta}^{(r)}}\{\mathcal{L}_c(\mathbf{y},\boldsymbol{z}|\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^{(r)}\}, \tag{5}$$

where $\mathcal{L}_c(\mathbf{y},\boldsymbol{z}|\boldsymbol{\theta}) := p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\theta})$ is the complete-data likelihood function. The $\mathcal{Q}$-function is the conditional expectation of $\mathcal{L}_c(\mathbf{y},\boldsymbol{z}|\boldsymbol{\theta})$ with respect to the conditional distribution $p(\boldsymbol{z}|\mathbf{y}, \boldsymbol{\theta}^{(r)})$ where $\boldsymbol{\theta}^{(r)}$ is a current fit of the parameter. The EM algorithm consists of two steps: the *expectation* (E) step and the *maximization* (M) step. The E-step evaluates $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$. The M-step determines a $\boldsymbol{\theta}^{(r)}$ that maximizes $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$. Under some mild regularity conditions, for large enough r, $\{\boldsymbol{\theta}^{(r)}\}$ obtained from the EM algorithm is the MLE, $\widehat{\boldsymbol{\theta}}$. For more details about the EM algorithm, see Dempster et al. (1977).

Although the EM algorithm is a good approach to dealing with latent variable models, the numerical optimization in the M-step is often unstable. Not surprisingly, the EM algorithm has been less popular to estimate latent variables models compared with the MCMC techniques. However, we will show that, without using the numerical optimization in the M-step, the theoretical properties of the EM algorithm can facilitate the computation of the proposed test for latent variable models.

Since $p(\mathbf{y}|\boldsymbol{\theta})$ and $\mathbf{s}(\mathbf{y}, \boldsymbol{\theta})$ are not analytically available for latent variable models, we propose to use the EM algorithm to compute $\mathbf{s}(\mathbf{y}, \boldsymbol{\theta})$. For any $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ in Θ, it was shown in Dempster et al. (1977) that

$$\mathbf{s}(\mathbf{y},\boldsymbol{\theta}) = \frac{\partial \mathcal{L}_o(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = E_{(\boldsymbol{z}|\mathbf{y},\boldsymbol{\theta})}\left\{\frac{\partial \mathcal{L}_c(\mathbf{y},\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$$
$$= \int \frac{\partial \mathcal{L}_c(\mathbf{y},\boldsymbol{z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} p(\boldsymbol{z}|\mathbf{y},\boldsymbol{\theta})\mathrm{d}\boldsymbol{z}.$$

If the analytical form of the $\mathcal{Q}$-function is available, we can replace the first derivatives of the log-likelihood function $\log p(\mathbf{y}|\boldsymbol{\theta})$ with the first derivatives of the $\mathcal{Q}$-function. A more general approach to evaluating the $\mathcal{Q}$-function is to use the following formula based on the MCMC output:

$$\mathbf{s}(\mathbf{y},\boldsymbol{\theta}) \approx \frac{1}{M}\sum_{m=1}^{M}\left\{\frac{\partial \log p(\mathbf{y},\boldsymbol{z}^{(m)}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\},$$

where $\{\boldsymbol{z}^{(m)}, m = 1, 2, \ldots, M\}$ is a random sample simulated from the posterior distribution $p(\boldsymbol{z}|\mathbf{y},\boldsymbol{\theta})$.

Although EM algorithm is a very general approach for analyzing latent variable models, it is very cumbersome to deal with dynamic latent variable models, such as, state space models. This is because we have to compute the $\mathbf{s}(\mathbf{y}^{1:t},\boldsymbol{\theta})$ recursively where the posterior sampling has to be implemented for n times (Doucet and Shephard, 2012). As a result, it is computationally demanding although some parallel computing techniques may be used. Alternatively, one can compute $\mathbf{s}(\mathbf{y},\boldsymbol{\theta})$ using the Kalman filter and the particle filters.

4.0.2 Computing BIMT by the Kalman filter

In economics, many time series models can be represented by a linear Gaussian state space form. The Kalman filter is an efficient recursive method for computing the optimal linear forecasts in such models. It also gives the exact likelihood function of the model. Here, we only present the basic idea of the Kalman filter for analyzing liner state space models. One may refer to Harvey (1989) for the detailed textbook treatment.

Consider a general linear state space models,

$$\begin{aligned} z_t &= Tz_{t-1} + R\varepsilon_t, \\ y_t &= D + Cz_t + \xi_t, \end{aligned}$$

where $\varepsilon_t \sim N(0, Q)$, $\xi_t \sim N(0, H)$, T is $n_s \times n_s$, R is $n_s \times n_e$, D is $n \times 1$, C is $n \times n_s$, Q is $n_e \times n_e$, H is $n \times n$. These six coefficient matrices are functions of a vector of parameters $\boldsymbol{\theta}$ which is $n_q \times 1$.

Let $\mathbf{y}^s = (y_1, y_2, \ldots, y_s)$, $z_t^s = E(z_t|\mathbf{y}^s)$, $\Sigma_t^s = E\{(z_t - z_t^s)(z_t - z_t^s)'|\mathbf{y}^s\}$. With the initial conditions, z_0^0 and Σ_0^0, for $t = 1, 2, \ldots, n$, the Kalman filter recursively implements the following steps

$$
\begin{aligned}
z_t^{t-1} &= T z_{t-1}^{t-1}, \\
\Sigma_t^{t-1} &= T \Sigma_{t-1}^{t-1} T' + RQR',
\end{aligned}
$$

and

$$
\begin{aligned}
z_t^t &= z_t^{t-1} + K_t \left(y_t - D - C z_t^{t-1} \right), \\
\Sigma_t^t &= \left[I_{n_s} - K_t C \right] \Sigma_t^{t-1},
\end{aligned}
$$

where

$$
K_t = \Sigma_t^{t-1} C' \left[C \Sigma_t^{t-1} C' + H \right]^{-1}.
$$

The observed-data log-likelihood is given by

$$
\begin{aligned}
\log p(\mathbf{y}|\boldsymbol{\theta}) &= -\sum_{t=1}^{n} \left[\frac{n}{2} \log 2\pi + \frac{1}{2} \log |F_t| + \frac{1}{2} \left(y_t - D - C z_t^{t-1} \right)' F_t^{-1} \left(y_t - D - C z_t^{t-1} \right) \right] \\
&= -\sum_{t=1}^{n} \left[\frac{n}{2} \log 2\pi + \frac{1}{2} \log |F_t| + \frac{1}{2} \omega_t' F_t^{-1} \omega_t \right],
\end{aligned}
$$

where $F_t = C P_t^{t-1} C' + H$, $\omega_t = y_t - D - C z_t^{t-1}$. Clearly, $\log p(\mathbf{y}|\boldsymbol{\theta})$ has to be calculated recursively since F_t and z_t^{t-1} are only available recursively. Similarly, $\mathbf{s}_t(\boldsymbol{\theta})$ has to be computed recursively. To calculate $\mathbf{s}_t(\boldsymbol{\theta})$ and the first order derivatives of $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$, we need to calculate the first order derivatives of $|F_t|$, $\omega_t' F_t^{-1} \omega_t$ recursively. In Appendix 4, we gives the expression of the relevant first order derivatives that are used to compute BIMT.

4.0.3 Computing BIMT by particle filters

In practice, the phenomenon of non-Gaussianity or non-linearity is often found. Consequently, the nonlinear non-Gausian state space models have been widely used in empirical works. However, they cannot be analyzed using the Kalman filter. Instead, one can use another recursive filtering algorithm known as particle filters. We only present the basic idea of particle filters here and refer the reader to recent review papers on particle filter by Doucet and Johansen (2009) and Creal (2012) for greater details.

Let $z_{t+1}|z_t \sim f\left(z_{t+1}|z_t, \boldsymbol{\theta}\right)$ and $y_t|z_t \sim g\left(y_t|z_t, \boldsymbol{\theta}\right)$. Let the initial density of z be $\mu\left(z|\boldsymbol{\theta}\right)$. The joint density of $\left(\mathbf{z}^t, \mathbf{y}^t\right)$ is

$$
p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right) = \mu\left(z_1|\boldsymbol{\theta}\right) \prod_{k=2}^{t} f\left(z_k|z_{k-1}, \boldsymbol{\theta}\right) \prod_{k=1}^{t} g\left(y_k|z_k, \boldsymbol{\theta}\right),
$$

and hence

$$p\left(\mathbf{y}^t|\boldsymbol{\theta}\right) = \int p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right) d\mathbf{z}^t.$$

For nonlinear and non-Gaussian state space models, neither $p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right)$ nor $p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)$ are available in closed-form. The goal here is to calculate $p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right)$, $p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)$, and $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$ sequentially for $t = 1, \ldots, n$. The idea of the using particle filters is to approximate the conditional probability distribution $p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t$ by its empirical measure. An example of particle filters is the Sequential Important Sampling and Resampling (SISR) algorithm which iterates the following step for $i = 1, \ldots, N$,

Step 1: At $t = 1$, $z_1^{(i)} \sim \mu(\cdot)$,

$$w_1\left(\mathbf{z}^{1(i)}\right) = \frac{\mu\left(z_1^{(i)}|\boldsymbol{\theta}\right) g\left(y_1|z_1^{(i)}, \boldsymbol{\theta}\right)}{q_1\left(z_1^{(i)}\right)}, \quad W_1^{(i)} = \frac{w_1\left(\mathbf{z}^{1(i)}\right)}{\sum_{i=1}^N w_1\left(\mathbf{z}^{1(i)}\right)},$$

$\mathbf{z}^{1(i)} = z_1^{(i)}$. Resample $\left(W_1^{(i)}, \mathbf{z}^{1(i)}\right)$ to obtain new particles $\left(\frac{1}{N}, \widetilde{\mathbf{z}}^{1(i)}\right)$.

Step 2: At $t \geq 2$, $z_t^{(i)} \sim q_n\left(\cdot|\widetilde{\mathbf{z}}^{t-1(i)}\right)$,

$$w_t\left(\mathbf{z}^{t(i)}\right) = \frac{f\left(z_t^{(i)}|\widetilde{z}_{t-1}^{(i)}, \boldsymbol{\theta}\right) g\left(y_t|\widetilde{z}_t^{(i)}, \boldsymbol{\theta}\right)}{q_t\left(z_t^{(i)}|\widetilde{\mathbf{z}}^{t-1(i)}\right)}, \quad W_t^{(i)} = \frac{w_t\left(\mathbf{z}^{t(i)}\right)}{\sum_{i=1}^N w_t\left(\mathbf{z}^{t(i)}\right)},$$

$\mathbf{z}^{t(i)} = \left(\widetilde{\mathbf{z}}^{t-1(i)}, z_t^{(i)}\right)$. Resample $\left(W_t^{(i)}, \mathbf{z}^{t(i)}\right)$ to obtain new particles $\left(\frac{1}{N}, \widetilde{\mathbf{z}}^{t(i)}\right)$.

Step 3: Approximate the conditional distribution $p_{\boldsymbol{\theta}}\left(d\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right)$ by its empirical measure

$$\widehat{p}\left(d\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) = \sum_{i=1}^N W_t^{(i)} \delta_{\mathbf{z}^{t(i)}}\left(d\mathbf{z}^t\right) \quad \text{or} \quad \widetilde{p}_{\boldsymbol{\theta}}\left(d\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^N \delta_{\widetilde{\mathbf{z}}^{t(i)}}\left(d\mathbf{z}^t\right),$$

and

$$\widehat{p}\left(y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^N w_t\left(\mathbf{z}^{t(i)}\right),$$

where N is the number of particles and $q_t(\cdot|\cdot)$ is the proposal density.

With the empirical measures $\left\{\widehat{p}\left(d\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right)\right\}_{t=1:n}$, we can approximate the integral

$$I_t = \int \varphi_t\left(\mathbf{z}^t\right) p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t,$$

by

$$\widehat{I}_t = \int \varphi_t\left(\mathbf{z}^t\right) \widehat{p}\left(d\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) = \sum_{i=1}^N W_t^{(i)} \varphi_t\left(\mathbf{z}^{t(i)}\right),$$

for $t = 1, \cdots, n$, where $\varphi_t\left(\mathbf{z}^t\right)$ is the target function. If one chooses $\varphi_t\left(\mathbf{z}^t\right) = \partial \log p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}$, then it is easy to show that

$$\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta}) = \int \varphi_t\left(\mathbf{z}^t\right) p\left(\mathbf{z}^t | \mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t.$$

Therefore, $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$ can be obtained recursively.

Based on the different proposal density $q_t(\cdot|\cdot)$, different particle filtering algorithms have been proposed in the literature, including the bootstrap particle filters of Gordon et al. (1993) and the auxiliary particle filters of Pitt and Shephard (1999). In this paper, we use the auxiliary particle filters to compute $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$ and the proposed test statistic. Appendix 5 gives the details about how to compute $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$ using the particle filters.

5 Empirical Examples

We now illustrate the proposed test to do specification analysis in three real examples. The first example is the well-known dynamic factor model. This is a linear state space model and the Kalman filter can be used to compute the proposed test statistic. The second is the linearized DSGE model. The third is the stochastic volatility model. This is a nonlinear non-Gaussian state space model and we use the particle filters to compute the test statistic.

5.1 A dynamic factor model

Stock and Watson (1992) developed a single-factor model to explain the comovements in many macroeconomic variables for the purpose of building a coincident economic indicator. Let Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t} be the logarithmic industrial production, personal income less transfer payments, total manufacturing and trade sales, and employees on nonagricultural payrolls. Stock and Watson (1992) considered the following dynamic factor model in the first difference form,

$$\begin{aligned}
\Delta Y_{it} &= D_i + \gamma_i \Delta C_t + e_{it}, \quad i = 1, 2, 3, \\
\Delta Y_{4t} &= D_4 + \gamma_{40}\Delta C_t + \gamma_{41}\Delta C_{t-1} + \gamma_{42}\Delta C_{t-2} + \gamma_{43}\Delta C_{t-3} + e_{4t}, \\
(\Delta C_t - \delta) &= \phi_1\left(\Delta C_{t-1} - \delta\right) + \phi_2\left(\Delta C_{t-2} - \delta\right) + w_t, \quad w_t \overset{i.i.d.}{\sim} N\left(0, 1\right), \\
e_{it} &= \psi_{i1} e_{it-1} + \psi_{i2} e_{it-2} + \varepsilon_{it}, \ \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_i^2\right), \ i = 1, 2, 3, 4,
\end{aligned}$$

where ΔC_t is the common factor. To avoid the identification problem, the model in the deviation form was considered,

$$\begin{aligned}
\Delta y_{it} &= \gamma_i \Delta c_t + e_{it}, \quad i = 1, 2, 3, \\
\Delta y_{4t} &= \gamma_{40}\Delta c_t + \gamma_{41}\Delta c_{t-1} + \gamma_{42}\Delta c_{t-2} + \gamma_{43}\Delta c_{t-3} + e_{4t}, \\
\Delta c_t &= \phi_1 \Delta c_{t-1} + \phi_2 \Delta c_{t-2} + w_t, \quad w_t \overset{i.i.d.}{\sim} N\left(0, 1\right), \\
e_{it} &= \psi_{i1} e_{it-1} + \psi_{i2} e_{it-2} + \varepsilon_{it}, \ \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_i^2\right), \ i = 1, 2, 3, 4,
\end{aligned}$$

where $\Delta y_{it} = Y_{it} - \Delta \overline{Y}_i$ and $\Delta c_t = \Delta C_t - \delta$. Following Kim and Nelson (1999), a state space representation of the model is given by

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \\ \Delta y_{4t} \end{bmatrix} = \begin{bmatrix} \gamma_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \gamma_3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \gamma_{40} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \\ e_{1t} \\ e_{1t-1} \\ e_{2t} \\ e_{2t-1} \\ e_{3t} \\ e_{3t-1} \\ e_{4t} \\ e_{4t-1} \end{bmatrix},$$

and

$$\begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \\ e_{1t} \\ e_{1t-1} \\ e_{2t} \\ e_{2t-1} \\ e_{3t} \\ e_{3t-1} \\ e_{4t} \\ e_{4t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \psi_{11} & \psi_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \psi_{21} & \psi_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \psi_{31} & \psi_{32} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \psi_{41} & \psi_{42} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta c_{t-1} \\ \Delta c_{t-2} \\ \Delta c_{t-3} \\ e_{1t} \\ e_{1t-1} \\ e_{2t} \\ e_{2t-1} \\ e_{3t} \\ e_{3t-1} \\ e_{4t} \\ e_{4t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \\ 0 \\ 0 \\ \varepsilon_{1t} \\ 0 \\ \varepsilon_{2t} \\ 0 \\ \varepsilon_{3t} \\ 0 \\ \varepsilon_{4t} \\ 0 \end{bmatrix},$$

where the parameter vector

$$\boldsymbol{\theta} = \left(\gamma_1, \gamma_2, \gamma_3, \gamma_{40}, \gamma_{41}, \gamma_{42}, \gamma_{43}, \phi_1, \phi_2, \psi_{11}, \psi_{12}, \psi_{21}, \psi_{22}, \psi_{31}, \psi_{32}, \psi_{41}, \psi_{42}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2\right)'.$$

To carry out Bayesian test of the hypothesis, we use the data that consist of the four coincident variables of U.S. from January 1959 to January 1995. The priors of parameters are specified as in Kim and Nelson (1999), we draw 10,000 samples from the posterior distribution, discard the first 2,000 as build-in period, and store the remaining samples as effective observations. The analytical derivatives of the linear Gaussian state space model are derived in Appendix 4. Based on the 80,00 random observations, we get BIMT = 74.3761.

To test whether this model is misspecified or not, based on the $\chi^2(p)$ distribution with $p = 21$, the symmetric credible interval is [11.59, 32.67] at the 10% significance level, [10.28, 35.48] at the 5% significance level, and [8.03, 41.40] at the 1% significance level.

Obviously BIMT falls outside of the credible intervals and rejects the null hypothesis at the three significance levels, suggesting that the model is misspecified.

5.2 Specification test in DSGE models

DSGE models are microfounded and optimization-based. They have become very popular in maceroeconomics over the last 30 years. Estimation and evaluation of the DSGE models require one to solve them and then to construct a linear or nonlinear state-space approximation. Bayesian time series methods have been widely applied to estimate the DSGE models. For a linear Gaussian approximation, the Kalman filter can be used to compute the likelihood function numerically; see Schorfheide (2001), Lubik and Schorfheide (2006), An and Schorfheide (2007). For a non-linear non-Gaussian approximation, Fernández-Villaverde and Rubio-Ramírez (2005) used the particle filter to calculate the likelihood numerically.

In this example, following An and Schorfheide (2007), we adopt a linear Gaussian approximation. We estimate a simple DSGE model of Clarida, Gali and Gertler (1999), which is also used in Andrews and Mikusheva (2013) to study the weak identification problem in DSGE models. The model is given by

$$bE_t\pi_{t+1} + \kappa x_t - \pi_t + \varepsilon_t = 0,$$

$$-[r_t - E_t\pi_{t+1} - \rho\Delta a_t] + E_t x_{t+1} - x_t = 0,$$

$$\lambda r_{t-1} + (1-\lambda)\,\phi_\pi \pi_t + (1-\lambda)\,\phi_x x_t + u_t = r_t,$$

$$\Delta a_t = \rho\Delta a_{t-1} + \varepsilon_{a,t},$$

$$\Delta u_t = \delta\Delta u_{t-1} + \varepsilon_{u,t}.$$

where π_t, r_t, x_t are the output growth rate, the inflation rate, and the interest rate, respectively, Δa_t and Δu_t are the unobserved shocks, and

$$(\varepsilon_t, \varepsilon_{a,t}, \varepsilon_{u,t})' \sim iidN\,(0, \Sigma) \text{ with } \Sigma = diag(\sigma^2, \sigma_a^2, \sigma_u^2).$$

Following Andrews and Mikusheva (2013), we set $b = 0.99$.

The data are from Lubik and Schorfheide (2006) with π_t, r_t, x_t being quaterly U.S. series on the GDP growth rates, the inflation rates, and the nominal interest rates from the first quarter of 1983 to the last quarter of 2002. The priors are the same as in Smets and Wouters (2007).

To aviod the well-documented weak identification problem in DSGE models (see, for example, Canova and sala (2009), Guerron-Quintana, Inoue and Kilian (2013), Iskrev (2010), and Mavroeidis (2005)), we fix parameters $\phi_x = 2.28$, $\phi_\pi = 2.02$, $\lambda = 0.898$, $\kappa =$

0.1 as they are potential sources of weak identification (Andrews and Mikusheva, 2013). Following Schorfheide (2001), we estimate the model by the random walk Metropolis-Hasting algorithm in the MATLAB-based DYNARE package (Adjemian et al., 2011). We draw 20,000 random observations from the posterior distribution with the first 10,000 draws being discarded as burning-in observations. All the quantities needed to compute BIMT are derived in Appendix 6. Based on the 10,000 random observations, we get BIMT = 50.1630.

To test whether this model is misspecified, based on the $\chi^2(p)$ distribution with $p = 5$, the symmetric credible interval is [1.145, 11.070] at the 10% significance level, [0.831, 12.833] at the 5% significance level, and [0.412, 16.750] at the 1% significance level. Obviously BIMT falls outside of the credible intervals and rejects the null hypothesis at the three significance levels, suggesting that the model is misspecified.

5.3 A stochastic volatility model

The stochastic volatility (SV) model introduced by Tauchen and Pitts (1983) and Talyor (1982) is used to describe financial time series. The SV model involves two noise processes, one for the observation, and one for the latent volatility. Given that the heavy tails are usually found in distributions of returns, following Abanto-Valle et al. (2010), we generalized the SV model with the normal distribution to the SV model with scale mixtures of normal distributions (SV-SMN):

$$\begin{aligned} y_t|h_t &= \exp(h_t/2)\,\lambda_t^{-1/2}u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1),\ \lambda_t \overset{i.i.d.}{\sim} \Gamma\left(\frac{\upsilon}{2},\frac{\upsilon}{2}\right),\ t=1,\ldots,n, \\ h_t|h_{t-1},\mu,\phi,\tau^2 &= \mu+\phi(h_{t-1}-\mu)+\tau\nu_t,\quad \nu_t \overset{i.i.d.}{\sim} N(0,1),\ t=1,\ldots,n, \end{aligned}$$

where y_t is the return at time t, h_t is the return volatility at period t, u_t and υ_t are uncorrelated, and $h_0 = \mu$, $\upsilon = 3$.

To carry out Bayesian test of the hypothesis, we fit the SV-SMN model to mean-corrected daily returns on Pound/Dollar exchange rates from 01/10/81 to 28/06/85. We first estimate the model using the Bayesian MCMC method using the following vague priors:

$$\mu \sim N[0,100],\quad \phi \sim Beta[1,1],\quad \tau^{-2} \sim \Gamma[0.001, 0.001].$$

We draw 110,000 from the posterior distribution and discard the first 10,000 as burning-in observations, and store the remaining samples as effective observations. On the basis of particle filters, using the approach shown in Appendix 5, we can compute out the Bayesian test statistic. Based on the 10,000 random observations, we can get BIMT = 38.7143.

To test whether this model is misspecified, based on the $\chi^2(p)$ distribution with $p = 3$, the symmetrical credible interval is [0.3518, 7.8147] at the 10% significance level, [0.2158, 9.3484] at the 5% significance level, and [0.0717, 12.8382] at the 1% significance level. Obviously we reject the null hypothesis that the model is correctly specified at the three significance levels.

6 Conclusions

In this paper, we have proposed a new Bayesian test statistic to assess the adequacy of specification of a model after the model is estimated by Bayesian MCMC methods. The test statistic is a Bayesian mean of a quadratic form. Under the regularity conditions, we show that it asymptotically approaches the IOS_a statistic of Presnell and Boos (2004).

The main advantages of the new statistic can be summarized as follows: (1) it is quite general and can be applied to a variety of models, including models that are difficult to estimate by frequentist methods such as models with latent variable; (2) it is easy to compute; (3) there is no need to specify an alternative hypothesis. We illustrate the new method in the context of three popular models, the dynamic factor model, the DSGE model and the heavy tailed stochastic volatility model. We can reject the single factor model, the simple DSGE model and the heavy tailed stochastic volatility model using real data.

7 Appendix

7.1 Appendix 1: Proof of Theorem 3.1

Under Assumption 6, we get

$$\frac{1}{n}L_n^{(2)}(\boldsymbol{\theta}) = \frac{1}{n}\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{1}{n}\frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{1}{n}\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + O_p(n^{-1}) = \hat{\mathbf{H}}(\boldsymbol{\theta}) + o_p(1),$$

Using the first order Taylor expansion, for $t = 1, 2, \ldots, n$, we can have

$$s_t(\hat{\boldsymbol{\theta}}) = s_t(\boldsymbol{\theta}_0) + h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\tilde{\boldsymbol{\theta}}_0$ lies on the segment between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Then, we can get that

$$\begin{aligned} s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})' &= \left[s_t(\boldsymbol{\theta}_0) + h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right]\left[s_t(\boldsymbol{\theta}_0) + h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right]' \\ &= s_t(\boldsymbol{\theta}_0)s_t(\boldsymbol{\theta}_0)' + 2h_t(\tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)s_t(\boldsymbol{\theta}_0)' + h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'h_t(\tilde{\boldsymbol{\theta}}_0). \end{aligned}$$

Since $\hat{\boldsymbol{\theta}}$ is the consistent estimator of $\boldsymbol{\theta}_0$, there exists a real sequence $k_n \to 0$ such that $\tilde{\boldsymbol{\theta}}_0 \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le k_n\}$ for enough large n. According to Assumption 5, we can get

$$\frac{1}{n}\sum_{t=1}^{n} h_t(\tilde{\boldsymbol{\theta}}_0) = \frac{1}{n}\sum_{t=1}^{n} h_t(\boldsymbol{\theta}_0) + o_p(1) = O_p(1).$$

Let $\hat{\boldsymbol{\theta}}_{ML}$ be the ML estimator of $\boldsymbol{\theta}$. Using the Taylor expansion, $p(\boldsymbol{\theta}) = O_p(1)$, we get

$$0 = \frac{\partial \ln p(\mathbf{y}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \ln p(\mathbf{y}, \hat{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \ln p(\mathbf{y}, \tilde{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{ML}),$$

where $\tilde{\boldsymbol{\theta}}_{ML}$ lies on the segment between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{ML}$. Thus, under the regularity conditions, we have

$$\begin{aligned}
\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{ML} &= \left[\frac{\partial^2 \ln p(\mathbf{y}, \tilde{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \frac{\partial \ln p(\mathbf{y}, \hat{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta}} \\
&= L_n^{-(2)} \left(\tilde{\boldsymbol{\theta}}_{ML} \right) \left[\frac{\partial \ln p(\mathbf{y} | \hat{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta}} + \frac{\partial \ln p(\hat{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta}} \right] \\
&= L_n^{-(2)} \left(\tilde{\boldsymbol{\theta}}_{ML} \right) \left[0 + \frac{\partial \ln p(\hat{\boldsymbol{\theta}}_{ML})}{\partial \boldsymbol{\theta}} \right] \\
&= O_p(n^{-1}) O_p(1) = O_p(n^{-1}).
\end{aligned}$$

Hence, according to the standard ML likelihood theory, $\hat{\boldsymbol{\theta}}$ is also the consistent estimator of $\boldsymbol{\theta}_0$. From the standard ML theory, we know that $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_{ML} + O_p(n^{-1/2})$, hence, $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}} + O_p(n^{-1}) + O_p(n^{-1/2}) = \hat{\boldsymbol{\theta}} + O_p(n^{-1/2})$.

Then, we can show that

$$\begin{aligned}
& \frac{1}{n} || \sum_{t=1}^{n} s_t(\hat{\boldsymbol{\theta}}) s_t(\hat{\boldsymbol{\theta}})' - \sum_{t=1}^{n} s_t(\boldsymbol{\theta}_0) s_t(\boldsymbol{\theta}_0)' || = \frac{1}{n} || \sum_{t=1}^{n} \left[s_t(\hat{\boldsymbol{\theta}}) s_t(\hat{\boldsymbol{\theta}})' - s_t(\boldsymbol{\theta}_0) s_t(\boldsymbol{\theta}_0)' \right] || \\
\leq\ & \frac{1}{n} \sum_{t=1}^{n} || 2 h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) s_t(\boldsymbol{\theta}_0)' || + \frac{1}{n} \sum_{t=1}^{n} || h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' h_t(\tilde{\boldsymbol{\theta}}_0) || \\
\leq\ & \left[\frac{1}{n} \sum_{t=1}^{n} || h_t(\tilde{\boldsymbol{\theta}}_0) || \right] \left[2 || (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) || \sup_{t \leq n} || s_t(\boldsymbol{\theta}_0)' || \right] \\
& + \left[\frac{1}{n} \sum_{t=1}^{n} || h_t(\tilde{\boldsymbol{\theta}}_0) || \right] \left[|| (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) ||^2 \sup_{t \leq n} || h_t(\tilde{\boldsymbol{\theta}}_0) || \right] \\
=\ & 2 O_p(1) O_p\left(n^{-1/2}\right) o_p\left(n^{1/2}\right) + O_p(1) O_p\left(n^{-1}\right) o_p\left(n\right) = o_p(1) \\
& \frac{1}{\sqrt{n}} \sup_{t \leq n} || s_t(\hat{\boldsymbol{\theta}}) - s_t(\boldsymbol{\theta}_0) || = \frac{1}{\sqrt{n}} \sup_{t \leq n} || h_t(\tilde{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) || = \left[\frac{1}{\sqrt{n}} || (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) || \sup_{t \leq n} || h_t(\tilde{\boldsymbol{\theta}}_0) || \right] \\
=\ & \frac{1}{\sqrt{n}} O_p(n^{-1/2}) o_p(n) = o_p(1).
\end{aligned}$$

Hence, using Assumption 6, we can get that

$$\begin{aligned}
\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) &= \frac{1}{n} \sum_{t=1}^{n} s_t(\hat{\boldsymbol{\theta}}) s_t(\hat{\boldsymbol{\theta}}_t)' = \frac{1}{n} \sum_{t=1}^{n} s_t(\theta_0) s_t(\theta_0)' + o_p(1) = O_p(1) \\
\frac{1}{\sqrt{n}} \sup_{t \leq n} || s_t(\hat{\boldsymbol{\theta}}) || &= \frac{1}{\sqrt{n}} \sup_{t \leq n} || s_t(\boldsymbol{\theta}_0) || + o_p(1) = o_p(1)
\end{aligned}$$

Similarly, using the first order Taylor expansion for $s_t(\bar{\boldsymbol{\theta}})$, we have

$$s_t(\bar{\boldsymbol{\theta}}) = s_t(\boldsymbol{\theta}_0) + h_t(\tilde{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\tilde{\boldsymbol{\theta}}$ lies on the segment between $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. It is noted that $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}} + O_p(n^{-1/2})$ and $\bar{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} + o_p(n^{-1/2})$. Hence, $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + O_p(n^{-1/2}) = \boldsymbol{\theta}_0 + o_p(1)$ so that $\bar{\boldsymbol{\theta}}$ is also the consistent estimator of $\boldsymbol{\theta}_0$, there exists a real sequence $k_n \to 0$ such that $\tilde{\boldsymbol{\theta}} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq k_n\}$ for large enough n. Hence, similarly with above proof, we can get

$$\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} s_t(\bar{\boldsymbol{\theta}})s_t(\bar{\boldsymbol{\theta}}_t)' = \frac{1}{n}\sum_{t=1}^{n} s_t(\theta_0)s_t(\theta_0)' + o_p(1) = O_p(1).$$

Furthermore, using the first order Taylor expansion, we have

$$s_t(\bar{\boldsymbol{\theta}}) = s_t(\hat{\boldsymbol{\theta}}) + h_t(\tilde{\boldsymbol{\theta}}_1)(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}),$$

where $\tilde{\boldsymbol{\theta}}_1$ lies on the segment between $\bar{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$. Then, we have

$$\begin{aligned} & s_t(\bar{\boldsymbol{\theta}})s_t(\bar{\boldsymbol{\theta}})' = \left[s_t(\hat{\boldsymbol{\theta}}) + h_t(\tilde{\boldsymbol{\theta}}_1)(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})\right]\left[s_t(\hat{\boldsymbol{\theta}}) + h_t(\tilde{\boldsymbol{\theta}}_1)(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})\right]' \\ = \; & s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})' + 2h_t(\tilde{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})' + h_t(\tilde{\boldsymbol{\theta}}_1)(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'h_t(\tilde{\boldsymbol{\theta}}_1). \end{aligned}$$

Using Remark 3.1, we can show that

$$\begin{aligned} & \frac{1}{n}||\sum_{t=1}^{n} s_t(\bar{\boldsymbol{\theta}})s_t(\bar{\boldsymbol{\theta}})' - \sum_{t=1}^{n} s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})'|| = \frac{1}{n}||\sum_{t=1}^{n}\left[s_t(\bar{\boldsymbol{\theta}})s_t(\bar{\boldsymbol{\theta}})' - s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})'\right]|| \\ \leq \; & \frac{1}{n}\sum_{t=1}^{n}||2h_t(\tilde{\boldsymbol{\theta}}_1)(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})'|| + \frac{1}{n}\sum_{t=1}^{n}||h_t(\tilde{\boldsymbol{\theta}}_1)(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'h_t(\tilde{\boldsymbol{\theta}}_1)|| \\ \leq \; & \left[\frac{1}{n}\sum_{t=1}^{n}||h_t(\tilde{\boldsymbol{\theta}}_1)||\right]\left[2||(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})||\sup_{t\leq n}||s_t(\hat{\boldsymbol{\theta}})'||\right] \\ & + \left[\frac{1}{n}\sum_{t=1}^{n}||h_t(\tilde{\boldsymbol{\theta}}_1)||\right]\left[||(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})||^2\sup_{t\leq n}||h_t(\tilde{\boldsymbol{\theta}}_1)||\right] \\ = \; & 2O_p(1)o_p\left(n^{-1/2}\right)o_p\left(n^{1/2}\right) + O_p(1)o_p\left(n^{-1}\right)o_p\left(n\right) = o_p(1). \end{aligned}$$

Then, we get

$$\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} s_t(\bar{\boldsymbol{\theta}})s_t(\bar{\boldsymbol{\theta}}_t)' = \frac{1}{n}\sum_{t=1}^{n} s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})' + o_p(1) = O_p(1)$$

According to Assumption 5, we get

$$\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} h_t(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} h_t(\boldsymbol{\theta}_0) + o_p(1) = O_p(1).$$

Furthermore, using Remark 3.1 once again, we get

$$
\begin{aligned}
& E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right] = E\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right] \\
= \; & E\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'|\mathbf{y}\right] + 2E\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})|\mathbf{y}\right](\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})' \\
= \; & E\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'|\mathbf{y}\right] + 2(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})' \\
= \; & E\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'|\mathbf{y}\right] - (\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}})' \\
= \; & -L_n^{-(2)}(\hat{\boldsymbol{\theta}}) + o_p(n^{-1}) + o_p(n^{-\frac{1}{2}})o_p(n^{-\frac{1}{2}}) \\
= \; & -\left[n\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right]^{-1} + o_p(n^{-1}) \\
= \; & -\frac{1}{n}\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}}) + o_p(n^{-1}).
\end{aligned}
$$

Hence, using Remark 3.1 once again, we have

$$
\begin{aligned}
& BIMT = n\mathbf{tr}\left\{\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}})E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]\right\} \\
= \; & n\mathbf{tr}\left\{\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}})E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]\right\} \\
= \; & n\mathbf{tr}\left\{\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) + o_p(1)\right]E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]\right\} \\
= \; & \mathbf{tr}\left\{\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) + o_p(1)\right]E\left[n(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]\right\} \\
= \; & \mathbf{tr}\left\{\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) + o_p(1)\right]\left[-\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}}) + o_p(1)\right]\right\} \\
= \; & -\mathbf{tr}\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})\right] + \mathbf{tr}\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})o_p(1)\right] + \mathbf{tr}\left[\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})o_p(1)\right] + o_p(1) \\
= \; & \mathbf{tr}\left[-\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})\right] + o_p(1) = IOS_a + o_p(1).
\end{aligned}
$$

Theorem 3.1 is proven.

7.2 Appendix 2: Proof of Corollary 3.2

When the model is correctly specified, the pseudo-true value $\boldsymbol{\theta}_0$ is the true value. Let $\mathbf{H}(\boldsymbol{\theta}_0) = \int \hat{\mathbf{H}}(\boldsymbol{\theta}_0)g(\mathbf{y})\mathrm{d}\mathbf{y}$. Using the central limit theorem, we get $\hat{\mathbf{H}}(\boldsymbol{\theta}_0) = \mathbf{H}(\boldsymbol{\theta}_0) + o_p(1)$. Furthermore, by Theorem 3.1, we get $\hat{\mathbf{H}}(\boldsymbol{\theta}_0) = \hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) + o_p(1)$ and

$$
\begin{aligned}
& -\frac{1}{n}\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}}) = -\frac{1}{n}\left[\hat{\mathbf{H}}(\boldsymbol{\theta}_0) + o_p(1)\right]^{-1} + o_p(n^{-1}) \\
= \; & -\frac{1}{n}\left[\hat{\mathbf{H}}(\boldsymbol{\theta}_0)\right]^{-1} + o_p(n^{-1}) \\
= \; & -\frac{1}{n}\left[\mathbf{H}(\boldsymbol{\theta}_0) + o_p(1)\right]^{-1} + o_p(n^{-1}) \\
= \; & -\frac{1}{n}\left[\mathbf{H}(\boldsymbol{\theta}_0)\right]^{-1} + o_p(n^{-1}).
\end{aligned}
$$

Hence, we can get that $\mathbf{H}(\boldsymbol{\theta}_0) = O(1)$.

Let $\mathbf{J}(\boldsymbol{\theta}_0) = \int \hat{\mathbf{J}}(\boldsymbol{\theta}_0)g(\mathbf{y})\mathrm{d}\mathbf{y}$. Using the central limit theorem, we get $\hat{\mathbf{J}}(\boldsymbol{\theta}_0) = \mathbf{J}(\boldsymbol{\theta}_0) + o_p(1)$. Then, using Assumption 5 and Theorem 3.1, we get

$$\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}})' = \hat{\mathbf{J}}(\boldsymbol{\theta}_0) + o_p(1),$$

and $\mathbf{J}(\boldsymbol{\theta}_0) = O(1)$.

When the model is correctly specified, according to the information matrix equality, we have $\mathbf{J}(\boldsymbol{\theta}_0) = -\mathbf{H}(\boldsymbol{\theta}_0)$ (White, 1996). Therefore, we get

$$\begin{aligned} & BIMT = \mathbf{tr}\left[-\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}^{-1}(\hat{\boldsymbol{\theta}})\right] + o_p(1) \\ = & \ \mathbf{tr}\left\{-\left[\hat{\mathbf{J}}(\boldsymbol{\theta}_0) + o_p(1)\right]\left[\mathbf{H}^{-1}(\boldsymbol{\theta}_0) + o_p(1)\right]\right\} + o_p(1) \\ = & \ \mathbf{tr}\left\{-\left[\mathbf{J}(\boldsymbol{\theta}_0) + o_p(1)\right]\left[\mathbf{H}^{-1}(\boldsymbol{\theta}_0) + o_p(1)\right]\right\} + o_p(1) \\ = & \ \mathbf{tr}\left[-\mathbf{J}(\boldsymbol{\theta}_0)\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right] + \mathbf{tr}\left[-\mathbf{J}(\boldsymbol{\theta}_0)o_p(1)\right] + \mathbf{tr}\left[-o_p(1)\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right] \\ = & \ \mathbf{tr}\left[-\mathbf{J}(\boldsymbol{\theta}_0)\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right] + o_p(1) = p + o_p(1). \end{aligned}$$

Corollary 3.2 is proven.

7.3 Appendix 3: Proof of Theorem 3.3

According to Theorem 3.1 and Corollary 3.2, we get

$$\begin{aligned} \hat{\mathbf{J}}(\bar{\boldsymbol{\theta}}) &= \frac{1}{n}\sum_{t=1}^{n} s_t(\bar{\boldsymbol{\theta}})s_t(\bar{\boldsymbol{\theta}}_t)' = \frac{1}{n}\sum_{t=1}^{n} s_t(\hat{\boldsymbol{\theta}})s_t(\hat{\boldsymbol{\theta}}_t)' + o_p(1) \\ &= \frac{1}{n}\sum_{t=1}^{n} s_t(\theta_0)s_t(\theta_0)' + o_p(1) = \mathbf{J}(\boldsymbol{\theta}_0) + o_p(1), \end{aligned}$$

where $\mathbf{J}(\boldsymbol{\theta}_0) = \int \hat{\mathbf{J}}(\boldsymbol{\theta}_0)g(\mathbf{y})\mathrm{d}\mathbf{y}$. According to Assumption 5, we get

$$\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} h_t(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n} h_t(\boldsymbol{\theta}_0) + o_p(1) = \mathbf{H}(\boldsymbol{\theta}_0) + o_p(1) = O_p(1),$$

where $-\mathbf{H}(\boldsymbol{\theta}_0) = \int \hat{\mathbf{H}}(\boldsymbol{\theta}_0)g(\mathbf{y})\mathrm{d}\mathbf{y}$.

When the model is correctly specified, according to the information matrix equality, we have $\mathbf{J}(\boldsymbol{\theta}_0) = -\mathbf{H}(\boldsymbol{\theta}_0)$. Hence, we get

$$-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}) = -\mathbf{H}(\boldsymbol{\theta}_0) + o_p(1) = \mathbf{J}(\boldsymbol{\theta}_0) + o_p(1) = \hat{\mathbf{J}}(\bar{\boldsymbol{\theta}}) + o_p(1).$$

Using the Bayesian large sample theory, we get $\sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \sim N\left[0, -L_n^{-(2)}(\hat{\boldsymbol{\theta}})\right]$ so that

$(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})=O_p(n^{-1/2})$. Hence, based on Remark 3.1, we get

$$\begin{aligned}
&f(\boldsymbol{\theta}|\mathbf{y}) = n(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'\hat{\mathbf{J}}(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}) = n(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})\\
=\;& n\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right]\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)\\
=\;& n(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) - 2n(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}})'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + n(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}})'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right](\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}})\\
=\;& n(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) - no_p(n^{-1/2})O_P(1)O_p(n^{-1/2}) + no_p(n^{-1/2})O_P(1)o_p(n^{-1/2})\\
=\;& n(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\left[-\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + o_p(1)\\
=\;& n(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + o_p(1)
\end{aligned}$$

Using the continuous mapping theorem, we can show that

$$n(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) = \sqrt{n}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\left[-L_n^{(2)}(\hat{\boldsymbol{\theta}})\right]\sqrt{n}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2(p).$$

Hence, we can show that

$$f(\boldsymbol{\theta}|\mathbf{y}) \xrightarrow{d} \chi^2(p).$$

7.4 Appendix 4: The derivation of BIMT for the linear state space model

The model latent variables x_t are linked to observed y_t via a state space system:

$$\begin{aligned}
x_t &= T(\theta)\,x_{t-1} + R(\theta)\,\varepsilon_t,\\
y_t &= D(\theta) + Z(\theta)\,x_t + \xi_t,
\end{aligned}$$

where y_t, D are $n_y \times 1$, T is $n_s \times n_s$, R is $n_s \times n_e$, Z is $n_y \times n_s$, θ is $n_q \times 1$.

Consider the state space system

$$\begin{aligned}
x_t &= Tx_{t-1} + R\varepsilon_t,\\
y_t &= D + Zx_t + \xi_t,
\end{aligned}$$

where $\varepsilon_t \sim N(0, Q)$, $\xi_t \sim N(0, H)$.

Let $Y_s = (y_1, y_2 ..., y_s)$, we can define

$$\begin{aligned}
x_t^s &= E(x_t|Y_s),\\
P_t^s &= E\{(x_t - x_t^s)(x_t - x_t^s)'|Y_s\}.
\end{aligned}$$

Then for the linear Gaussian state-space model specified in above equation, with initial condition x_0^0 and P_0^0, for $t = 1, 2...n$, the Kalman Filter algorithm is as follows:

$$\begin{aligned}
x_t^{t-1} &= Tx_{t-1}^{t-1},\\
P_t^{t-1} &= TP_{t-1}^{t-1}T' + RQR',
\end{aligned}$$

with

$$\begin{aligned} x_t^t &= x_t^{t-1} + K_t \left(y_t - D - Z x_t^{t-1} \right), \\ P_t^t &= \left[I_{n_s} - K_t Z \right] P_t^{t-1}, \end{aligned}$$

where

$$K_t = P_t^{t-1} Z' \left[Z P_t^{t-1} Z' + H \right]^{-1}.$$

From the Kalman Filter, the likelihood of the data is as follows:

$$\begin{aligned} \log \ell &= -\sum_{t=1}^{n} \left[\frac{n_y}{2} \log 2\pi + \frac{1}{2} \log |F_t| + \frac{1}{2} \left(y_t - D - Z x_t^{t-1} \right)' F_t^{-1} \left(y_t - D - Z x_t^{t-1} \right) \right] \\ &= -\sum_{t=1}^{n} \left[\frac{n_y}{2} \log 2\pi + \frac{1}{2} \log |F_t| + \frac{1}{2} \omega_t' F_t^{-1} \omega_t \right], \end{aligned}$$

where

$$\begin{aligned} F_t &= Z(\theta) P_t^{t-1} Z(\theta)' + H(\theta), \\ \omega_t &= y_t - D(\theta) - Z(\theta) x_t^{t-1}. \end{aligned}$$

Before we get the derivatives of the model, we first introduce some notations from Magnus and Neudecker (2002) about the matrix derivative.

Definition 7.1 *Let $F = (f_{st})$ be an $m \times p$ matrix function of an $n \times q$ matrix of variables $X = (x_{ij})$. Any $mp \times nq$ matrix A containing all the partial derivatives such that each row contains the partial derivatives of one function with respect to all variables, and each column contains the partial derivatives of all functions with respect to one variable x_{ij}, is called a derivative of F. We define the α-derivative as:*

$$DF(X) = \frac{\partial vecF(X)}{\partial (vecX)'}.$$

In our case, $\partial (vec\boldsymbol{\theta})' = \partial \boldsymbol{\theta}'$ since $\boldsymbol{\theta}$ is a vector.

Definition 7.2 *Let A be an $m \times n$ matrix. There exists a unique $mn \times mn$ permutation matrix K_{mn} which is defined as:*

$$K_{mn} \cdot vecA = vec\left(A' \right).$$

Since K_{mn} is a permutation matrix, it is orthogonal and $K_{mn}^{-1} = K_{mn}'$.

To compute the first order derivative of the likelihood, we have the following

$$\frac{\partial vec\left(\omega_t\right)}{\partial \boldsymbol{\theta}'} = -\frac{\partial vec\left(D\right)}{\partial \boldsymbol{\theta}'} - \left(x_t^{t-1\prime} \otimes I_{n_y}\right)\frac{\partial vec\left(C\right)}{\partial \boldsymbol{\theta}'} - \left(I_1 \otimes C\right)\frac{\partial vec\left(z_t^{t-1}\right)}{\partial \boldsymbol{\theta}'},$$

$$\begin{aligned}\frac{\partial vec\left(F_t\right)}{\partial \boldsymbol{\theta}'} &= \left(\left(P_t^{t-1}C'\right)' \otimes I_{n_y} + \left(I_{n_y} \otimes \left(CP_t^{t-1}\right)\right)K_{n_y n_s}\right)\frac{\partial vec\left(C\right)}{\partial \boldsymbol{\theta}'} \\ &\quad + \left(C \otimes C\right)\frac{\partial vec\left(P_t^{t-1}\right)}{\partial \boldsymbol{\theta}'} + \frac{\partial vecH}{\partial \boldsymbol{\theta}'},\end{aligned}$$

$$\frac{\partial vec\left(F_t^{-1}\right)}{\partial \boldsymbol{\theta}'} = -\left(\left(F_t^{-1}\right)' \otimes F_t^{-1}\right)\frac{\partial vec\left(F_t\right)}{\partial \boldsymbol{\theta}'},$$

$$\frac{\partial vec\left(\log\left|F_t\right|\right)}{\partial \boldsymbol{\theta}'} = \left(vec\left[\left(F_t^{-1}\right)'\right]\right)'\frac{\partial vec\left(F_t\right)}{\partial \boldsymbol{\theta}'},$$

$$\begin{aligned}\frac{\partial vec\left(\omega_t' F_t^{-1}\omega_t\right)}{\partial \boldsymbol{\theta}'} &= \left[\left(F_t^{-1}\omega_t\right)' \otimes I_1\right]K_{n_y 1}\frac{\partial vec\left(\omega_t\right)}{\partial \boldsymbol{\theta}'} + \left(\omega_t' \otimes \omega_t'\right)\frac{\partial vec\left(F_t^{-1}\right)}{\partial \boldsymbol{\theta}'} \\ &\quad + \left[I_1 \otimes \left(\omega_t' F_t^{-1}\right)\right]\frac{\partial vec\left(\omega_t\right)}{\partial \boldsymbol{\theta}'}.\end{aligned}$$

In the above equations, the first order derivatives of the matrix D, Z, Q, H, R are easy to get.

Given the initial conditions P_0^0 and x_0^0, we have the following recursive equations

$$\frac{\partial vec\left(z_t^{t-1}\right)}{\partial \boldsymbol{\theta}'} = \left(I_1 \otimes T\right)\frac{\partial vec\left(z_{t-1}^{t-1}\right)}{\partial \boldsymbol{\theta}'} + \left(z_{t-1}^{t-1\prime} \otimes I_{n_s}\right)\frac{\partial vec\left(T\right)}{\partial \boldsymbol{\theta}'},$$

$$\begin{aligned}\frac{\partial vec\left(P_t^{t-1}\right)}{\partial \boldsymbol{\theta}'} &= \left(\left(P_{t-1}^{t-1}T'\right)' \otimes I_{n_s}\right)\frac{\partial vec\left(T\right)}{\partial \boldsymbol{\theta}'} + \left(T \otimes T\right)\frac{\partial vec\left(P_{t-1}^{t-1}\right)}{\partial \boldsymbol{\theta}'} \\ &\quad + \left(I_{n_s} \otimes TP_{t-1}^{t-1}\right)K_{n_s n_s}\frac{\partial vec\left(T\right)}{\partial \boldsymbol{\theta}'} + \frac{\partial vec\left(RQR'\right)}{\partial \boldsymbol{\theta}'},\end{aligned}$$

$$\begin{aligned}\frac{\partial vec\left(z_t^t\right)}{\partial \boldsymbol{\theta}'} &= \frac{\partial vec\left(z_t^{t-1}\right)}{\partial \boldsymbol{\theta}'} + \left[\left(y_t - D - Cz_t^{t-1}\right)' \otimes I_{n_s}\right]\frac{\partial vec\left(K_t\right)}{\partial \boldsymbol{\theta}'} \\ &\quad - \left(I_1 \otimes K_t\right)\frac{\partial vec\left(D\right)}{\partial \boldsymbol{\theta}'} - \left(z_t^{t-1\prime} \otimes K_t\right)\frac{\partial vec\left(C\right)}{\partial \boldsymbol{\theta}'} - \left(I_1 \otimes K_t C\right)\frac{\partial vec\left(z_t^{t-1}\right)}{\partial \boldsymbol{\theta}'},\end{aligned}$$

$$\begin{aligned}\frac{\partial vec\left(P_t^t\right)}{\partial \boldsymbol{\theta}'} &= -\left(\left(CP_t^{t-1}\right)' \otimes I_{n_s}\right)\frac{\partial vec\left(K_t\right)}{\partial \boldsymbol{\theta}'} - \left(P_t^{t-1\prime} \otimes K_t\right)\frac{\partial vec\left(C\right)}{\partial \boldsymbol{\theta}'} \\ &\quad + \left(I_{n_s} \otimes \left(I_{n_s} - K_t C\right)\right)\frac{\partial vec\left(P_t^{t-1}\right)}{\partial \boldsymbol{\theta}'},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial vec\left(K_t\right)}{\partial \boldsymbol{\theta}'} &= \left[\left(C'F_t^{-1}\right)' \otimes I_{n_s}\right] \frac{\partial vec\left(P_t^{t-1}\right)}{\partial \boldsymbol{\theta}'} + \left[\left(F_t^{-1}\right)' \otimes P_t^{t-1}\right] K_{n_y n_s} \frac{\partial vec\left(C\right)}{\partial \boldsymbol{\theta}'} \\ &+ \left[I_{n_y} \otimes P_t^{t-1}C'\right] \frac{\partial vec\left(F_t^{-1}\right)}{\partial \boldsymbol{\theta}'},\end{aligned}$$

and

$$\frac{\partial vec\left(RQR'\right)}{\partial \boldsymbol{\theta}'} = \left[\left(RQ' \otimes I_{n_s}\right) + \left(I_{n_s} \otimes RQ\right) K_{n_s n_e}\right] \frac{\partial vecR}{\partial \boldsymbol{\theta}'} + \left(R \otimes R\right) \frac{\partial vecQ}{\partial \boldsymbol{\theta}'}.$$

The initial condition is given as

$$\begin{aligned}x_0^0 &= 0, \\ P_0^0 &= TP_0^0T' + RQR'.\end{aligned}$$

From the above, we have

$$vec\left(P_0^0\right) = \left(I_{n_s^2} - T \otimes T\right)^{-1} vec\left(RQR'\right).$$

Then

$$\frac{\partial vec\left(P_0^0\right)}{\partial \boldsymbol{\theta}'} = \left[\left(TP_0^0 \otimes I_{n_s}\right) + \left(I_{n_s} \otimes TP_0^0\right) K_{n_s n_s}\right] \frac{\partial vec\left(T\right)}{\partial \boldsymbol{\theta}'} + \left(T \otimes T\right) \frac{\partial vec\left(P_0^0\right)}{\partial \boldsymbol{\theta}'} + \frac{\partial vec\left(RQR'\right)}{\partial \boldsymbol{\theta}'}.$$

7.5 Appendix 5: The derivation of BIMT for the nonlinear non-Gaussian state space model with particle filters

Let $\varphi_t\left(\mathbf{z}^t\right)$ be the first order derive of the complete likelihood function with respect to the parameter $\boldsymbol{\theta}$. This is just the integrand in Fisher's identity (Cappé et al., 2005)

$$\frac{\partial \log p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} = \int \frac{\partial \log p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t.$$

Then we have the following recursive form

$$\varphi_t\left(\mathbf{z}^t\right) = \varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t\left(z_t, z_{t-1}\right),$$

where

$$\varphi_t\left(\mathbf{z}^t\right) = \frac{\partial \log p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}, \quad u_t\left(z_t, z_{t-1}\right) = \frac{\partial \log g\left(y_t|z_t, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} + \frac{\partial \log f_{\boldsymbol{\theta}}\left(z_t|z_{t-1}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}.$$

Hence, following Doucet and Shephard (2012), we get the sample score $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$ as

$$\begin{aligned}\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta}) &= \int \varphi_t\left(\mathbf{z}^t\right) p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t \\ &= \int\int \left(\varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t\left(z_t, z_{t-1}\right)\right) p\left(\mathbf{z}^{t-1}|z_t, \mathbf{y}^{t-1}, \boldsymbol{\theta}\right) d\mathbf{z}^{t-1} p\left(z_t|\mathbf{y}^t, \boldsymbol{\theta}\right) dz_t \\ &= \int S_t\left(z_t\right) p\left(z_t|\mathbf{y}^t, \boldsymbol{\theta}\right) dz_t,\end{aligned}$$

where

$$
\begin{aligned}
S_t(z_t) &= \int \left(\varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t(z_t, z_{t-1})\right) p\left(\mathbf{z}^{t-1}|z_t, \mathbf{y}^{t-1}, \boldsymbol{\theta}\right) d\mathbf{z}^{t-1} \\
&= \int \left(\varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t(z_t, z_{t-1})\right) p\left(\mathbf{z}^{t-2}|z_{t-1}, \mathbf{y}^{t-2}, \boldsymbol{\theta}\right) d\mathbf{z}^{t-2} p\left(z_{t-1}|z_t, \mathbf{y}^{t-2}, \boldsymbol{\theta}\right) dz_{t-1} \\
&= \frac{\int \left(S_{t-1}(z_{t-1}) + u_t(z_t, z_{t-1})\right) f(z_t|z_{t-1}, \boldsymbol{\theta}) p\left(z_{t-1}|\mathbf{y}^t, \boldsymbol{\theta}\right) dz_{t-1}}{\int f(z_t|z_{t-1}, \boldsymbol{\theta}) p(z_{t-1}|\mathbf{y}^t, \boldsymbol{\theta}) dz_{t-1}}.
\end{aligned}
$$

Then we have

$$
\widehat{S}_t(z_t) = \frac{\sum_{j=1}^N W_{t-1}^{(j)} f\left(z_t|z_{t-1}^{(i)}, \boldsymbol{\theta}\right)}{\sum_{j=1}^N f\left(z_t|z_{t-1}^{(i)}, \boldsymbol{\theta}\right)} \left(S_{t-1}\left(z_{t-1}^{(i)}\right) + \frac{\partial \log g(y_t|z_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \log f\left(z_t|z_{t-1}^{(i)}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right)
$$

Let $\varphi_t\left(\mathbf{z}^t\right)$ be the first order derive of the complete likelihood function with respect to the parameter $\boldsymbol{\theta}$. This is just the integrand in Fisher's identity (Cappé et al., 2005)

$$
\frac{\partial \log p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} = \int \frac{\partial \log p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t.
$$

Then we have the following recursive form

$$
\varphi_t\left(\mathbf{z}^t\right) = \varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t(z_t, z_{t-1}),
$$

where

$$
\varphi_t\left(\mathbf{z}^t\right) = \frac{\partial \log p\left(\mathbf{z}^t, \mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}, \quad u_t(z_t, z_{t-1}) = \frac{\partial \log g(y_t|z_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \log f_{\boldsymbol{\theta}}(z_t|z_{t-1}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.
$$

Hence, following Doucet and Shephard (2012), we get the sample score $\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta})$ as

$$
\begin{aligned}
\mathbf{s}(\mathbf{y}^t, \boldsymbol{\theta}) &= \int \varphi_t\left(\mathbf{z}^t\right) p\left(\mathbf{z}^t|\mathbf{y}^t, \boldsymbol{\theta}\right) d\mathbf{z}^t \\
&= \int\int \left(\varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t(z_t, z_{t-1})\right) p\left(\mathbf{z}^{t-1}|z_t, \mathbf{y}^{t-1}, \boldsymbol{\theta}\right) d\mathbf{z}^{t-1} p\left(z_t|\mathbf{y}^t, \boldsymbol{\theta}\right) dz_t \\
&= \int S_t(z_t) p\left(z_t|\mathbf{y}^t, \boldsymbol{\theta}\right) dz_t,
\end{aligned}
$$

where

$$
\begin{aligned}
S_t(z_t) &= \int \left(\varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t(z_t, z_{t-1})\right) p\left(\mathbf{z}^{t-1}|z_t, \mathbf{y}^{t-1}, \boldsymbol{\theta}\right) d\mathbf{z}^{t-1} \\
&= \int \left(\varphi_{t-1}\left(\mathbf{z}^{t-1}\right) + u_t(z_t, z_{t-1})\right) p\left(\mathbf{z}^{t-2}|z_{t-1}, \mathbf{y}^{t-2}, \boldsymbol{\theta}\right) d\mathbf{z}^{t-2} p\left(z_{t-1}|z_t, \mathbf{y}^{t-2}, \boldsymbol{\theta}\right) dz_{t-1} \\
&= \frac{\int \left(S_{t-1}(z_{t-1}) + u_t(z_t, z_{t-1})\right) f(z_t|z_{t-1}, \boldsymbol{\theta}) p\left(z_{t-1}|\mathbf{y}^t, \boldsymbol{\theta}\right) dz_{t-1}}{\int f(z_t|z_{t-1}, \boldsymbol{\theta}) p(z_{t-1}|\mathbf{y}^t, \boldsymbol{\theta}) dz_{t-1}}.
\end{aligned}
$$

Then we have

$$\widehat{S}_t(z_t) = \frac{\sum_{j=1}^{N} W_{t-1}^{(j)} f\left(z_t|z_{t-1}^{(i)}, \boldsymbol{\theta}\right)}{\sum_{j=1}^{N} f\left(z_t|z_{t-1}^{(i)}, \boldsymbol{\theta}\right)} \left(S_{t-1}\left(z_{t-1}^{(i)}\right) + \frac{\partial \log g(y_t|z_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \log f\left(z_t|z_{t-1}^{(i)}, \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right)$$

and

$$\widehat{\mathbf{s}}(\mathbf{y}^t, \boldsymbol{\theta}) = \sum_{j=1}^{N} W_t^{(j)} \widehat{S}_t\left(z_t^{(j)}\right),$$

where $\left(W_t^{(j)}, z_t^{(i)}\right)$ are the particles to approximate $p\left(z_t|\mathbf{y}^t\right) dz_t$. Then the individual scores is estimated by

$$\widehat{\mathbf{s}}_t(\boldsymbol{\theta}) = \widehat{\mathbf{s}}(\mathbf{y}^t, \boldsymbol{\theta}) - \widehat{\mathbf{s}}(\mathbf{y}^{t-1}, \boldsymbol{\theta}).$$

For the asymptotic properties of $\widehat{\mathbf{s}}_t(\boldsymbol{\theta})$, see Poyiadjis (2011) and Doucet and Shephard (2012).

7.6 Appendix 6: The derivation of BIMT for the DSGE model

The equilibrium object for a DSGE model is a collection of the nonlinear equations defining optimality conditions, markets clearing conditions, etc. We follow the standard practice and linearize these conditions around a steady state. Then, the model can be written as linear expectation system,

$$\Gamma_0(\theta) x_t = \Gamma_1(\theta) E_t[x_{t+1}] + \Gamma_2(\theta) x_{t-1} + \Gamma_3(\theta) \varepsilon_t, \tag{6}$$

where x_t are the state variables, ε_t the exogenous shocks, θ the structural parameters of interest, and $\{\Gamma_1\}$ matrix functions that map the equilibrium conditions of the model, where $\Gamma_0(\theta), \Gamma_1(\theta), \Gamma_2(\theta)$ are $n_s \times n_s$, $\Gamma_3(\theta)$ is $n_s \times n_e$. The solution to the system takes the form of a $VAR(1)$,

$$x_t = T(\theta) x_{t-1} + R(\theta) \varepsilon_t, \tag{7}$$

The mapping from θ to T and R must be solved numerically for all models of interest, where T is $n_s \times n_s$, R is $n_s \times n_e$. The model variables x_t are linked to observed y_t via a state space system:

$$\begin{aligned} x_t &= T(\theta) x_{t-1} + R(\theta) \varepsilon_t, \\ y_t &= D(\theta) + Z(\theta) x_t + \xi_t, \end{aligned}$$

where y_t, D are $n_y \times 1$, Z is $n_y \times n_s$, θ is $n_q \times 1$. Then the likelihood function is the same as in Appendix 4. It is different from the dynamic factor model that $T(\theta)$ and $R(\theta)$ do not have closed form in DSGE models.

Following Iskrev (2008), we can get the first order derivatives of matrix T and R, substitute (7) into (6), we have

$$\Gamma_0(\theta)\, x_t = \Gamma_1(\theta)\, T(\theta)\, x_t + \Gamma_2(\theta)\, x_{t-1} + \Gamma_3(\theta)\, \varepsilon_t.$$

Furthermore

$$\left(\Gamma_0(\theta) - \Gamma_1(\theta)\, T(\theta)\right) x_t = \Gamma_2(\theta)\, x_{t-1} + \Gamma_3(\theta)\, \varepsilon_t. \tag{8}$$

From (8)

$$\left(\Gamma_0(\theta) - \Gamma_1(\theta)\, T(\theta)\right) x_t = \left(\Gamma_0(\theta) - \Gamma_1(\theta)\, T(\theta)\right) T(\theta)\, x_{t-1} + \left(\Gamma_0(\theta) - \Gamma_1(\theta)\, T(\theta)\right) R(\theta)\, \varepsilon_t. \tag{9}$$

Comparing (8) and (9), we have

$$\left(\Gamma_0(\theta) - \Gamma_1(\theta)\, T(\theta)\right) T(\theta) - \Gamma_2(\theta) = 0. \tag{10}$$

$$\left(\Gamma_0(\theta) - \Gamma_1(\theta)\, T(\theta)\right) R(\theta) - \Gamma_3(\theta) = 0. \tag{11}$$

Consider Equation(10), we can get the derivatives of matrix T by solving the following equation

$$\begin{aligned}\left[(I_{n_s} \otimes \Gamma_0) - (I_{n_s} \otimes \Gamma_1 T) - \left(T' \otimes \Gamma_1\right)\right] \frac{\partial vec(T)}{\partial \theta'} &- \left(T'^2 \otimes I_{n_s}\right) \frac{\partial vec(\Gamma_1)}{\partial \theta'} \\ &+ \left(T' \otimes I_{n_s}\right) \frac{\partial vec(\Gamma_0)}{\partial \theta'} - \frac{\partial vec(\Gamma_2)}{\partial \theta'} = 0.\end{aligned}$$

From (11), the first order derivatives of matrix R is as follows:

$$\frac{\partial vec(R)}{\partial \theta'} = -\left(\Gamma_3' \otimes I_{n_s}\right)\left(W'^{-1} \otimes W^{-1}\right) \frac{\partial vec(W)}{\partial \theta'} + \left(I_{n_e} \otimes W^{-1}\right) \frac{\partial vec(\Gamma_3)}{\partial \theta'}.$$

From Herbst (2011), where

$$W = \Gamma_0 - \Gamma_1 T,$$

$$\frac{\partial vec(W)}{\partial \theta'} = \frac{\partial vec(\Gamma_0)}{\partial \theta'} - \left(T' \otimes I_{n_s}\right) \frac{\partial vec(\Gamma_1)}{\partial \theta'} - (I_{n_s} \otimes \Gamma_1) \frac{\partial vec(T)}{\partial \theta'}.$$

References

Abanto-Valle, C.A., Bandyopadhyay, D., Lachos, V.H. and Enriquez, I. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis*, **54(12)**, 2883-2898.

Adjemian, S., Bastani,H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot, S. (2011). Dynare: Reference manual, version 4, *Dynare Working Papers*, 1.

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9(2)**, 385-426.

An, S. and Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric Reviews*, **26**, 113-172.

Andrews, D.W. (1997). A conditional Kolmogorov test. *Econometrica*, **65(5)**, 1097-1128.

Andrews, I. and Mikusheva, A. (2013), Maximum likelihood inference in weakly identified DSGE models. *Quantitative Economics, forthcoming.*

Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in hidden Markov models.* New York: Springer.

Canova, F. and Sala,L. (2009). Back to square one: identification issues in DSGE models. *Journal of Monetary Economics,* **56,** 431-449.

Chen, C.F. (1985). On Asymptotic Normality of Limiting Density Function with Bayesian Implications. *Journal of the Royal Statistical Society, Series B*, **47(3)**, 540-546.

Chesher, A. and Spady, R. (1991). Asymptotic expansions of the information matrix test statistic. *Econometrica*, **59**, 787–815.

Clarida, R., Gali, J. and Gertler, M. (1999). The science of monetary policy: a new Keynesian perspective. *Journal of Economic Literature,* **37,** 1661-1707.

Corradi, V. and Swanson, N.R. (2004). Some recent developments in predictive accuracy testing with nested models and (generic) nonlinear alternatives. *International Journal of Forecasting*, **20(2)**, 185-199.

Cox, D.R. (1961). Tests of separate families of hypotheses. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability.* Vol. 1, 105-123.

Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B*, **24(2)**, 406-424.

Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews*, **31(3)**, 245-296..

Davidson, R. and MacKinnon, J.G. (1992). A new form of information test. *Econometrica*, **60(1)**, 145-157.

Dempster, A.P., Laird, N.M. and Rublin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39(1)**, 1-38.

Dhaene, G. and Hoorelbeke, D. (2004). The information matrix test with bootstrap-based covaiance matrix estimation. *Economics Letters*, **82**, 341-347.

Doucet, A. and Johansen, A.M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, **12**, 656-704.

Doucet, A. and Shephard, N. (2012). Robust inference on parameters via particle filters and sandwich covariance matrices. *Working paper*, University of Oxford, Department of Economics.

Durbin, J., Koopman, S.J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Eubank, R.L. and Spiegelman, C.H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, **85**, 387-392.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, 865-890.

Fernádez-Villaverde, J. and J. Rubio-Ramíez (2005). Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics,* **20**, 891-910.

Geweke, J.F., Koop, G. and van Dijk, H.K. (2011). *Handbook of Bayesian Econometrics*. Oxford: Oxford University Press.

Geweke, J.F. and McCauland, W. (2001). Bayesian specification analysis in econometrics. *American Journal of Agricultural Economics,* **83(5),** 1181-1186.

Gozalo, P.L. (1993). A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, **9**, 451-451.

Guerron-Quintana, P., A. Inoue, and L. Lilian (2013). Frequentist inference in weakly identified DSGE models. *Quantitative Economics,* **4(2)**, 197-229.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50(4)**, 1029-1054.

Harvey, A. C. (1991). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, **46**, 1251-1271.

Herbst, E. (2011). Essays on Bayesian Marcroeconomics. Ph.D. dissertation, University of Pennsylvania.

Hong, Y., & Li, H. (2005). Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*, **18(1)**, 37-84.

Horowitz, J. L. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, **61**, 395–411.

Imai, S., Jain, N., and Ching, A. (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica*, **77(6)**, 1865-1899.

Iskrev, N. (2008). Evaluating the information matrix in linearized DSGE Models. *Economics Letters*, **99**, 607-610.

Iskrev, N. (2010). Evaluating the strength of identification in DSGE models. An a priori approach. *Bank of Portugal working paper.*

Jacquier, E., Polson N., Rossi P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics,* **122,** 185-212.

Johnson, V.E. (2004). A Bayesian χ^2 test for goodness-of-fit. *The Annal of Statistics,* **32(6),** 2361-2384.

Kim, C J, Nelson, C. R. (1999). *State-space models with regime switching.* Cambridge, The MIT Press.

Kim, J. Y. (1994). Bayesian asymptotic theory in a time series model with a possible nonstationary process. *Econometric Theory*, **10(3-4)**, 764-773.

Kim, J. Y. (1998). Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models. *Econometrica*, **66**, 359-380.

Kim S., Shepard N., Chib S. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies,* **65**, 361-393.

Koopman, S. J. and Shephard, N. (1992). Exact score for time series models in state space form. *Biometrika*, 823-826.

Lancaster, T. (1984). The covariance matrix of the information matrix test. *Econometrica*, **52(4)**, 1051-1053.

Li, Y. and Yu, J. (2012). Bayesian Hypothesis Testing in Latent Variable Models. *Journal of Econometrics*, **166(2)**, 237-246.

Li, Y., Zeng, T. and Yu, J. (2014). Robust Deviation Information Criterion for Latent Variable Models. *Working paper*, Singapore Management University.

Li, Y., Zeng, T. and Yu, J. (2014). A new approach to Bayesian hypothesis testing. *Journal of Econometrics*, **178**, 602-612.

Lubik, T. and Schorfheide, F. (2006). A Bayesian look at the new open economy macroeconomics, in *NBER Macroeconomics Annual 2005, Volume 20,* MIT Press, 313-382.

Magnus, J. R. and Neudecker, H. (2002). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

Mavroeidis, S. (2005), Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve. *Journal of Money, Credit and Banking*, **37**, 421-499.

Meyer, R., and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *The Econometrics Journal*, **3(2)**, 198-215.

Newey, W.K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, **53(5)**, 1047-1070.

Müller,U.K. (2013). Risk of Bayesian inference in misspecifed models and the sandwich covariance matrix. *Econometrica*, **81(5)**, 1805-1849.

Orme, C. (1990). The small sample performance of the information-matrix test. *Journal of Econometrics*, **46**, 309-331.

Pitt, M.K., and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, **94**, 590-599.

Poyiadjis, G., Doucet, A., Singh, S.S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, **98(1)**, 65-80.

Presnell, B. and Boos, D.D. (2004). The IOS Test for model misspecification. *Journal of the American Statistical Association*, **99**, 216-227.

Raftery, A. and Lewis,S. (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, **4**, 763-773.

Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, **26**, 393-415.

Schorfheide, F. (2001). Loss function-based evaluation of DSGE models, *Journal of Applied Econometrics,* **15**, 645-70.

Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review,* **97**(3), 586-606.

Stock, J.H. and Watson, M. W. (1992). A probability model of the coincident economic indicators. *Leading Economic Indicators: New Approaches and Forecasting Records*, 63.

Taylor, S.J. (1982). Financial returns modelled by the product of two stochastic processes - A study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, **1**, 203-226.

Takeuchi, K. (1976). Distribution of Informational Statistics and a Criterion of Model Fitting. *Suri-Kagaku (Mathematic Sciences)*, **153**, 12-18.(in Japanese).

Tanner, T.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528-540.

Tauchen G.E. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, **30(1)**, 415-443.

Tauchen G.E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, **51(2)**, 485-505.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50(1)**, 1-25.

White, H. (1987). Specification testing in dynamic models. *Advances in Econometrics*, Fifth world congress. **1**: 1-58.

White, H. (1996). *Estimation, inference and specification analysis.* Cambridge university press.

Wooldridge, J.M. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory*, **8(4)**, 452-475.

Zheng, X. (2000). A consistent test of conditional parametric distributions. *Econometric Theory*, **16(5)**, 667-691.