A Specification Test based on the MCMC Output*

Yong Li
Renmin University of China

Jun Yu
Singapore Management University

Tao Zeng
Wuhan University

May 10, 2017

Abstract

A test statistic is proposed to assess the model specification after the model is estimated by Bayesian MCMC methods. The new test is motivated from the power enhancement technique of Fan, Liao and Yao (2015). It combines a component (J_1) that tests a null point hypothesis in an expanded model and a power enhancement component (J_0) obtained from the null model. It is shown that J_0 converges to zero when the null model is correctly specified and diverges when the null model is misspecified. Also shown is that J_1 is asymptotically χ^2-distributed, suggesting that the proposed test is asymptotically pivotal, when the null model is correctly specified. The proposed test has several properties. First, its size distortion is small and hence bootstrap methods can be avoided. Second, it is easy to compute from the MCMC output and hence is applicable to a wide range of models, including latent variable models for which frequentist methods are difficult to use. Third, when the test statistic rejects the specification of the null model and J_1 takes a large value, the test suggests the source of misspecification of the null model. The finite sample performance is investigated using simulated data. The method is illustrated in a linear regression model, a linear state-space model, and a stochastic volatility model using real data.

JEL classification: C11, C12, G12
Keywords: Specification test; Point hypothesis test; Latent variable models; Markov chain Monte Carlo; Power enhancement technique; Information matrix.

*We would like to thank three referees, an AE, Yacine Aït-Sahalia, Jianqing Fan for their constructive comments which have led to a significant improvement of the paper. 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 3 under the grant number MOE2013-T3-1-009. Yong Li, Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing, 1000872, 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. Tao Zeng, Economics and Management School, Wuhan University, Wuhan, 430072, China.

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 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 conditional moment test of Newey (1985), the information matrix test of White (1982), the IOS test of Presnell and Boos (2004), the information ratio (IR) test of Zhou et al (2012). 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 parametric counterparts. The idea originated from the Kolmogorov-Smirnov test or the closely related family such as the Cramer-von Mises and Anderson-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 density.

For many widely used models in economics, such as latent variable models and structural dynamic choice models (Imai, Jain and Ching, 2009; Norets, 2009), it is not easy to obtain the ML estimate or construct a nonparametric estimate. 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 in the Bayesian paradigm.

In addition, it is well-known that specification tests that are based on the information matrix, including the information matrix test (IMT) of White (1982), the IOS test of Presnell and Boos (2004), the IR test of Zhou et al (2012), are subject to severe size distortions. To reduce the size distortion, bootstrap methods have been used; see for example, Horowitz (1994), Presnell and Boos (2004), Zhou et al (2012). For models where MCMC is a popular estimation method, it is computationally infeasible to do bootstrap.

Given the increasing popularity of MCMC in practical applications, it is therefore natural to introduce a specification test to assess the adequacy of a candidate model after it is estimated by MCMC. We seek to answer two questions in the present paper. First, how we can assess the validity of the model specification? Second, is it possible to tell the source of model misspecification if the null model is misspecified? Motivated by the power enhancement technique of Fan, et al (2015) and based on a model expansion strategy, we propose a new specification test based on the MCMC output. It combines a component (J_1) that tests a null point hypothesis in an expanded model and a power enhancement component (J_0) obtained from the null model. It is shown that J_0 converges to zero when the null model is correctly specified and diverges when the null model is misspecified. Also shown is that J_1 is asymptotically χ^2-distributed, suggesting that the proposed test is asymptotically pivotal, when the null model is correctly specified.

The proposed test has several nice properties. First, its size distortion is small and hence bootstrap methods can be avoided. Second, it is easy to compute from the MCMC output and hence is applicable to a wide range of models, including latent variable models for which ML and bootstrap methods are difficult to use. Third, when the test statistic rejects the specification of the null model and J_1 takes a large value, our test suggests the source of misspecification of the null model. However, the proposed test as a lower local power. This is the price we pay for avoiding using bootstrap methods.

The paper is organized as follows. Section 2 briefly reviews the literature on the specification tests. Section 3 proposes the test statistic based on the MCMC output and establishes the properties of the proposed test. Section 4 illustrates the new method using two simulation studies and three empirical studies. Section 5 concludes the paper. Appendix collects the proof of the theoretical results in the paper and discusses how to compute the test statistic in the context of state-space models.

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 q. Let $P_{\boldsymbol{\theta}}$ denote P indexed by $\boldsymbol{\theta}$. 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 information matrix equivalence due to White (1982). Let $p(\mathbf{y}|\boldsymbol{\theta})$ denote the likelihood function of Model $P_{\boldsymbol{\theta}}$ and

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

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 r $(= q(q+1)/2)$ dimensional vector. Let $\mathbf{y} = (y_1, \ldots, y_n)$ denote the i.i.d. observations and

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

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

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

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

$$\text{IMT} = nD_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)V_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)D_n\left(\hat{\boldsymbol{\theta}}_{ML}\right), \tag{1}$$

where

$$\begin{aligned}V_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) &= \frac{1}{n}\sum_{t=1}^{n}\nu_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)\nu_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)',\\ \nu_t\left(\hat{\boldsymbol{\theta}}_{ML}\right) &= d\left(y_t, \hat{\boldsymbol{\theta}}_{ML}\right) - \dot{D}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)\hat{\mathbf{H}}_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\mathbf{s}\left(y_t, \hat{\boldsymbol{\theta}}_{ML}\right).\end{aligned}$$

Under a set of regularity conditions, White (1982) showed that $\text{IMT} \xrightarrow{d} \chi^2(r)$ as $n \to \infty$ under the null hypothesis. White (1987) extended the method to cover dynamic models. Lancaster (1984) pointed out that the covariance matrix of IMT can be estimated without computing the third derivatives of the density function analytically.

Presnell and Boos (2004) proposed an alternative test – the "in-and-out" likelihood ratio (IOS) test for models with i.i.d. observations. Let $\hat{\boldsymbol{\theta}}_{ML}^{(t)}$ 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 $p\left(y_t, \hat{\boldsymbol{\theta}}_{ML}^{(t)}\right)$ can be regarded as the predictive likelihood by the other observations. Presnell and Boos (2004) defined the "in-and-out" likelihood ratio test as:

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

and showed that the asymptotic form of IOS is

$$\mathrm{IOS}_A = \mathbf{tr}\left[-\hat{\mathbf{H}}_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)\right], \tag{2}$$

and $\mathrm{IOS} - \mathrm{IOS}_A = o_p\left(n^{-1/2}\right)$. Like IMT, IOS_A also compares $\hat{H}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)$ with $\hat{J}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)$, but in a ratio form instead of an additive form. Under the null hypothesis, $\mathrm{IOS}_A \xrightarrow{p} q$ and $n^{1/2}\left(\mathrm{IOS}_A - q\right)$ converges to a normal distribution with zero mean and finite variance. Clearly, IOS and IOS_A are asymptotically equivalent.

Zhou, et al (2012) considered the model misspecification problem that the first moment of a candidate model is correctly specified, but the second moment is misspecified. The proposed test statistic takes the form of IOS_A/q which is denoted as the information ratio (IR) test. Zhou, et al (2012) established the asymptotic distribution of IR. Under the null hypothesis, it was shown that $n^{1/2}\left(\mathrm{IR} - 1\right)$ converges to a normal distribution with zero mean and finite variance.

It is well documented that the asymptotic distributions poorly approximate their finite sample counterparts for IMT, IOS, IOS_A, and IR. As a result, they all suffer serious bias distortions if the asymptotic distributions are used to obtain critical values. See Orme (1990), Chesher and Spady (1991), Davidson and Mackinnon (1992), Horowitz (1994) for evidence of size distortions for IMT. The poor finite sample performance of these tests is not surprising as the asymptotic theory relies on the convergence of the sample high order moments which is slow. To reduce the size distortion of IMT, Horowitz (1994) advocated the use of bootstrap methods to obtain better critical values for implementing IMT. Presnell and Boos (2004) suggested using a bootstrap method for implementing IOS and IOS_A. Zhou et al (2012) suggested using a different bootstrap method for the IR test.

It is not necessary to base a specification test on ML. 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 Anderson-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 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 Bayesian specification analysis.

3 A Specification Test based on the MCMC Output

After a candidate model is estimated by a Bayesian MCMC method, a natural way to check the validity of the model is to construct a MCMC-based version of a ML-based specification test. This is a reasonable way to proceed as both ML and MCMC are full-likelihood-based approaches.

3.1 A naïve MCMC-based information matrix test

In this subsection, we propose a naïve MCMC-based information matrix test. First we need to introduce some notations. Define $l_t(\boldsymbol{\theta}) = \log p\left(\mathbf{y}^t|\boldsymbol{\theta}\right) - \log p\left(\mathbf{y}^{t-1}|\boldsymbol{\theta}\right)$ to be the conditional likelihood for t observation and $\nabla^j l_t(\boldsymbol{\theta})$ as the jth derivative of $l_t(\boldsymbol{\theta})$, we suppress the subscript when $j=1$. Let $\mathbf{y}^t := (y_1, \ldots, y_t)$, and

$$
\begin{aligned}
&\mathbf{s}\left(\mathbf{y}^t, \boldsymbol{\theta}\right) := \frac{\partial \log p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{t} \nabla l_i(\boldsymbol{\theta}),\ \mathbf{h}\left(\mathbf{y}^t, \boldsymbol{\theta}\right) := \frac{\partial^2 \log p\left(\mathbf{y}^t|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \sum_{i=1}^{t} \nabla^2 l_i(\boldsymbol{\theta}),\\
&\mathbf{s}_t(\boldsymbol{\theta}) := \nabla l_t(\boldsymbol{\theta}) = \mathbf{s}\left(\mathbf{y}^t, \boldsymbol{\theta}\right) - \mathbf{s}\left(\mathbf{y}^{t-1}, \boldsymbol{\theta}\right),\ \mathbf{h}_t(\boldsymbol{\theta}) := \nabla^2 l_t(\boldsymbol{\theta}) = \mathbf{h}\left(\mathbf{y}^t, \boldsymbol{\theta}\right) - \mathbf{h}\left(\mathbf{y}^{t-1}, \boldsymbol{\theta}\right),\\
&\hat{\mathbf{J}}_n(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{s}_t(\boldsymbol{\theta})\mathbf{s}_t'(\boldsymbol{\theta}),\ \hat{\mathbf{H}}_n(\boldsymbol{\theta}) := \frac{1}{n}\sum_{t=1}^{n} \mathbf{h}_t(\boldsymbol{\theta}),\\
&\mathbf{J}_n(\boldsymbol{\theta}) := \int \hat{\mathbf{J}}_n(\boldsymbol{\theta})\, g(\mathbf{y})d\mathbf{y}, \mathbf{H}_n(\boldsymbol{\theta}) := \int \hat{\mathbf{H}}_n(\boldsymbol{\theta})\, g(\mathbf{y})d\mathbf{y}\\
&L_n(\boldsymbol{\theta}) := \log p(\boldsymbol{\theta}|\mathbf{y}),\ L_n^{(j)}(\boldsymbol{\theta}) := \partial^j \log p(\boldsymbol{\theta}|\mathbf{y}) / \partial \boldsymbol{\theta}^j.
\end{aligned}
$$

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

Assumption 1: Let $\hat{\boldsymbol{\theta}}$ be the posterior mode such that $L_n^{(1)}(\hat{\boldsymbol{\theta}}) = 0$. There exists an integer N_1 and some $\delta > 0$ such that for $n > N_1$ and $\boldsymbol{\theta} \in H(\hat{\boldsymbol{\theta}}, \delta) = \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|| \leq \delta\}$, $L_n^{(2)}(\hat{\boldsymbol{\theta}})$ is negative definite with probability approaching one.

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

Assumption 3: For any $\varepsilon > 0$, there exists a positive number δ, such that

$$\lim_{n\to\infty} P\left[\sup_{\boldsymbol{\theta}\in B\left(\hat{\boldsymbol{\theta}},\,\delta\right)} \left\| \left[-L_n^{(2)}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1} \left[L_n^{(2)}(\boldsymbol{\theta}) - L_n^{(2)}\left(\hat{\boldsymbol{\theta}}\right)\right] \right\| < \varepsilon\right] = 1. \tag{3}$$

where $B\left(\hat{\boldsymbol{\theta}},\ \delta\right)$ is the neighborhood of $\hat{\boldsymbol{\theta}}$.

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

$$\int_{\boldsymbol{\Theta}-B\left(\hat{\boldsymbol{\theta}},\,\delta\right)} p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta} = O_p\left(n^{-3}\right),$$

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

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

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

where $\boldsymbol{\theta}_0$ is a unique minimizer.

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

Assumption 7: Assume

$$\mathbf{H}(\boldsymbol{\theta}_0) := \lim_{n\longrightarrow\infty} \mathbf{H}_n(\boldsymbol{\theta}_0) \text{ and } \mathbf{J}(\boldsymbol{\theta}_0) := \lim_{n\longrightarrow\infty} \mathbf{J}_n(\boldsymbol{\theta}_0)$$

exist and are nonsingular, and $\lim_{n\longrightarrow\infty} n^{-1} \int \sum_{t=1}^n \triangledown^3 l_t(\boldsymbol{\theta}_0)\, g(\mathbf{y}) d\mathbf{y}$ exists.

Assumption 8: $\boldsymbol{\theta}_0 \in int(\boldsymbol{\Theta})$ where $\boldsymbol{\Theta}$ is a compact, separable metric space.

Assumption 9: $\{y_t, t = 1, 2, 3, \ldots\}$ is an α mixing sequence that satisfies, for $\mathcal{F}_{-\infty}^t = \sigma(y_t, y_{t-1}, \ldots)$ and $\mathcal{F}_{t+m}^\infty = \sigma(y_{t+m}, y_{t+m+1}, \ldots)$, the mixing coefficient $\alpha(m) = O\left(m^{\frac{-2r}{r-2}-\varepsilon}\right)$ for some $\varepsilon > 0$ and $r > 2$.

Assumption 10: There exists a function $M_t(y_t)$ such that for $0 \leqslant j \leqslant 8$, all $\boldsymbol{\theta} \in \mathcal{G}$ where $\mathcal{G}$ is an open, convex set containing $\boldsymbol{\Theta}$, $\triangledown^j l_t(\boldsymbol{\theta})$ exists, $\sup_{\boldsymbol{\theta}\in\mathcal{G}} \left\|\triangledown^j l_t(\boldsymbol{\theta})\right\| \leqslant M_t(y_t)$, and $\sup_t E\left\|M_t(y_t)\right\|^{r+\delta} \leq M < \infty$ for some $\delta > 0$.

Assumption 11: $\left\{\triangledown^j l_t(\boldsymbol{\theta})\right\}$ is L_2-near epoch dependent with respect to $\{\mathbf{y}_t\}$ of size -1 for $0 \leqslant j \leqslant 1$ and $-\frac{1}{2}$ for $j = 2, 3$ uniformly on $\boldsymbol{\Theta}$.

Assumption 12: For all $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, $\left\| \triangledown^j l_t(\boldsymbol{\theta}) - \triangledown^j l_t(\boldsymbol{\theta}') \right\| \leq c_t(\mathbf{y}^t) \left\| \boldsymbol{\theta} - \boldsymbol{\theta}' \right\|$ for $0 \leqslant j \leqslant 3$ in probability, where $c_t(\mathbf{y}^t)$ is a positive random variable, $\sup_t E \left\| c_t(\mathbf{y}^t) \right\| < \infty$ and $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} (c_t - Ec_t) \xrightarrow{p} 0$.

Remark 3.1 *Assumption 1-4 have been used to develop the Bayesian large sample theory; see, for example, Chen (1985), Kim (1994, 1998), Geweke (2005). Similar assumptions have been used to develop the asymptotic properties of the Laplace type estimator in Chernozhukov and Hong (2003). The order condition in Assumption 4 is used to develop higher order expansions; see, for example, Miyata (2004, 2010). Based on these assumptions, Li, Zeng and Yu (2015) showed that,*

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

Assumption 5 is a standard regularity condition on the Hessian; see Müller (2013). Assumption 6 ensures that when the sample size increases, the likelihood information dominates the prior information so that the prior information can be ignored asymptotically. Assumption 7-10 are similar to those made in Rilstone et al (1996), Newey and Smith (2004), and Bester and Hansen (2006) for developing higher order expansions.

Based on Remark 3.1 and the expression for IOS_A given in Equation (2), if we replace $-\hat{H}_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)$ with $V\left(\hat{\boldsymbol{\theta}}\right)$, a natural MCMC-based informative matrix test statistic can be defined as:

$$\mathrm{BIMT} = \mathbf{tr}\left[nV\left(\bar{\boldsymbol{\theta}}\right)\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right)\right] = n\int \left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)'\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right) p(\boldsymbol{\theta}|\mathbf{y})\,\mathrm{d}\boldsymbol{\theta}. \tag{4}$$

Proposition 3.1 *Under Assumptions 1-12, we have*

$$BIMT = IOS_A + o_p\left(n^{-1/2}\right) = q \times IR + o_p\left(n^{-1/2}\right),$$

where q is the dimension of parameter $\boldsymbol{\theta}$. If the model is correctly specified, we have

$$BIMT = q + O_p\left(n^{-1/2}\right).$$

Remark 3.2 *Following Proposition 3.1, we can see that $n^{1/2}(BIMT/q - 1)$ has the same asymptotic distribution as $n^{1/2}(IOS_A/q - 1)$ and $n^{1/2}(IR - 1)$. Hence, BIMT may be regarded as the MCMC-based version of IOS_A and IR. Different from IMT, IOS, IOS_A, BIMT is based on the MCMC output and hence is easier to obtain for some complex models, such as latent variable models.*

Remark 3.3 *However, since* $n^{1/2}\left(BIMT/q-1\right)$ *has the same asymptotic distribution as* $n^{1/2}\left(IOS_A/q-1\right)$ *and* $n^{1/2}\left(IR-1\right)$*, BIMT inherits the size distortion problem of* IOS_A *and IR and bootstrap methods must be used. This is why we do not use BIMT for specification testing directly. Instead it is used to construct the power enhancement function in our proposed test statistic.*

3.2 Power enhancement technique

Before we introduce our test statistic, it is important to review the power enhancement technique of Fan, et al (2015). Fan, et al considered the hypothesis testing problem of $H_0 : \boldsymbol{\theta} = \mathbf{0}$ where $\boldsymbol{\theta}$ is a high-dimensional vector. The alternative hypothesis H_1 is sparse so that the null hypothesis is violated by only a few components. They showed that traditional tests, such as the Wald test, have a low power. To enhance the power, they introduced a power enhancement component which is zero under the null hypothesis with high probability and diverges quickly under sparse alternatives.

Their new test statistic (call it J) has the form of

$$J = J_0 + J_1,$$

where J_1 is an asymptotically pivotal test statistic, such as Wald test, and J_0 is the power enhancement component. J_0 needs to satisfy three properties: (a) $J_0 \geq 0$ almost surely; (b) under H_0, $\Pr(J_0 = 0|H_0) \to 1$; (c) J_0 diverges in probability under some specific regions of H_1. Clearly, property (a) ensures that J is at least as powerful as J_1; property (b) guarantees that the asymptotic distribution of J under H_0 is determined by J_1 and hence the size of J is asymptotically equivalent to that of J_1; property (c) guarantees that the power of J improves that of J_1.

Motivated by this power enhancement technique, we propose a specification test based on the MCMC output. This new test combines a component (J_1) that tests a null point hypothesis in an expanded model and a power enhancement component (J_0) obtained from the original model to which we wish to perform the specification test.

3.3 A specification test based on the MCMC output

As in Fan et al (2015), our proposed test has two components, J_0 and J_1. To introduce J_1, we expand $p(\mathbf{y}|\boldsymbol{\theta})$, the model in concern, to a larger model denoted by $p\left(\mathbf{y}|\boldsymbol{\theta}_L\right)$ where $\boldsymbol{\theta}_L = \left(\boldsymbol{\theta}', \boldsymbol{\theta}_E'\right)'$ with $\boldsymbol{\theta}_E$ being a q_E-dimensional vector. So the expanded model $p\left(\mathbf{y}|\boldsymbol{\theta}_L\right)$ nests the original model $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$. We assume that if the specification $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$ is correct, then the true value of $\boldsymbol{\theta}_E$ is zero.

Let

$$
\begin{aligned}
s(\mathbf{y},\boldsymbol{\theta}_L) &= \frac{\partial \log p(\mathbf{y}|\boldsymbol{\theta}_L)}{\partial \boldsymbol{\theta}_L}, \\
C(\mathbf{y},\boldsymbol{\theta}_L) &= s(\mathbf{y},\boldsymbol{\theta}_L)\, s(\mathbf{y},\boldsymbol{\theta}_L)', \\
V(\bar{\boldsymbol{\theta}}_L) &= E\left[\left(\boldsymbol{\theta}_L-\bar{\boldsymbol{\theta}}_L\right)\left(\boldsymbol{\theta}_L-\bar{\boldsymbol{\theta}}_L\right)'|\mathbf{y}\right] = \int \left(\boldsymbol{\theta}_L-\bar{\boldsymbol{\theta}}_L\right)\left(\boldsymbol{\theta}_L-\bar{\boldsymbol{\theta}}_L\right)' p(\boldsymbol{\theta}_L|\mathbf{y})\mathrm{d}\boldsymbol{\theta}_L,
\end{aligned}
$$

where $\bar{\boldsymbol{\theta}}_L$ is the posterior mean of $\boldsymbol{\theta}_L$ in the expanded model. The J_1 component is designed to test the point null hypothesis $\boldsymbol{\theta}_E = 0$ after the expanded model is estimated by a Bayesian MCMC method. In particular, we follow Li, et al (2015) by considering a test statistic given by

$$
J_1 = \mathbf{tr}\left\{C_E\left(\mathbf{y},\left(\bar{\boldsymbol{\theta}},\boldsymbol{\theta}_E=0\right)\right) V_E\left(\bar{\boldsymbol{\theta}}_L\right)\right\}, \tag{5}
$$

where $C_E\left(\mathbf{y},\left(\bar{\boldsymbol{\theta}},\boldsymbol{\theta}_E=0\right)\right)$ is the submatrix of $C(\mathbf{y},\boldsymbol{\theta}_L)$ corresponding to $\boldsymbol{\theta}_E$ evaluated at $\left(\bar{\boldsymbol{\theta}},\boldsymbol{\theta}_E=0\right)$ and $V_E\left(\bar{\boldsymbol{\theta}}_L\right)$ is the submatrix of $V_E(\boldsymbol{\theta}_L)$ corresponding to $\boldsymbol{\theta}_E$ evaluated at $\bar{\boldsymbol{\theta}}_L$. As shown in Li, et al (2015), J_1 is a Bayesian version of Lagrange multiplier (LM; Breusch and Pagan, 1980) test and $J_1 \xrightarrow{d} \chi^2(q_E)$ when $\boldsymbol{\theta}_E = 0$. Typically, J_1 has good size property as it is designed to test the point null hypothesis.

If J_1 rejects the hypothesis $\boldsymbol{\theta}_E = 0$, it suggests that the original model $p(\mathbf{y}|\boldsymbol{\theta})$ is misspecified and indicates a source of model misspecification in $p(\mathbf{y}|\boldsymbol{\theta})$. Unfortunately, if J_1 fails to reject the hypothesis $\boldsymbol{\theta}_E = 0$, no conclusion can be drawn about the validity of the original model $p(\mathbf{y}|\boldsymbol{\theta})$. This is because, in practice, there are many different paths to expand the model. While J_1 may have good powers in some paths, it may have low powers in other paths. This problem is similar to that in the Wald statistic in the context of testing a high-dimensional vector against sparse alternatives.

To deal with this problem of low power, we introduce the following power enhancement component,

$$
J_0 = \sqrt{n}(\mathrm{BIMT}/q-1)^2, \tag{6}
$$

and propose a MCMC-based test statistic for model misspecification

$$
\mathrm{BMT} = J_1 + J_0 = \mathbf{tr}\left\{C_E\left(\mathbf{y},\left(\bar{\boldsymbol{\theta}},\boldsymbol{\theta}_E=0\right)\right) V_E\left(\bar{\boldsymbol{\theta}}_L\right)\right\} + \sqrt{n}(\mathrm{BIMT}/q-1)^2. \tag{7}
$$

In the following theorem, we establish the large sample properties of J_0 and J_1.

Theorem 3.1 *Under Assumptions 1-12 and if the model is correctly specified, we have,*

$$
J_1 \xrightarrow{d} \chi^2(q_E), \; J_0 = o_p(1), \; BMT \xrightarrow{d} \chi^2(q_E).
$$

If the model is misspecified with $q^* \neq q$, we have

$$
J_0 = \sqrt{n}\left(q^*/q-1\right)^2 + o_p(\sqrt{n}), \;\; BMT \sim O_p(\sqrt{n}),
$$

where $q^* = \mathbf{tr}\left[-\mathbf{H}\left(\boldsymbol{\theta}^*\right)^{-1}\mathbf{J}\left(\boldsymbol{\theta}^*\right)\right]$ *with* $\boldsymbol{\theta}^*$ *being the pseudo true value of* $\boldsymbol{\theta}$ *(Huber, 1967; White, 1982).*

Remark 3.4 *From (6) and Theorem 3.1, it is easy to see that* J_0 *is nonnegative almost surely and* $J_0 = o_p(1)$ *under* H_0*. In addition, Theorem 3.1 suggests that whenever* $q^* \neq q$*, as* $n \to \infty$*,* $J_0 \to \infty$*. Hence,* J_0 *satisfies the three power enhancement properties listed in Fan, et al. (2015). Since* $J_1 \xrightarrow{d} \chi^2(q_E)$ *and* $J_0 = o_p(1)$*, BMT is asymptotically pivotal (*χ^2*) under* H_0 *and the size distortion in BMT due to adding* J_0 *is asymptotically negligible. Under the alternative hypothesis in the region where* $q^* \neq q$*,* J_0 *diverges and dominates* J_1*, serving nicely as the power enhancement component.*

Remark 3.5 *It was noted earlier that* IOS_A*, IR and BIMT all have a complex asymptotic variance under* H_0*. In BMT, we do not use* J_0 *as the test statistic but as the power enhancement component. The asymptotic distribution of BMT under* H_0 *is determined by that of* J_1*. Since the establishment of asymptotic distribution of* J_1 *under* H_0 *requires relatively mild regularity conditions, BMT is applicable to a wide range of models.*

Remark 3.6 *BMT has several nice properties. First, compared with IM, IOS,* IOS_A *and IR, BMT is based on the MCMC output. When the likelihood function is difficult to optimize but the MCMC draws from the posterior distribution are available, BMT is easier to compute than IM, IOS,* IOS_A *and IR. Second, when* J_1 *does not have the size distortion problem, it is most likely that BMT will not suffer from size distortion. As a result, no bootstrap method is needed and intensive computational effort is avoided.*

Remark 3.7 *It is well documented in the specification test literature that most specification tests do not provide guidance to the possible source of model misspecification when the null hypothesis is rejected. Since our test relies on selecting particular paths for model expansion, if both BMT and* J_1 *are larger than the critical value, our approach not only suggests that the original model* $p(\mathbf{y}|\boldsymbol{\theta})$ *is misspecified but also indicates a source of model misspecification in* $p(\mathbf{y}|\boldsymbol{\theta})$*.*

Remark 3.8 *While* J_1 *depends on the path of model expansion,* J_0 *is always independent of paths. According to Theorem 3.1, as long as* $q^* \neq q$*,* $J_0 = O_p(\sqrt{n})$*. Hence, no matter which path the model is expanded in, even in the path where* J_1 *takes a very small value, BMT can still detect the model misspecification due to the power enhancement component.*

Remark 3.9 *Relative to* IOS_A*, IR and BIMT, the proposed test has a lower local power. This is the price we pay for avoiding using bootstrap methods. From Proposition 3.1 and*

Theorem 3.1, it is easy to show that IOS_A, IR and BIMT can detect the local misspecification that shrinks to the null at the rate of $n^{-1/2}$ (i.e. $q^* - q = O_p(n^{-1/2})$). Since J_0 is $O_p(1)$ when $q^* - q = O_p(n^{-1/4})$, BMT can detect the local misspecification that shrinks to the null at the rate of $n^{-1/4}$. This comparison suggests that one may define an alternative power enhancement function such as $J_0 = n^{\alpha}(BIMT/q - 1)^2$ for $\alpha \in (1/2, 1)$ to improve the local power. While the new J_0 can raise the local power, it introduces more size distortion to BMT. The analysis of such a trade-off is beyond the scope of the present paper.

Remark 3.10 *BMT will depend on the choice of prior. In general, a highly informative prior may have a strong influence on BMT. When BMT is used to test the model misspecification, we suggest the use of noninformative priors to avoid the dependence of BMT on priors in finite sample.*

Remark 3.11 *An important class models for which MCMC has been heavily used is state-space models. In Appendix 3, we discuss how to compute BMT in state-space models when the MCMC output is available.*

4 Simulation and Empirical Studies

In this section, we first design two simulation studies to check the finite sample performance of the proposed test. In the first simulation study, we test for heteroskedasticity in a linear regression model. This study aims to compare BMT with other popular tests in terms of size and power. In the second simulation study, we test the specification of a linear state-space model where existing misspecification tests are difficult to use but BMT is easier to obtain. Then, we consider empirical studies to examine the specification of three models and to highlight the usefulness of our test. The first model is a linear regression model. The second model is a linear state-space model where the existing tests are difficult to use. This third model is a stochastic volatility model where the existing tests are impossible to use.

4.1 Simulation Studies

4.1.1 Test for heteroskedasticity in a linear regression model

To do a Monte Carlo comparison of the IR test with other popular misspecification tests, Zhou et al (2012) considered the heteroskedasticity testing problem in a linear regression model. In our first simulation study, we adopt the simulation design of Zhou et al (2012) and compare the size and the power of BMT with those of the alternative tests. The linear regression model is specified as,

$$y_i = 1 + 2x_{i1} + 2x_{i2} + \epsilon_i, \epsilon_i = \sigma_i \xi_i, \xi_i \overset{i.i.d.}{\sim} N(0,1),$$

For this model, the covariates x_{i1} and x_{i2} are independently generated from the $U[-3,3]$ distribution. We would like to test the following null hypothesis of homoskedasticity, i.e.,

$$H_0 : Var(\epsilon_i) = \sigma_i^2 = \sigma^2, i = 1, 2, \cdots, n.$$

The DGP under the null hypothesis and the alterative hypothesis is, respectively,

$$H_0 : \sigma_i^2 = 1; \quad H_1 : \sigma_i^2 = \exp(x_{i1} + x_{i2}).$$

For the expanded model, we use

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i, \epsilon_i = \sigma \xi_i, \xi_i \sim N(0,1).$$

Following Zhou et al (2012), we run 2,000 replications, each of which has three different sample sizes, 50, 100, 200.

Table 1: The empirical size for IOS_A

	IOS_A	
n	Asymptotic distribution	Bootstrap distribution
50	0.216	0.049
100	0.147	0.050
200	0.136	0.056

Table 2: The empirical size for alternative tests

n	IR	IM	IOS	BMT
50	0.044	0.050	0.060	0.051
100	0.045	0.059	0.056	0.055
200	0.046	0.065	0.048	0.050

Table 3: Empirical power under the alternative hypothesis

n	IR	IM	IOS	BMT
50	0.85	0.11	0.9837	0.797
100	0.95	0.46	1	0.976
200	1.00	0.93	1	1.000

We first design an experiment to check the size distortion problem in IOS_A. Table 1 reports the size of IOS_A based on the asymptotic distribution and on the bootstrap distribution. The method used to obtain the asymptotic variance was proposed by Lancaster

(1984). It can be seen clearly that the size distortion is very large when the asymptotic distribution is used and the bootstrap method can solve the size distortion problem. In this example, MLE is trivial to compute and hence bootstrap methods are feasible.

To implement the proposed test, we need to use the Bayesian MCMC method to estimate the model under the null hypothesis and the expanded model. The conjugate vague priors for the hyper-parameters are set as

$$\mu_\beta = 0,\ V_\beta = 100 \times I, a = 0.01,\ b = 0.01,$$

where β is the vector of intercept and slope parameters and I is the identity matrix with dimension 3 for the null model, and with dimension 4 for the expanded model. In this example, since the posterior distribution is available analytically, we simply make 2,000 draws from the posterior directly.

Table 2 reports the empirical size of IR, IM, IOS and BMT under the null hypothesis and at the 5% significance level. The results of the first three tests are extracted from Zhou et al (2012) where the critical values are obtained from bootstrap methods. The BMT test entertains similar performance to the other test and shows the small size distortion in all cases.

Table 3 reports the empirical power of IR, IM, IOS and BMT at the 5% significance level. The results of the first two tests are extracted from Zhou et al (2012). From this table, it can be seen that the power of IOS is always the highest, followed closely by BMT and IR, while the power of IM can be very low (when n =50). The power of BMT is compatible with that of IR.

From this experiment we can conclude that the finite sample performance of BMT is satisfactory with small size distortion and good power. We should emphasize that the critical value of BMT is obtained from $\chi^2(1)$ and no bootstrap method is used.

4.1.2 A linear state-space model

The model under the null hypothesis is the following linear state-space model

$$\begin{aligned} R_t &= \beta_t R_{0t} + \varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\varepsilon^2\right), \\ \beta_{t+1} &= \bar{\beta} + \phi\left(\beta_t - \bar{\beta}\right) + \eta_t, \eta_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\eta^2\right). \end{aligned} \tag{8}$$

This random coefficient model has found many applications in economics and finance. While MLE of this model can be obtained by using the Kalman filter, the bootstrap method will be computationally costly for obtaining critical values for IM, IOS_A and IR. Consequently, we only implement BMT in this example.

The expanded model is

$$R_t = \alpha + \beta_t R_{0t} + \varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\varepsilon^2\right) \tag{9}$$

$$\beta_{t+1} = \bar{\beta} + \phi\left(\beta_t - \bar{\beta}\right) + \eta_t, \eta_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\eta^2\right),$$

where an intercept is added to the observation equation. If Model (8) is correctly specified, $\alpha = 0$ in the expanded model.

For Bayesian estimation, we use the following vague priors for the hyper-parameters,

$$\alpha \sim N(0, 10^3),\ \bar{\beta} \sim N(0, 10^3),\ \phi \sim Beta(1, 1),\ \sigma_\varepsilon^{-2} \sim \Gamma(10^{-3}, 10^{-3}),\ \sigma_\eta^{-2} \sim \Gamma(10^{-3}, 10^{-3}).$$

Based on 20,000 MCMC samples after 2,000 burning-in observations from the posterior distribution, we compute BMT. We run 1,000 replications, each of which has three different sample sizes, n =200, 400, 800.

To compute the empirical size, we set the parameter values at $\sigma_\varepsilon^2 = 0.000307$, $\bar{\beta} = 0.96$, $\phi = 0.5$, $\sigma_\eta^2 = 0.208$ and R_{0t} are generated from a i.i.d. normal distribution with mean 0 and variance 0.001. To compute the empirical power, we consider two different DGPs. The first DGP (denoted by $M1$) is given by

$$\begin{aligned} R_t &= \beta_t R_{0t} + \frac{\sigma_\varepsilon}{\sqrt{3}}\varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} t_3, \\ \beta_{t+1} &= \bar{\beta} + \phi\left(\beta_t - \bar{\beta}\right) + \eta_t, \eta_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\eta^2\right), \end{aligned} \tag{10}$$

where t_3 is a t distribution with 3 degrees of freedom, $\sigma_\varepsilon^2 = 0.000307$, $\bar{\beta} = 0.96$, $\phi = 0.5$, $\sigma_\eta^2 = 0.208$ and R_{0t} are generated from an i.i.d. normal distribution with mean 0 and variance 0.001. The second DGP for computing the power of BMT (denoted by $M2$) is given by

$$\begin{aligned} R_t &= \alpha + \beta_t R_{0t} + \frac{\sigma_\varepsilon}{\sqrt{3}}\varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} t_3, \\ \beta_{t+1} &= \bar{\beta} + \phi\left(\beta_t - \bar{\beta}\right) + \eta_t, \eta_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\eta^2\right), \end{aligned} \tag{11}$$

where $\alpha = 0.002$, $\sigma_\varepsilon^2 = 0.000307$, $\bar{\beta} = 0.96$, $\phi = 0.5$, $\sigma_\eta^2 = 0.208$ and R_{0t} are generated from an i.i.d. normal distribution with mean 0 and variance 0.001.

Table 4: Empirical size and empirical power

n	Empirical size	Empirical power ($M1$)		Empirical power ($M2$)	
		J_1	BMT	J_1	BMT
200	0.074	0.032	0.518	0.300	0.723
400	0.063	0.041	0.804	0.544	0.942
800	0.054	0.050	0.973	0.801	0.998

Table 4 reports the empirical size (at the 5% significance level) and the empirical power of BMT. To check whether or not J_1 is useful to provide the guidance about the source of misspecification, we also report the proportion of the 2,000 replications where J_1 rejects $\alpha = 0$ in the expanded model (9).

Several interesting findings come from Table 4. First, the size distortion is small and becomes better and better as the sample size increases, suggesting there is no need to use bootstrap methods. Second, the power is good and becomes higher and higher as the sample size increases. Third, the good power of BMT may not come from J_1 at all. In fact, J_1 loses power under $M1$. This finding is not surprising because $M1$ implies that $E(R_t|\beta_t, R_{0t}) = \beta_t R_{0t}$, suggesting the mean structure specified in the null model is correct and hence $\alpha = 0$. That is why J_1 only rejects $\alpha = 0$ at about 5% rate in the experiment. The power of BMT comes from the power enhancement component. In this case, unfortunately, J_1 does not provide the source of misspecification. Fourth, when the DGP is $M2$, $E(R_t|\beta_t, R_{0t}) = 0.002 + \beta_t R_{0t}$. The mean structure specified in the null model is wrong and hence $\alpha \neq 0$. In this case, J_1 rejects $\alpha = 0$ more often. When J_1 indeed rejects $\alpha = 0$, it suggests that the mean structure is the source of misspecification in Model (8).

4.2 Empirical studies

4.2.1 A linear regression model

In the first empirical study, we test misspecification of a model that explains arrest records. The data set contains data on arrests during the year 1986 and other information on 2,725 men born in either 1960 or 1961 in California. Each man in the sample was arrested at least once prior to 1986. Let y be the number of times the man was arrested during 1986, x_1, x_2, x_3, x_4 be the proportion (not percentage) of arrests prior to 1986 that led to conviction, average sentence length served for prior convictions, the months spent in prison in 1986, and the number of quarters during which the man was employed in 1986. See Wooldridge (2014) for more details. The null model is the following linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i, \varepsilon_i \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right). \tag{12}$$

The conjugated prior distributions for β ($:= \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix}'$) and σ^2 are set at

$$\beta \sim N\left(\mu_\beta, \sigma^2 V_\beta\right), \sigma^{-2} \sim \Gamma(a, b).$$

We use vague priors where the hyper-parameters in the priors are set at

$$\mu_\beta = 0,\ V_\beta = 100 \times I_5,\ a = 0.01,\ b = 0.01.$$

For the expanded model, we use

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i1}^2 + \varepsilon_i, \varepsilon_i \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right). \tag{13}$$

If Model (12) is correctly specified, $\beta_5 = 0$ in Model (13).

For the Bayesian MCMC analysis, 20,000 random draws are sampled from the posterior distribution. The posterior mean, standard deviation, 2.5% quantile, and 97.5% quantile of all the parameters are reported in Table 5 for both models.

Table 5: Posterior quantities of the null model and the expanded model

	Linear Regression Model				Expanded Model			
	Mean	SD	2.5 Percent	97.5 Percent	Mean	SD	2.5 Percent	97.5 Percent
β_0	0.7063	0.0330	0.6404	0.7729	0.6327	0.0355	0.5639	0.7027
β_1	-0.1515	0.0411	-0.2299	-0.0724	0.7851	0.1530	0.4801	1.0800
β_2	0.0074	0.0046	-0.0014	0.0164	0.0041	0.0048	-0.0053	0.0138
β_3	-0.0374	0.0088	-0.0545	-0.0203	-0.0441	0.0086	-0.0609	-0.0268
β_4	0.1032	0.0107	-0.1229	-0.0815	-0.0934	0.0105	-0.1134	-0.0735
σ^2	0.7068	0.0193	0.6686	0.7460	0.6971	0.0187	-0.9811	0.0786
β_5	-	-	-	-	-0.9811	0.1554	0.6609	0.7329

The critical value of $\chi^2(1)$ is 6.63 at the 1% significance level. In this study, the BMT statistic is 347.0783, suggesting that Model (13) is misspecified. It is easy to find out that J_1 is 37.7853 (i.e., J_0=309.2930) which is also greater than the 1% critical value of $\chi^2(1)$. Note that using J_1 we can reject $\beta_5 = 0$ in Model (13), suggesting that the misspecification of Model (13) comes from the wrong functional form in x_{i1}.

For this model it is easy to obtain IMT and feasible to obtain the critical value using a bootstrap method. IMT is 1732 and the 95% bootstrap critical value is 46.0734. Hence, IMT also suggest that Model (12) is misspecified, reinforcing the result from BMT. However, IMT does not tell the user how to improve the model.

4.2.2 A linear state-space model

In this section, we consider a capital asset pricing model (CAPM) with time-varying beta in a state-space form. Following Mergner and Bulla (2008), we specify the following model

$$\begin{aligned} R_{it} &= \beta_{it} R_{0t} + \varepsilon_{it}, \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_{i\varepsilon}^2\right), \\ \beta_{it+1} &= \bar{\beta}_i + \phi\left(\beta_{it} - \bar{\beta}_i\right) + \eta_{it}, \eta_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_{i\eta}^2\right), \end{aligned} \tag{14}$$

where R_{0t} denotes the excess return of the market portfolio and R_{it} denotes the excess return to sector i for period $t = 1, \ldots, T$. R_{0t} is the DJ STOXX 600 return index, which includes the 600 largest stocks in Europe, serves as a proxy for the overall market. The dataset used are weekly excess returns calculated from the total return indices for pan-European industry portfolios, covering the period from 2 December 1987 to 14 January 2016. The sample size is 1467. Here we choose the sector to be the insurance industry.

The expanded model is

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{it} R_{0t} + \varepsilon_{it}, \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma^2_{i\varepsilon}\right), \\ \beta_{it+1} &= \bar{\beta}_i + \phi_i \left(\beta_{it} - \bar{\beta}_i\right) + \eta_{it}, \eta_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma^2_{i\eta}\right), \end{aligned} \tag{15}$$

where an intercept is added to the mean equation. If Model (14) is correctly specified, $\alpha_i = 0$ in Model (15).

For Bayesian estimation, we use the vague priors for the hyper-parameters which are set as

$$\alpha_i \sim N(0, 10^3),\ \bar{\beta}_i \sim N(0, 10^3),\ \phi_i \sim Beta(1, 1),\ \sigma^{-2}_{i\varepsilon} \sim \Gamma(10^{-3}, 10^{-3}), \sigma^{-2}_{i\eta} \sim \Gamma(10^{-3}, 10^{-3}),$$

and draw 20,000 MCMC samples after 2,000 burning-in observations from the posterior distribution and compute BMT. We do not implement other tests as bootstrap methods are computationally too expensive. The posterior mean, standard deviation, 2.5% quantile, and 97.5% quantile of all the parameters are reported in Table 6 for both models (both α_i and $\sigma^2_{i\varepsilon}$ are multiplied by 10,000).

Table 6: Posterior quantities of the null model and the expanded model

	Linear State Space Model				Expanded Model			
	Mean	SD	2.5 Percent	97.5 Percent	Mean	SD	2.5 Percent	97.5 Percent
$\sigma^2_{i\varepsilon}$	1.3613	0.0740	1.2234	1.5140	1.3593	0.0756	1.2181	1.5153
$\bar{\beta}_i$	1.2157	0.0270	1.1620	1.2680	1.2186	0.0273	1.1650	1.2720
ϕ_i	0.4261	0.0945	0.2338	0.6007	0.4227	0.0948	0.2294	0.6031
$\sigma^2_{i\eta}$	0.1622	0.0255	0.1123	0.2134	0.1629	0.0260	0.1121	0.2151
α_i	-	-	-	-	-3.9009	3.6097	-11.0550	3.1275

The critical value of $\chi^2(1)$ is 6.63 at the 1% significance level. BMT is 124.4333, suggesting that Model (15) is misspecified. It is easy to find out that J_1 is 1.2553 (i.e., J_0=123.1780) which is less than the critical values of $\chi^2(1)$. Interestingly, using J_1 alone suggests that we cannot reject $\alpha_i = 0$ in Model (15).

4.2.3 A stochastic volatility (SV) model

The dataset used here contains the daily returns on AUD/USD exchange rates from January 2005 to December 2012. Following a suggestion of a referee, before we apply BMT to the SV model, we first test the i.i.d. normal model with constant mean and constant variance given by

$$y_t = \alpha + \varepsilon_t,\ \varepsilon_t \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right). \tag{16}$$

An AR(1) model is used as the expanded model

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t,\ \varepsilon_t \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right). \tag{17}$$

The Bayesian MCMC method is implemented to estimate the parameters with the following vague prior

$$\alpha \sim N(0, 100\sigma^2),\ \beta \sim N(0, 100\sigma^2),\ \sigma^{-2} \sim \Gamma(0.001, 0.001).$$

For the above two models, we draw 20,000 MCMC samples from the posterior distribution and compute BMT. The posterior mean, standard deviation, 2.5% quantile, and 97.5% quantile of all the parameters are reported in Table 7.

Table 7: Posterior quantities of the null model and the expanded model

	IID Normal				AR(1) Model			
	Mean	SD	2.5 Percent	97.5 Percent	Mean	SD	2.5 Percent	97.5 Percent
α	-0.0140	0.0201	-0.0536	0.0263	-0.0137	0.0204	-0.0539	0.0270
σ^2	0.8026	0.0259	0.7689	0.8727	0.8208	0.0255	0.7726	0.8737
β	-	-	-	-	-0.0115	0.0216	-0.0524	0.0287

The critical value of $\chi^2(1)$ is 6.63 at the 1% significance level. BMT is 251.52, rejecting the i.i.d. normal model. This conclusion is not surprising as the volatility of stock returns is stochastic. However, J_1 is 0.2858 (i.e., J_0=251.23) which is less than the critical value of $\chi^2(1)$. Using J_1 alone only suggests that we cannot reject $\beta = 0$ in Model (17). This conclusion is also not surprising as the weekly returns have very weak serial correlations.

Next, we change the null model to the following basic SV model,

$$\begin{aligned} y_t &= \alpha + \exp(h_t/2)\, u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1), \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \tau\nu_t,\ \ \nu_t \overset{i.i.d.}{\sim} N(0,1). \end{aligned} \tag{18}$$

The expanded model is as follows,

$$\begin{aligned} y_t &= \alpha + \beta_1 y_{t-1} + \exp(h_t/2)\, u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1). \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \tau\nu_t,\ \ \nu_t \overset{i.i.d.}{\sim} N(0,1). \end{aligned} \tag{19}$$

The following vague priors are used

$$\alpha \sim N(0, 100),\ \mu \sim N(0, 100),\ \ \phi \sim Beta(1, 1),\ \tau^{-2} \sim \Gamma(0.001, 0.001), \beta_1 \sim N(0.5, 100).$$

To obtain BMT, we draw 110,000 MCMC samples from the posterior distribution and discard the first 10,000 as burning-in observations, and store the remaining samples as effective observations in both models. The posterior mean, standard deviation, 2.5% quantile, and 97.5% quantile of all the parameters are reported in Table 8. BMT is calculated based on particle filters for which the calculation details are given in Appendix 5. In this case, BMT=0.4279 which is less than the critical value of $\chi^2(1)$, suggesting that the basic SV model is not misspecified.

Table 8: Posterior quantities of the null model and the expanded model

	Basic SV Model				Expanded Model			
	Mean	SD	2.5 Percent	97.5 Percent	Mean	SD	2.5 Percent	97.5 Percent
α	-0.0290	0.0138	-0.0560	-0.0021	-0.0300	0.0139	-0.0574	-0.0027
μ	-0.7518	0.2571	-1.2650	-0.2421	-0.7143	0.2499	-1.2180	-0.2142
ϕ	0.9905	0.0039	0.9821	0.9974	0.9901	0.0041	0.9813	0.9973
τ^2	0.0163	0.0035	0.0105	0.0239	0.0167	0.0036	0.0106	0.0245
β_1	-	-	-	-	-0.0153	0.0226	-0.0595	0.0291

5 Conclusions

In this paper, we have proposed a new specification test statistic based on the MCMC output to assess the adequacy of specification of a model. It combines a component (J_1) that tests a null point hypothesis in an expanded model and a power enhancement component (J_0) obtained from the null model. It is shown that J_0 converges to zero when the null model is correctly specified and diverges when the null model is misspecified. Also shown is that J_1 is asymptotically χ^2-distributed, suggesting that the proposed test is asymptotically pivotal, when the null model is correctly specified.

When J_1 does not suffer from the size distortion problem, the proposed test will have good size. Consequently, no bootstrap method is needed to correct the size. When J_1 loses power, the power enhancement component (J_0) raises the power of the proposed test. If J_1 rejects the null point hypothesis in an expanded model, it provides guidance of source of misspecification.

An important feature of the proposed test is that it is based on the MCMC output. While several specification tests based on the information matrix are available in the literature, they all require MLE as the input. Moreover, since the asymptotic distribution of these test performs poorly in finite sample, bootstrap methods have been suggested to calculate critical values, increasing the computational cost. For models where MCMC is a popular method, MLE is very difficult to obtain and bootstrap methods are computationally too expensive. This may help explain why no specification test has been carried out to these models in practice.

It is possible to introduce a ML-based test statistic of the same spirit. When MLE is not difficult to obtain but it is not easy to find a suitable bootstrap method or all bootstrap methods are too costly to implement, one can use a ML-based specification test with the power enhancement technique. This alternative test will be reported in a separate study.

6 Appendix

6.1 Appendix 1: Proof of Proposition 3.1

From the definition of BMT, we have,

$$\begin{aligned}
&\text{BIMT} = n\mathbf{tr}\left\{\hat{\mathbf{J}}_n(\bar{\boldsymbol{\theta}})E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]\right\} \\
= \;& \mathbf{tr}\left\{\left[\hat{\mathbf{J}}_n(\hat{\boldsymbol{\theta}})+o_p(n^{-1/2})\right]E\left[n(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]\right\} \\
= \;& \mathbf{tr}\left\{\left[\hat{\mathbf{J}}_n(\hat{\boldsymbol{\theta}})+o_p(n^{-1/2})\right]\left[-\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})+O_p(n^{-1})\right]\right\} \\
= \;& \mathbf{tr}\left[-\hat{\mathbf{J}}_n(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})\right]+o_p(n^{-1/2}) = \text{IOS}_A+o_p(n^{-1/2}) = q\times\text{IR}+o_p(n^{-1/2}),
\end{aligned}$$

since we have

$$\begin{aligned}
\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) &= \frac{1}{n}\sum_{t=1}^{n}s_t\left(\bar{\boldsymbol{\theta}}\right)s_t\left(\bar{\boldsymbol{\theta}}\right)' \\
&= \frac{1}{n}\sum_{t=1}^{n}\left[s_t\left(\hat{\boldsymbol{\theta}}\right)+h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)\right]\left[s_t\left(\hat{\boldsymbol{\theta}}\right)+h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)\right]' \\
&= \frac{1}{n}\sum_{t=1}^{n}s_t\left(\hat{\boldsymbol{\theta}}\right)s_t\left(\hat{\boldsymbol{\theta}}\right)'+\frac{2}{n}\sum_{t=1}^{n}h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)s_t\left(\hat{\boldsymbol{\theta}}\right)' \\
&\quad +\frac{1}{n}\sum_{t=1}^{n}h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)'h_t\left(\tilde{\boldsymbol{\theta}}_1\right)' \\
&= \hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}\right)+o_p(n^{-1/2}),
\end{aligned} \tag{20}$$

where $\tilde{\boldsymbol{\theta}}_1$ lies between $\bar{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$. To obtain (20), note that

$$\begin{aligned}
vec\left(\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right)\right) &= \frac{1}{n}\sum_{t=1}^{n}vec\left(s_t\left(\hat{\boldsymbol{\theta}}\right)s_t\left(\hat{\boldsymbol{\theta}}\right)'\right)+\frac{2}{n}\sum_{t=1}^{n}\left[s_t\left(\hat{\boldsymbol{\theta}}\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\right]vec\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right) \\
&\quad +\frac{1}{n}\sum_{t=1}^{n}\left[h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\right]vec\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right),
\end{aligned}$$

where vec is the column-wise vectorization. By Assumption 10, we have

$$\frac{2}{n}\sum_{t=1}^{n}\left[s_t\left(\hat{\boldsymbol{\theta}}\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\right]=O_p\left(1\right),\quad \frac{1}{n}\sum_{t=1}^{n}h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_1\right)=O_p\left(1\right),$$

and $\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}=o_p\left(n^{-1/2}\right)$ from Remark 3.1. Then we can get (20). And under Assumptions 1-12, following Li, Zeng and Yu (2015), we have

$$-\frac{1}{n}\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})=E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]+O_p(n^{-2}).$$

Proposition 3.1 is proven.

Next we give the proof for the order of BIMT$-q$. Note that, by Assumption 7,

$$\begin{aligned}\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right) &= \frac{1}{n}\sum_{t=1}^{n} h_t\left(\bar{\boldsymbol{\theta}}\right) = \frac{1}{n}\sum_{t=1}^{n} h_t\left(\boldsymbol{\theta}_0\right) + \frac{1}{n}\sum_{t=1}^{n} \bigtriangledown l^{(3)}\left(\tilde{\boldsymbol{\theta}}_2\right)\left[I_q \otimes \left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} h_t\left(\boldsymbol{\theta}_0\right) + O_p\left(n^{-1/2}\right) = \mathbf{H}\left(\boldsymbol{\theta}_0\right) + O_p\left(n^{-1/2}\right), \end{aligned} \tag{21}$$

where $\bigtriangledown l^{(3)}\left(\tilde{\boldsymbol{\theta}}_2\right)$ is the third order derivative of $l_t\left(\boldsymbol{\theta}\right)$ evaluated at $\tilde{\boldsymbol{\theta}}_2$, $\tilde{\boldsymbol{\theta}}_2$ lies between $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, $\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p\left(n^{-1/2}\right)$, $\mathbf{H}\left(\boldsymbol{\theta}_0\right) = \lim_{n\to\infty}\int \frac{1}{n}\sum_{t=1}^{n} h_t\left(\boldsymbol{\theta}_0\right) g\left(\mathbf{y}\right) d\mathbf{y}$.

$$\begin{aligned}\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) &= \frac{1}{n}\sum_{t=1}^{n} s_t\left(\bar{\boldsymbol{\theta}}\right) s_t\left(\bar{\boldsymbol{\theta}}\right)' \\ &= \frac{1}{n}\sum_{t=1}^{n}\left[s_t\left(\boldsymbol{\theta}_0\right) + h_t\left(\tilde{\boldsymbol{\theta}}_3\right)\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right]\left[s_t\left(\boldsymbol{\theta}_0\right) + h_t\left(\tilde{\boldsymbol{\theta}}_3\right)\left(\bar{\theta} - \boldsymbol{\theta}_0\right)\right]' \\ &= \frac{1}{n}\sum_{t=1}^{n} s_t\left(\boldsymbol{\theta}_0\right) s_t\left(\boldsymbol{\theta}_0\right)' + \frac{2}{n}\sum_{t=1}^{n} h_t\left(\tilde{\boldsymbol{\theta}}_3\right)\left(\bar{\theta} - \boldsymbol{\theta}_0\right) s_t\left(\boldsymbol{\theta}_0\right)' \\ &\quad + \frac{1}{n}\sum_{t=1}^{n} h_t\left(\tilde{\boldsymbol{\theta}}_3\right)\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' h_t\left(\tilde{\boldsymbol{\theta}}_3\right)', \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_3$ lies between $\bar{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Note that

$$\begin{aligned} vec\left(\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right)\right) &= \frac{1}{n}\sum_{t=1}^{n} vec\left(s_t\left(\boldsymbol{\theta}_0\right) s_t\left(\boldsymbol{\theta}_0\right)'\right) + \frac{2}{n}\sum_{t=1}^{n}\left[s_t\left(\boldsymbol{\theta}_0\right) \otimes h_t\left(\tilde{\boldsymbol{\theta}}_3\right)\right] vec\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\ &\quad + \frac{1}{n}\sum_{t=1}^{n}\left[h_t\left(\tilde{\boldsymbol{\theta}}_3\right) \otimes h_t\left(\tilde{\boldsymbol{\theta}}_3\right)\right] vec\left(\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\left(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)'\right). \end{aligned}$$

Hence, similar to Equation (20) and by Assumption 7, we have

$$\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) = \mathbf{J}\left(\boldsymbol{\theta}_0\right) + O_p\left(n^{-1/2}\right), \tag{22}$$

where $\mathbf{J}\left(\boldsymbol{\theta}_0\right) = \lim_{n\to\infty}\int \frac{1}{n}\sum_{t=1}^{n} s_t\left(\boldsymbol{\theta}_0\right) s_t\left(\boldsymbol{\theta}_0\right)' g\left(\mathbf{y}\right) d\mathbf{y}$. If the model is correctly specified, $-\mathbf{H}\left(\boldsymbol{\theta}_0\right) = \mathbf{J}\left(\boldsymbol{\theta}_0\right)$, then we have

$$-\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right) = \hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) + O_p\left(n^{-1/2}\right). \tag{23}$$

From Li, Zeng and Yu (2015), under Assumptions 1-12, by the Laplace expansion,

$$\mathbf{tr}\left(-n\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right) V\left(\bar{\boldsymbol{\theta}}\right)\right) = q + O_p\left(n^{-1}\right). \tag{24}$$

Hence, from (23) and (24), we have

$$\begin{aligned}\mathrm{BIMT} &= \mathbf{tr}\left(n\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) V\left(\bar{\boldsymbol{\theta}}\right)\right) = \mathbf{tr}\left(n\left(-\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right) + O_p\left(n^{-1/2}\right)\right) V\left(\bar{\boldsymbol{\theta}}\right)\right) \\ &= \mathbf{tr}\left(-n\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right) V\left(\bar{\boldsymbol{\theta}}\right)\right) + O_p\left(n^{-1/2}\right) \\ &= q + O_p\left(n^{-1/2}\right). \end{aligned}$$

6.2 Appendix 2: Proof of Theorem 3.1

According to Li, et al (2015), if $\boldsymbol{\theta}_E = 0$ in the expanded model, as $n \to \infty$,

$$J_1 = \mathbf{tr}\left\{C_E\left(\mathbf{y}, (\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}_E = 0)\right) V_E\left(\bar{\boldsymbol{\theta}}_L\right)\right\} \xrightarrow{d} \chi^2(q_E)$$

By Proposition 3.1, we have

$$\begin{aligned}
& \text{BIMT} = \mathbf{tr}\left[-\hat{\mathbf{J}}_n(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})\right] + o_p(1) \\
= \;& \mathbf{tr}\left\{-\left[\hat{\mathbf{J}}_n(\boldsymbol{\theta}^*) + o_p(1)\right]\left[\hat{\mathbf{H}}_n^{-1}(\boldsymbol{\theta}^*) + o_p(1)\right]\right\} + o_p(1) \\
= \;& \mathbf{tr}\left\{-\hat{\mathbf{J}}_n(\boldsymbol{\theta}^*)\hat{\mathbf{H}}_n^{-1}(\boldsymbol{\theta}^*)\right\} + o_p(1) \\
= \;& \mathbf{tr}\left\{-\left[\mathbf{J}(\boldsymbol{\theta}^*) + o_p(1)\right]\left[\mathbf{H}^{-1}(\boldsymbol{\theta}^*) + o_p(1)\right]\right\} + o_p(1) \\
= \;& \mathbf{tr}\left[-\mathbf{J}(\boldsymbol{\theta}^*)\mathbf{H}^{-1}(\boldsymbol{\theta}^*)\right] + o_p(1) = q^* + o_p(1),
\end{aligned}$$

where $\boldsymbol{\theta}^*$ is the pseudo true value (Huber, 1967, White, 1982). Clearly, $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ when the model is correctly specified.

Hence, if the model is misspecified, we have

$$J_1 = \sqrt{n}(\text{BIMT}/q - 1)^2 = \sqrt{n}\left(q^*/q - 1 + o_p(1)\right)^2 = \sqrt{n}\left(q^*/q - 1\right)^2 + o_p(\sqrt{n}).$$

6.3 Appendix 3: Computing BMT in Latent Variable Models

MCMC has been popular for estimate an important class of latent variable models – the state-space models. We now discuss how to compute BMT for the state-space models after they are estimated by MCMC. To introduce the state-space model, let $\mathbf{y}$ be the observed variables and $\mathbf{z} = (z_1, \ldots, z_n)$ be the latent variables. The model is given by

$$\left\{\begin{array}{c} y_t = F(z_t, u_t, \boldsymbol{\theta}) \\ z_t = G(z_{t-1}, v_t, \boldsymbol{\theta}) \end{array}\right. . \tag{25}$$

The first equation is the observation equation while the second equation is the state equation. When the distribution of u_t and v_t is Gaussian and 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

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}. \tag{26}$$

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 (26) 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.

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 $p(\boldsymbol{\theta}, \boldsymbol{z}|\mathbf{y})$.

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 particle filters to calculate $p(\mathbf{y}|\boldsymbol{\theta})$ and its derivatives with respect to $\boldsymbol{\theta}$.

6.3.1 Computing BMT 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

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = E_{\boldsymbol{\theta}^{(r)}}\{\mathcal{L}_c(\mathbf{y},\boldsymbol{z}|\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^{(r)}\}, \tag{27}$$

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 the 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 particle filters.

6.3.2 Computing BMT 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. One may refer to Harvey (1989) for the detailed textbook treatment of the linear Gaussian state-space model and the calculation of the observed-data log-likelihood recursively.

Similarly, the first order derivative of the observed-data log-likelihood, $\mathbf{s}_t(\boldsymbol{\theta})$, has to be computed recursively. In Appendix 4, we give the expression of the relevant first order derivatives that are used to compute BMT.

6.3.3 Computing BMT by particle filters

In practice, the phenomenon of non-Gaussianity or non-linearity is often found. Consequently, the nonlinear non-Gaussian state-space models have been widely used in empirical studies. 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 filters 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 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 $t = 1$, $z_1^{(i)} \sim \mu\left(\cdot\right)$,

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

$\mathbf{z}^{1(i)} = z_1^{(i)}$. 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 $q_t\left(\cdot|\cdot\right)$ 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 $q_t\left(\cdot|\cdot\right)$, 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 filter 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 particle filters.

6.4 Appendix 4: The derivation of BMT for the linear state-space model

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 $\varepsilon_t\sim N\left(0,Q\right)$, $\xi_t\sim N\left(0,H\right)$. Let $Y_s=(y_1,y_2...,y_s)$. We define

$$\begin{aligned}x_{t|s} &= E\left(x_t|Y_s\right),\\ P_{t|s} &= E\left[\left(x_t-x_{t|s}\right)\left(x_t-x_{t|s}\right)'|Y_s\right].\end{aligned}$$

With the initial condition $x_{0|0}$ and $P_{0|0}$, 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 $K_t = P_{t|t-1} Z' \left[Z P_{t|t-1} Z' + H \right]^{-1}$, for $t = 1, 2 \ldots n$.

From the Kalman filter, the observed data likelihood 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 6.1 *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, that contains 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 6.2 *Let A be an $m \times n$ matrix. There exists a unique $mn \times mn$ permutation matrix K_{mn} which is defined as:*

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

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

$$
\frac{\partial vec(\omega_t)}{\partial \boldsymbol{\theta}'} = -\frac{\partial vec(D)}{\partial \boldsymbol{\theta}'} - \left(x_{t|t-1}' \otimes I_{n_y} \right) \frac{\partial vec(Z)}{\partial \boldsymbol{\theta}'} - (I_1 \otimes Z) \frac{\partial vec\left(z_{t|t-1} \right)}{\partial \boldsymbol{\theta}'},
$$

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

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

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

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

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

Given the initial conditions $x_{0|0}$ and $P_{0|0}$, we have the following recursions

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

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

$$\begin{aligned}\frac{\partial vec\left(x_{t|t}\right)}{\partial \boldsymbol{\theta}'} &= \frac{\partial vec\left(x_{t|t-1}\right)}{\partial \boldsymbol{\theta}'} + \left[\left(y_t - D - Zx_{t|t-1}\right)' \otimes I_{n_s}\right] \frac{\partial vec\left(K_t\right)}{\partial \boldsymbol{\theta}'} \\ &\quad - \left(I_1 \otimes K_t\right) \frac{\partial vec\left(D\right)}{\partial \boldsymbol{\theta}'} - \left(z_{t|t-1}' \otimes K_t\right) \frac{\partial vec\left(Z\right)}{\partial \boldsymbol{\theta}'} - \left(I_1 \otimes K_t Z\right) \frac{\partial vec\left(z_{t|t-1}\right)}{\partial \boldsymbol{\theta}'},\end{aligned}$$

$$\begin{aligned}\frac{\partial vec\left(P_{t|t}\right)}{\partial \boldsymbol{\theta}'} &= -\left(\left(ZP_{t|t-1}\right)' \otimes I_{n_s}\right) \frac{\partial vec\left(K_t\right)}{\partial \boldsymbol{\theta}'} - \left(P_{t|t-1}' \otimes K_t\right) \frac{\partial vec\left(Z\right)}{\partial \boldsymbol{\theta}'} \\ &\quad + \left(I_{n_s} \otimes \left(I_{n_s} - K_t Z\right)\right) \frac{\partial vec\left(P_{t|t-1}\right)}{\partial \boldsymbol{\theta}'},\end{aligned}$$

where

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

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

$$\begin{aligned} x_{0|0} &= 0, \\ P_{0|0} &= TP_{0|0}T' + RQR'. \end{aligned}$$

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

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

6.5 Appendix 5: The derivation of BMT 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 recursion

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

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

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

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

Then we have

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

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 recursion

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

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

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

Then we have

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

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).

References

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

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

Bester, C.A. and Hansen, C. (2006). Bias reduction for Bayesian and frequentist estimators. *SSRN Working Paper Series*

Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, **47**, 239-253.

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

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.

Chernozhukov, V. and Hong, H. (2003), An MCMC approach to classical estimation. *Journal of Econometrics*, **115(2)**, 293–346

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

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.

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.

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, J. Q., Liao,Y. and Yao, J. W. (2015). Power Enhancement in High-Dimensional Cross-Sectional Tests. *Econometrica*,**83(4)**, 1497-1541.

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

Geweke, J. F. (2005). *Contemporary Bayesian Econometrics and Statistics.* Vol. 537, Wiley-Interscience

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

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F*, **140**, 107–113.

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

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.

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.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, **1(1)**, 221-233.

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

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

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.

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

Li, Y., Liu, X. B. and Yu, J. (2015). A Bayesian chi-squared test for hypothesis testing. *Journal of Econometrics*, **189**, 54-69.

Li, Y., Zeng, T. and Yu, J. (2015). Robust deviation information criterion for latent variable models. *Working paper*, Singapore Management University.

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

McFadden, D. (1989) A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, **57(5)**, 995-1026.

Mergner, S. and Bulla, J. (2008). Time-varying beta risk of Pan-European industry portfolios: A comparison of alternative modeling techniques. *The European Journal of Finance*, **14(8)**, 771-802.

Miyata, Y. (2004) Fully exponential Laplace approximations using asymptotic modes. *Journal of the American Statistic Association*, **81**, 82-86.

Miyata, Y. (2010) Laplace approximations to means and variances with asymptotic modes. *Journal of Statistical Planning and Inference*, **140**, 382-392.

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

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

Newey, W. K., Smith, R. J. (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, **72(1)**, 219-255.

Norets, A. (2009). Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables. *Econometrica*, **77(5)**, 1665-1682.

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.

Rilstone, P., Srivatsava, V. K., and Ullah, A. (1996). The second order bias and MSE of nonlinear estimators . *Journal of Econometrics*, **75**, 369-395.

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

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.

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 of Econometric Society. **1**: 1-58.

Wooldridge, J M. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory*, **8(04)**, 452-475.

Wooldridge, J. M. (2014). *Introductory Econometrics: A Modern Approach.* Nelson Education.

Zhou, Q. M., Song, P. X. K., and Thompson, M. E. (2012). Information ratio test for model misspecification in quasi-likelihood inference. *Journal of the American Statistical Association*, **107(497)**, 205-213.

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