Specification Tests based on MCMC Output*

Yong Li
Renmin University of China

Jun Yu
Singapore Management University

Tao Zeng
Zhejiang University

July 26, 2018

Abstract

Two test statistics are proposed to determine model specification after a model is estimated by an MCMC method. The first test is the MCMC version of IOS_A test and its asymptotic null distribution is normal. The second 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 first test. 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 second test is asymptotically pivotal, when the null model is correctly specified. The main feature of the first test is that no alternative model is needed. The second test has several properties. First, its size distortion is small and hence bootstrap methods can be avoided. Second, it is easy to compute from 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 null model and J_1 takes a large value, the test suggests the source of misspecification. 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, Xiaobin Liu, Han Chen, Xin Tang 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.71773130) 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. Zeng gratefully acknowledges the financial support from the Chinese Natural Science fund (No.1202-2109000045 and 91646206). 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, School of Economics and Academy of Financial Research, Zhejiang University, Zhejiang, 310058, China. Email: ztzt6512@gmail.com. Author for correspondence: Tao Zeng.

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 reduced forms 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 Lagrange 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), Wooldridge (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 (MLE) 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 Markov chain Monte Carlo (MCMC) methods to conduct Bayesian posterior analysis of econometric models. With the rapid growth in computer capability, fitting models of increasing complexity has become easier and easier by MCMC.

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 specification tests 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 a model specification? Second, is it possible to tell the source of model misspecification if the null model is rejected?

We propose two new specification tests based on MCMC output. The first test is the MCMC version of IOS_A of Presnell and Boos (2004) and its asymptotic null distribution is normal. The second test is our main statistic which is motivated by the power enhancement technique of Fan, et al (2015) and based on a model expansion strategy. 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 first test. 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 main feature of the first test is that no alternative model is needed. The second test has several properties. First, its size distortion is small and hence bootstrap methods can be avoided. Second, it is easy to compute from 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 a null model and J_1 takes a large value, our test suggests the source of misspecification. However, the proposed test has a lower local power. This is the price we pay for avoiding using a bootstrap method.

The paper is organized as follows. Section 2 proposes the two test statistics based on MCMC output and establishes their asymptotic properties. Section 3 illustrates the method using two simulation studies and three empirical studies. Section 4 concludes the paper. Appendix collects the proof of the theoretical results in the paper and discusses how to compute the two test statistics in the context of state-space models. Proofs of Theorem 2.2 are provided in an online supplement.

2 Two Specification Tests based on 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 an MCMC version of an ML-based specification test. This is a reasonable way to proceed as both ML and MCMC are full-likelihood-based approaches.

2.1 An MCMC-based information matrix test

In this subsection, we propose an MCMC-based information matrix test. First we need to introduce some notations. 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. 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

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

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

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

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

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,

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

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

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

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

Assumption 8: $\boldsymbol{\theta}_0 \in int\left(\boldsymbol{\Theta}\right)$ 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 $\mathcal{F}_{-\infty}^{t} = \sigma\left(y_t, y_{t-1}, \ldots\right)$ and $\mathcal{F}_{t+m}^{\infty} = \sigma\left(y_{t+m}, y_{t+m+1}, \ldots\right)$, the mixing coefficient $\alpha\left(m\right) = O\left(m^{\frac{-2r}{r-2}-\varepsilon}\right)$ 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 $\boldsymbol{\Theta}$, $\bigtriangledown^j l_t\left(\boldsymbol{\theta}\right)$ exists, $\sup_{\boldsymbol{\theta}\in\mathcal{G}} \left\|\bigtriangledown^j l_t\left(\boldsymbol{\theta}\right)\right\| \leqslant M_t(y_t)$, and $\sup_t E\left\|M_t\left(y_t\right)\right\|^{r+\delta} \leq M < \infty$ for some $\delta > 0$.

Assumption 11: $\left\{\bigtriangledown^j l_t\left(\boldsymbol{\theta}\right)\right\}$ 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 $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$, $\left\|\bigtriangledown^j l_t\left(\boldsymbol{\theta}\right) - \bigtriangledown^j l_t\left(\boldsymbol{\theta}'\right)\right\| \leq c_t\left(\mathbf{y}^t\right)\left\|\boldsymbol{\theta} - \boldsymbol{\theta}'\right\|$ for $0 \leqslant j \leqslant 3$ in probability, where $c_t\left(\mathbf{y}^t\right)$ is a positive random variable, $\sup_t E\left\|c_t\left(\mathbf{y}^t\right)\right\| < \infty$ and $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}\left(c_t - Ec_t\right) \xrightarrow{p} 0$.

Remark 2.1 *Assumption 1-4 have been used to develop Bayesian large sample theory; see, for example, Chen (1985), Kim (1994, 1998), Geweke (2005). Similar assumptions have been used to develop 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). Assumption 5 is a standard regularity condition to define the pseudo-true value; see Huber (1967), White (1982) and 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 7-12 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 these assumptions, Li, Yu and Zeng (2017) showed that,*

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

Before we introduce our test statistics, it is important to review some leading specification tests based on MLE. One of the earliest specification tests is based on the information matrix equivalence due to White (1982). Under the null hypothesis that the model is correctly specified, it is well-known that $\mathbf{H}(\boldsymbol{\theta}) + \mathbf{J}(\boldsymbol{\theta}) = 0$. White (1982) proposed the following information matrix test

$$\text{IMT} = nD_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) V_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right) D_n\left(\hat{\boldsymbol{\theta}}_{ML}\right), \tag{2}$$

where $\hat{\boldsymbol{\theta}}_{ML}$ is the MLE of $\boldsymbol{\theta}$, and

$$V_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) = \frac{1}{n}\sum_{t=1}^{n} \nu_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)\nu_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)',$$

$$\nu_t\left(\hat{\boldsymbol{\theta}}_{ML}\right) = d\left(y_t, \hat{\boldsymbol{\theta}}_{ML}\right) - \dot{D}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)\hat{\mathbf{H}}_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\mathbf{s}\left(y_t, \hat{\boldsymbol{\theta}}_{ML}\right),$$

$$D_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) = \frac{1}{n}\sum_{t=1}^{n} d\left(y_t, \hat{\boldsymbol{\theta}}_{ML}\right), \dot{D}_n = \frac{\partial D_n}{\partial \boldsymbol{\theta}}, d(\mathbf{y}, \boldsymbol{\theta}) = vech\left[\mathbf{h}(\mathbf{y}, \boldsymbol{\theta}) + \mathbf{s}(\mathbf{y}, \boldsymbol{\theta})\mathbf{s}'(\mathbf{y}, \boldsymbol{\theta})\right].$$

Based on a set of regularity conditions, White (1982) showed that $\text{IMT} \xrightarrow{d} \chi^2$ as $n \to \infty$ under the null hypothesis.

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:

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

and showed that the asymptotic form of IOS is

$$\text{IOS}_A = \mathbf{tr}\left[-\hat{\mathbf{H}}_n^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)\right], \tag{3}$$

and $\text{IOS} - \text{IOS}_A = o_p\left(n^{-1/2}\right)$. Like IMT, IOS_A also compares $\hat{\mathbf{H}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)$ with $\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)$, but in a ratio form instead of an additive form. Under the null hypothesis, $\text{IOS}_A \xrightarrow{p} q$ and $n^{1/2}\left(\text{IOS}_A - q\right)$ converges to a normal distribution with zero mean and a very complicated variance. Clearly, IOS and IOS_A are asymptotically equivalent. Zhou, et al (2012) proposed a test statistic that 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 $n^{1/2}\left(\text{IR} - 1\right)$ converges to a normal distribution with zero mean and a very complicated variance.

Unfortunately, it is well-documented that the asymptotic distributions poorly approximate their finite sample counterparts for IMT, IOS, IOS_A. As a result, they all suffer from serious bias distortions if asymptotic distributions are used to obtain critical values. See Orme (1990), Chesher and Spady (1991), Davidson and Mackinnon (1992), Horowitz (1994) for evidence of severe oversized problem for IMT. Presnell and Boos (2004) showed that the convergence of IOS statistic to normality is slow by simulation so they proposed to obtain the critical values by parametric bootstrap. 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. Naturally, to reduce the size distortion, the bootstrap methods can be advocated to be implemented for calibrating better critical values, see Horowitz (1994), Presnell and Boos (2004) and Zhou et al (2012).

Based on Remark 2.1 and the expression of IOS_A given in Equation (3), if we replace $-\hat{\mathbf{H}}_{\mathbf{n}}^{-1}\left(\hat{\boldsymbol{\theta}}_{ML}\right)$ with $nV\left(\bar{\boldsymbol{\theta}}\right)$ and $\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)$ with $\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right)$, a natural MCMC-based information matrix test (which is our first test statistic) can be defined as:

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

Proposition 2.1 *Under Assumptions 1-12, we have*

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

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 2.2 *Following Proposition 2.1 and the discussion in Section 2, we can see that $n^{1/2}\left(BIMT/q - 1\right)$ has the same asymptotic distribution as $n^{1/2}\left(IOS_A/q - 1\right)$ and $n^{1/2}\left(IR - 1\right)$. Hence, BIMT may be regarded as the MCMC version of IOS_A. As IMT and IOS, BIMT does not require an alternative model be specified. Different from IMT, IOS and IOS_A, BIMT is based on MCMC output and hence is easier to obtain for some complex models, such as latent variable models. However, simulation studies that will be reported in Section 3.1.1 show that BIMT suffers from severe size distortion. Hence, bootstrap methods must be used, greatly increasing the computational cost.*

2.2 Power enhancement technique

The size problem and the computational cost for the first statistic point to a need for another test statistic. Before we introduce our second 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 a 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 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 first test.

2.3 The main specification test

As in Fan et al (2015), our second 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 $p(\mathbf{y}|\boldsymbol{\theta}_L)$ 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 $p(\mathbf{y}|\boldsymbol{\theta}_L)$ nests the original model $p(\mathbf{y}|\boldsymbol{\theta})$. We assume that if the specification $p(\mathbf{y}|\boldsymbol{\theta})$ 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 an 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 $C\left(\mathbf{y}, \boldsymbol{\theta}_L\right)$ 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 $V\left(\boldsymbol{\theta}_L\right)$ corresponding to $\boldsymbol{\theta}_E$ evaluated at $\bar{\boldsymbol{\theta}}_L$. As shown in Li, et al (2015), J_1 is a MCMC-version of LM test (Breusch and Pagan, 1980) and $J_1 \xrightarrow{d} \chi^2\left(q_E\right)$ 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 $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$ is misspecified and indicates the source of model misspecification in $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$. 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 $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$. 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, as well explained in Fan et al (2015).

To deal with this problem of low power, we introduce a power enhancement component to improve the power based on BIMT, that is,

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

and propose the following MCMC-based test 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 large sample properties for J_0, J_1 and BMT.

Theorem 2.1 *Under Assumptions 1-12, when the model is correctly specified, we have,*

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

Let $q^* := \mathbf{tr}\left[-\mathbf{H}\left(\boldsymbol{\theta}_0\right)^{-1} \mathbf{J}\left(\boldsymbol{\theta}_0\right)\right]$. *If the model is misspecified with* $q^* \neq q$, *we have*

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

so that the order of the power of BMT is no less than $O_p(\sqrt{n})$.

Remark 2.3 *From Equation (6) and Theorem 2.1, it is easy to see that* J_0 *satisfies the three power enhancement properties listed in Fan, et al. (2015). Since* $J_1 \xrightarrow{d} \chi^2\left(q_E\right)$ *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* H_1 *in the region where* $q^* \neq q$, J_0

diverges and dominates J_1, serving nicely as a power enhancement component. 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 the source of model misspecification in $p(\mathbf{y}|\boldsymbol{\theta})$.

Remark 2.4 *BMT has several nice properties. First, compared with IMT, IOS, IOS_A and IR, BMT is based on 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 the others. 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. In addition, BMT can be obtained under other simulation-based approaches, such as sequential Monte Carlo methods, as suggested by a referee. In addition, by incorporating BIMT into J_0, there is no need to calculate the complicated asymptotic variance of BIMT. These important properties make BMT applicable to a wide range of models.*

Remark 2.5 *While J_1 depends on the path of model expansion, J_0 is always independent of model expansion. According to Theorem 2.1, 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 2.6 *Relative to IOS_A, IR and BIMT, BMT has a lower local power. This is the price we pay for avoiding using bootstrap methods. From Proposition 2.1 and Theorem 2.1, 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 2.7 *Informative priors impose tight constraints on parameters so that the posterior covariance matrix and hence BIMT and BMT can be sensitive to priors. To minimize the impact of priors, we suggest the use of non-informative priors or flat priors when implementing our tests.*[1]

BMT requires selecting an auxiliary model to expand the original model. When the model is misspecified such that $q^* \neq q$, *BMT* can always detect the misspecification

[1]Loosely speaking, a non-informative prior is referred to a prior with big variance in our paper.

asymptotically regardless of choice of auxiliary model. However, auxiliary models will affect the size and power properties of BMT in finite samples. In general, it is very difficult to specify the "optimal" auxiliary model, as in indirect inference (Gourieroux, et al, 1993) where an auxiliary model is also needed. Here we provide some practical guidelines on how to choose an auxiliary model.

First, it is generally preferable to expand the conditional mean than to expand the conditional variance. Assume that, for a statistical model, the conditional mean and conditional variance are $E(y|F)$ and $Var(y|F)$, respectively, where F is the information set. Under mild regularity conditions, it is known that if $E(y|F)$ is misspecified, parameters in $E(y|F)$ are often inconsistently estimated, whereas if $Var(y|F)$ is misspecified, parameters in $E(y|F)$ can be consistently estimated (White, 1982). According to this property, if $E(y|F)$ is correctly specified, whether $Var(y|F)$ is correctly specified or not, J_1 will take a small value. Now consider the following two cases. In the first case $E(y|F)$ is expanded while in the second case $Var(y|F)$ is expanded. First consider the case when $Var(y|F)$ is misspecified, $E(y|F)$ is correctly specified, and $E(y|F)$ is expanded. In this case J_1 takes a small value. Since BMT rejects H_0, together with a small value for J_1, it suggests that the source of misspecification is in the conditional variance but not in the conditional mean. Second consider the case when $E(y|F)$ is misspecified, $Var(y|F)$ is correctly specified, and $Var(y|F)$ is expanded. In this case, J_1 takes a large value and rejects H_0, incorrectly indicating that the source of misspecification is in the conditional variance. This strategy for expanding the conditional mean even when the conditional variance is misspecified will be implemented in the third empirical example in Section 3.

Second, choice of an auxiliary model can be guided by economic theory. In the second and third empirical examples, we show how to choose auxiliary models using asset pricing theories. Third, choice of an auxiliary model can be guided by computational cost. It is important to specify an auxiliary model that can be quickly estimated. If not, BMT will be difficult to compute. Fourth, as usual, the law of parsimony is applicable. That is, when alternative auxiliary models with the same structure but different number of parameters are available, the model with the smallest number of parameters should be tried first because the simplest solution tends to be the right one. Of course, there is a size-power tradeoff here.

2.4 The proposed tests based on MCMC output

Asymptotic properties of BMT have been established based on $\bar{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{y})$ and $V\left(\bar{\boldsymbol{\theta}}\right) = E\left[(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})'|\mathbf{y}\right]$. In practice, however, analytical expressions for $\bar{\boldsymbol{\theta}}$ and $V\left(\bar{\boldsymbol{\theta}}\right)$ are often not available and some consistent estimates of $\bar{\boldsymbol{\theta}}$ and $V\left(\bar{\boldsymbol{\theta}}\right)$ based on MCMC output have to be used to approximate $\bar{\boldsymbol{\theta}}$ and $V\left(\bar{\boldsymbol{\theta}}\right)$. Let $\left\{\boldsymbol{\theta}_n^{(m)}\right\}_{m=1}^{M}$ be MCMC draws from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ of the null model. A consistent estimate of the posterior mean $\bar{\boldsymbol{\theta}}$ and the posterior variance $V\left(\bar{\boldsymbol{\theta}}\right)$ is given by

$$\widetilde{\boldsymbol{\theta}} = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\theta}_n^{(m)}, \widetilde{V}(\widetilde{\boldsymbol{\theta}}) = \frac{1}{M}\sum_{m=1}^{M}\left(\boldsymbol{\theta}_n^{(m)} - \widetilde{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}_n^{(m)} - \widetilde{\boldsymbol{\theta}}\right)'.$$

Similarly, let $\left\{\boldsymbol{\theta}_{Ln}^{(m)}\right\}_{m=1}^{M}$ be MCMC draws from the posterior distribution $p(\boldsymbol{\theta}_L|\mathbf{y})$ of the expanded model. Then,

$$\widetilde{\boldsymbol{\theta}}_L = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\theta}_{Ln}^{(m)}, \widetilde{V}(\widetilde{\boldsymbol{\theta}}_L) = \frac{1}{M}\sum_{m=1}^{M}\left(\boldsymbol{\theta}_{Ln}^{(m)} - \widetilde{\boldsymbol{\theta}}_L\right)\left(\boldsymbol{\theta}_{Ln}^{(m)} - \widetilde{\boldsymbol{\theta}}_L\right)'.$$

Based on these estimates of posterior moments, MCMC-based estimates of BIMT and BMT can be obtained as

$$\begin{aligned}
\widetilde{BIMT} &= n\mathbf{tr}\left[\hat{\mathbf{J}}_n(\widetilde{\boldsymbol{\theta}})\widetilde{V}(\widetilde{\boldsymbol{\theta}})\right], \\
\widetilde{BMT} &= \widetilde{J}_1 + \widetilde{J}_0 = \mathbf{tr}\left\{C_E\left[\mathbf{y}, (\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_E = 0)\right]\left[\widetilde{V}_E(\widetilde{\boldsymbol{\theta}}_L)\right]\right\} + \sqrt{n}\left(\widetilde{BIMT}/q - 1\right)^2,
\end{aligned}$$

where $C_E\left[\mathbf{y}, (\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_E = 0)\right]$ is the submatrix of $C\left[\mathbf{y}, (\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}_E = 0)\right]$ and $\widetilde{V}_E(\widetilde{\boldsymbol{\theta}}_L)$ is the submatrix of $\widetilde{V}(\widetilde{\boldsymbol{\theta}}_L)$.

The number of MCMC draws (M) should be chosen so that $\widetilde{BIMT}$ and $\widetilde{BMT}$ enjoy the same asymptotic properties of BIMT and BMT, respectively, when n is allowed to go to infinity. To derive the correct orders for M, we need to add Assumption 13 below. The same assumption was also used in Cheng, et al (2017), Robert and Casella (2004) and Jones (2004).

Assumption 13: Assume $\left\{\boldsymbol{\theta}_n^{(m)}\right\}_{m=1}^{M}$ and $\left\{\boldsymbol{\theta}_{Ln}^{(m)}\right\}_{m=1}^{M}$ are two Markov chains which are aperiodic, ψ-irreducible, positive Harris recurrent and geometrically ergodic with the stationary distribution being $p(\boldsymbol{\theta}|\mathbf{y})$ and $p(\boldsymbol{\theta}_L|\mathbf{y})$, and $\max_{n\geq 1} E\left[\left|\boldsymbol{\theta}_n^{(1)}\right|^{4+\varepsilon_0}|\mathbf{y}\right] < \infty$, $\max_{n\geq 1} E\left[\left|\boldsymbol{\theta}_{Ln}^{(1)}\right|^{4+\varepsilon_0}|\mathbf{y}\right] < \infty$, for some $\varepsilon_0 > 0$.

For $a = 1, 2, \cdots, q$, let $\boldsymbol{\theta}_a$ be the a^{th} component of $\boldsymbol{\theta}$, and $\sigma_{1n,a}^2$ be the long run variance of Markov chain, $\left\{\boldsymbol{\theta}_a^{(m)}\right\}_{m=1}^{M}$. That is,

$$\sigma_{1n,a}^2 = Var\left(\boldsymbol{\theta}_a|\mathbf{y}\right) + 2\sum\nolimits_{k=1}^{\infty}\gamma_{1n,a}\left(k|\mathbf{y}\right),$$

where $\gamma_{n,a}\left(k|\mathbf{y}\right)$ is the k^{th} order autocovariance given by

$$\gamma_{1n,a}\left(k|\mathbf{y}\right) = Cov\left(\boldsymbol{\theta}_a^{(1)}, \boldsymbol{\theta}_a^{(1+k)}|\mathbf{y}\right) = E\left(\boldsymbol{\theta}_a^{(1)}\boldsymbol{\theta}_a^{(1+k)}|\mathbf{y}\right) - E\left(\boldsymbol{\theta}_a^{(1)}|\mathbf{y}\right)E\left(\boldsymbol{\theta}_a^{(1+k)}|\mathbf{y}\right).$$

Similarly, if we let $\boldsymbol{\vartheta} = vech\left[\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\right)'\right]$ and $\boldsymbol{\vartheta}_L = vech\left[\left(\boldsymbol{\theta}_L - \bar{\boldsymbol{\theta}}_L\right)\left(\boldsymbol{\theta}_L - \bar{\boldsymbol{\theta}}_L\right)'\right]$ where $\bar{\boldsymbol{\theta}} = E\left[\boldsymbol{\theta}|\mathbf{y}\right]$ and $\bar{\boldsymbol{\theta}}_L = E\left[\boldsymbol{\theta}_L|\mathbf{y}\right]$, then we can define $\sigma_{2n,b}^2$ to be the long run variance of Markov chain, $\left\{\boldsymbol{\vartheta}_b^{(m)}\right\}_{m=1}^M$ for $b = 1, 2, \cdots, r$ $(= q(q+1)/2)$. Moreover, for $b = 1, 2, \cdots, r_L$ $(= q_L(q_L+1)/2)$ with $q_L = q + q_E$, we can define $\sigma_{Ln,b}^2$ to be the long run variance of Markov chain, $\left\{\boldsymbol{\vartheta}_{Ln,b}^{(m)}\right\}_{m=1}^M$.

Theorem 2.2 *Let*

$$\sigma_{1n}^{2*} = \max_{a\in\{1,2,\ldots,q\}} \sigma_{1n,a}^2, \sigma_{2n}^{2*} = \max_{b\in\{1,2,\ldots,r\}} \sigma_{2n,b}^2, \sigma_{Ln}^{2*} = \max_{b\in\{1,2,\ldots,r_L\}} \sigma_{Ln,b}^2.$$

Let M_{BIMT} be the number of MCMC draws from $p(\boldsymbol{\theta}|\mathbf{y})$ for $\widetilde{BIMT}$. Let M_{BMT} and M_L be the number of MCMC draws from $p(\boldsymbol{\theta}|\mathbf{y})$ and $p(\boldsymbol{\theta}_L|\mathbf{y})$ for $\widetilde{BMT}$. Under Assumptions 1-13, for any $c_i^* > 0$ with $i = 1, 2$, if we choose

$$M_{BIMT} = \max\left\{n^{1+c_1^*}\sigma_{1n}^{2*}, n^{3+c_2^*}\sigma_{2n}^{2*}\right\}, \tag{8}$$

then, when the model is correctly specified, we have

$$\sqrt{n}\left(\widetilde{BIMT} - BIMT\right) = o_p(1).$$

Furthermore, for any $c_i^* > 0$ with $i = 3, 4, 5$, if we choose

$$M_{BMT} = \max\left\{n^{1+c_3^*}\sigma_{1n}^{2*}, n^{2.5+c_4^*}\sigma_{2n}^{2*}\right\},\ M_L = n^{2+c_5^*}\sigma_{Ln}^{2*}, \tag{9}$$

then, when the model is correctly specified, we have

$$\widetilde{J}_1 = \widetilde{J}_1 + o_p(1), \widetilde{J}_0 = o_p(1),\ \widetilde{BMT} = BMT + o_p(1).$$

When the model is misspecified such that $q^* \neq q$, we have

$$\widetilde{J}_0 = J_0 + o_p(1) = \sqrt{n}\left[q^*/q - 1\right]^2 + 2\sqrt{n}\left(q^*/q - 1\right)o_p(1) + O_p(n^{-1/2}) = O_p(\sqrt{n}).$$

Remark 2.8 *Theorem 2.2 gives the order for the number of MCMC draws in (8) to ensure that $\widetilde{BIMT}$ has the same asymptotic distribution as BIMT and that in (9) to ensure that $\widetilde{BMT}$ has the same asymptotic distribution as BMT. In addition, it gives the condition under which $\widetilde{BMT}$ has the same order of power as BMT.*

Remark 2.9 *In practice, the sample size n is often large enough so that $M_{BIMT} = \max\left\{n^{1+c_1^*}\sigma_{1n}^{2*}, n^{3+c_2^*}\sigma_{2n}^{2*}\right\} = n^{3+c_2^*}\sigma_{2n}^{2*}$ and $M_{BMT} = \max\left\{n^{1+c_3^*}\sigma_{1n}^{2*}, n^{2.5+c_4^*}\sigma_{2n}^{2*}\right\} = n^{2.5+c_2^*}\sigma_{2n}^{2*}$. In this case, M_{BMT} is of a smaller order than M_{BIMT} and the difference in order is $\sqrt{n}$. When the number of MCMC draws is set at $M_{BIMT} = n^{3+c_2^*}\sigma_{2n}^{2*}$, Theorem 2.2 suggests that $\widetilde{BIMT}$−BIMT$= o_p(n^{-1/2})$. According to Proposition 2.1, BIMT$= q + O_p\left(n^{-1/2}\right)$ under H_0. These two properties imply that both $\widetilde{BIMT}$ and BIMT converge to the same distribution. However, if we only choose $M_{BIMT} = n^{2.5+c_2^*}\sigma_{2n}^{2*}$, then $\widetilde{BIMT}$−BIMT$= o_p(n^{-1/4})$,* *suggesting that* $\widetilde{BIMT}$ *and BIMT may not converge to the same distribution. When the number of MCMC draws is set at* $M_{BMT} = n^{2.5+c_3^*}\sigma_{2n}^{2*}$ *in the original model and at* $M_L = n^{2+c_5^*}\sigma_{Ln}^{2*}$ *in the expanded model, Theorem 2.2 suggests that* $\widetilde{BMT} - BMT = o_p(1)$*. According to Theorem 2.1,* $BMT \xrightarrow{d} \chi^2(q_E)$ *under* H_0*. These two properties imply that* $\widetilde{BMT}$ *and BMT converge to the same distribution. Hence, for* $\widetilde{BIMT}$ *to have the same asymptotic distribution as BIMT, a stronger order condition is needed for* M *than that for* $\widetilde{BMT}$ *to have the same asymptotic distribution as BMT. The orders differ by* $\sqrt{n}$*. This is additional advantage in using BMT over BIMT. For example, if* $n = 2000$*,* $\sqrt{n} \approx 45$*. It means the number of MCMC draws required for* $\widetilde{BIMT}$ *is about 45 times as large as that for* $\widetilde{BMT}$*.*

Remark 2.10 *In practice,* σ_{1n}^{2*}*,* σ_{2n}^{2*} *and* σ_{Ln}^{2*} *are unknown. Hence, one has to estimate them from MCMC output. For example, we can estimate them sequentially by consistent batch means or spectral methods.*[2] *Once the order achieves the desirable one, we may stop MCMC drawing. Let consistent estimates of* σ_{1n}^{2*}*,* σ_{2n}^{2*} *and* σ_{Ln}^{2*} *be* $\hat{\sigma}_{1n}^{2*}$*,* $\hat{\sigma}_{2n}^{2*}$ *and* $\hat{\sigma}_{Ln}^{2*}$*. Suppose BMT is used. For the null model, we should choose* $M \geq n^{2.5+c_4^*}\hat{\sigma}_{2n}^{2*}$*. For the expanded model, we should choose* $M \geq n^{2+c_5^*}\hat{\sigma}_{Ln}^{2*}$*. Since* $\{c_i^*\}_{i=4}^5$ *are any positive constants, the lower bound of* $n^{c_i^*}$ *is one. In practice, we may set* $n^{c_i^*}$ *to be a number slightly larger than 1.*

3 Simulation and Empirical Studies

In this section, we first design two simulation studies to check the finite sample performance of BMT. 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. We also investigate the performance of BIMT in this model. 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 existing tests are difficult to use. This third model is a stochastic volatility model where existing tests are impossible to use.

[2] The conditions under which these estimators are strongly consistent are established in Flegal and Jones (2010) and Jones, et al (2006).

3.1 Simulation Studies

3.1.1 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 some 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 a $U[-3,3]$ 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,

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

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

For the expanded model, we use

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

Hence, $\boldsymbol{\theta}_E = \beta_3$.

To implement the proposed test, we need to use the MCMC method to estimate the model under the null hypothesis and the expanded model. To check the robustness of priors, we consider two sets of non-informative prior specifications. The first prior is proper but very vague and given by

$$\boldsymbol{\beta} = (\beta_0, \beta_1, ..\beta_3)' \sim N[\boldsymbol{\mu}_\beta = 0, \sigma^2 \boldsymbol{V}_\beta = 100 \times \mathbf{I}_4], \sigma^{-2} \sim Gamma(a = 0.01, b = 0.01),$$

where $\mathbf{I}_4$ is the identity matrix with dimension 4, (a, b) are hyperparameters of the Gamma distribution. Both the normal prior distribution and the Gamma prior distribution have large spread so that they are non-informative. The second is an improper flat prior, i.e., $p(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}$. In this example, since the posterior distribution is available analytically, we simply make 2,000 draws from the posterior directly.

We first check the size distortion problem in IOS_A and BIMT when the flat prior is used. Table 1 reports the empirical size of IOS_A and BIMT based on the asymptotic distribution and the parametric bootstrap distribution. In this example, the ML method is trivial to

Table 1: Empirical size for IOS_A and BIMT under the asymptotic distribution and the bootstrap distribution

	IOS_A		BIMT	
n	Asymptotic	Bootstrap	Asymptotic	Bootstrap
50	0.216	0.049	0.5420	0.0570
100	0.147	0.050	0.3270	0.0525
200	0.136	0.056	0.2155	0.0570

Table 2: Empirical size for alternative tests

n	IR	IMT	IOS	BMT_v	BMT_f
50	0.044	0.050	0.060	0.051	0.046
100	0.045	0.059	0.056	0.055	0.050
200	0.046	0.065	0.048	0.050	0.052

implement and hence the bootstrap method is feasible. The method used to obtain the asymptotic variance was proposed by Lancaster (1984). It can be seen clearly that the oversized problem for both IOS_A and BIMT is severe when the asymptotic distribution is used. The size distortion is even larger for BIMT than for IOS_A, especially when n is small. For both tests, the bootstrap method can solve the size distortion problem. These results reinforce the theory developed earlier in the paper.

Let BMT_v be BMT under the vague prior and BMT_f be BMT under the flat prior. Table 2 reports the empirical size of IR, IMT, IOS, BMT_v and BMT_f under H_0 and at the 5% significance level. The results of the first three tests are extracted from Zhou et al (2012) where critical values are obtained from the bootstrap distribution. The BMT test entertains similar performance to other tests and shows small size distortions in all cases. Moreover, the size of BMT is robust against the change in prior. Table 3 reports the empirical power of IR, IMT, 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 IMT can be quite low (when n =50). The power of BMT is compatible with that of IR. Again, the prior does not have significant influence on the power of BMT.

From this experiment we can conclude that the finite sample performance of BMT is

Table 3: Empirical power under the alternative hypothesis

n	IR	IMT	IOS	BMT_v	BMT_f
50	0.85	0.11	0.9837	0.797	0.750
100	0.95	0.46	1.000	0.976	0.961
200	1.00	0.93	1.000	1.000	1.000

satisfactory with small size distortion and good power. Both the size and the power of BMT are not sensitive to priors. We should emphasize that critical values of BMT are obtained from χ^2 and hence no bootstrap method is needed.

3.1.2 A linear state-space model

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

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

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 IMT, IOS$_A$, IR and BIMT. Consequently, we only implement BMT in this example.

The expanded model is

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

where an intercept is added to the observation equation. If Model (10) is correctly specified, $\alpha = 0$ in the expanded model.

For the MCMC analysis, we use the following vague priors for 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 burn-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 empirical size, we set 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 an i.i.d. normal distribution with mean 0 and variance 0.001. To compute empirical power, we consider two different DGPs. The first DGP (denoted by $M1$) is given by

$$\begin{aligned} R_t &= \beta_t R_{0t} + \frac{\sigma_\varepsilon}{\sqrt{3}} \varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} t_3, \\ \beta_{t+1} &= \bar{\beta} + \phi\left(\beta_t - \bar{\beta}\right) + \eta_t, \eta_t \overset{i.i.d.}{\sim} N\left(0, \sigma_\eta^2\right), \end{aligned} \tag{12}$$

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

$$R_t = \alpha + \beta_t R_{0t} + \frac{\sigma_\varepsilon}{\sqrt{3}} \varepsilon_t, \varepsilon_t \overset{i.i.d.}{\sim} t_3, \tag{13}$$

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

where $\alpha = 0.002$, $\sigma_\varepsilon^2 = 0.000307$, $\bar{\beta} = 0.96$, $\phi = 0.5$, $\sigma_\eta^2 = 0.208$ 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 on the possible source of misspecification, we also report the proportion of the 2,000 replications where J_1 rejects $\alpha = 0$ in the expanded model (11).

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. 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. 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 (10).

3.2 Empirical studies

3.2.1 A linear regression model

In the first empirical study, we test the specification 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.

As to the data, the sample size 2,725. For more details, one can refer to Wooldridge (2014).

The null model is the following linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i, \varepsilon_i \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right). \tag{14}$$

For the expanded model, we use

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{1i}^2 + \varepsilon_i, \varepsilon_i \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right). \tag{15}$$

If Model (14) is correctly specified, $\beta_5 = 0$ in Model (15). Conjugated vague priors for β ($:= \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \end{pmatrix}'$) 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 very vague priors where hyper-parameters in the priors are set at

$$\mu_\beta = 0,\ V_\beta = 100 \times \mathbf{I}_6,\ a = 0.01,\ b = 0.01.$$

For the 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	0.7067	0.0332	0.6415	0.7717	0.6317	0.0350	0.5634	0.7009
β_1	-0.1506	0.0409	-0.2306	-0.0712	0.7897	0.1556	0.4869	1.0942
β_2	0.0074	0.0047	-0.0019	0.0167	0.0040	0.0048	-0.0053	0.0134
β_3	-0.0374	0.0088	-0.0546	-0.0202	-0.0439	0.0088	-0.0611	-0.0267
β_4	0.1033	0.0104	-0.1236	-0.0828	-0.0933	0.0105	-0.1141	-0.0729
σ^2	0.7069	0.0193	0.6700	0.7461	0.6970	0.0189	0.6611	0.7347
β_5	-	-	-	-	-0.9855	0.1576	-1.2981	-0.6776

The critical value of $\chi^2(1)$ is 6.63 at the 1% significance level. In this study, the BMT statistic is 346.6568, suggesting that Model (15) is misspecified. It is easy to find out that J_1 is 38.6919 (i.e., J_0=307.9649) 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 (15), suggesting that the misspecification of Model (15) comes from the wrong functional form in x_{1i}.

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 5% bootstrap critical value is 46.0734. Hence, IMT also suggest that Model (14) is misspecified, reinforcing the result from BMT. However, IMT does not tell how to improve the model.

In this example, if we use the consistent batch means of Jones (2006) to estimate the long run variances by setting the number of batches at $\sqrt{M}$, then we have $\hat{\sigma}_{1n}^{2*} = 1.51 \times 10^{-3}$, $\hat{\sigma}_{2n}^{2*} = 5.55 \times 10^{-6}$, $\hat{\sigma}_{Ln}^{2*} = 1.10 \times 10^{-3}$. According to Remark 3.14, with $n^v = 1$, the lower bound for the number of MCMC draws is $M_{BMT} = 2153$ for the null model and $M_L = 8168$ for the expanded model. Hence, our choice of 20,000 MCMC draws for both models is large enough to ensure the validity of the asymptotic theory for $\widetilde{BMT}$.

3.2.2 A linear state-space model

The capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) is a fundamental theory in finance. When investors can borrow and lend at a risk-free rate, the intercept is expected to be zero in the CAPM. Another important feature of the CAPM is that beta is constant over time. However, it is well-documented that the systematic risk of an asset depends on microeconomic factors as well as macroeconomic factors. Hence, allowing time-varying beta is an important way to generalize the CAPM.

In this section, we extend the traditional CAPM by allowing for time-varying beta in a state-space form. Following Mergner and Bulla (2008), a CAPM without intercept but with time-varying beta is given by

$$\begin{aligned} R_{it} &= \beta_{it} R_{0t} + \varepsilon_{it}, \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_{i\varepsilon}^2\right), \\ \beta_{it+1} &= \bar{\beta}_i + \phi\left(\beta_{it} - \bar{\beta}_i\right) + \eta_{it}, \eta_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_{i\eta}^2\right), \end{aligned} \tag{16}$$

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. This asset pricing model is used to show that the investor cannot obtain extra return from investing in the insurance industry.

In this example, we would like to test if the CAPM without intercept and with time-varying beta can describe a dataset. Naturally, the following CAPM with intercept and time-varying beta can be chosen as the expanded model,

$$\begin{aligned} R_{it} &= \alpha_i + \beta_{it} R_{0t} + \varepsilon_{it}, \varepsilon_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_{i\varepsilon}^2\right), \\ \beta_{it+1} &= \bar{\beta}_i + \phi_i\left(\beta_{it} - \bar{\beta}_i\right) + \eta_{it}, \eta_{it} \overset{i.i.d.}{\sim} N\left(0, \sigma_{i\eta}^2\right), \end{aligned} \tag{17}$$

where an intercept is added to the mean equation. If Model (16) is correctly specified, $\alpha_i = 0$ in Model (17).

For the MCMC analysis, we use the following non-informative priors for hyper-parameters

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

We draw 500,000 MCMC samples after 50,000 burn-in observations from the posterior distribution for the null model, and 150,000 MCMC samples after 20,000 burn-in for the expanded model to compute BMT. 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 $\sigma_{i\varepsilon}^2$ are multiplied by 10,000). We do not implement other tests as bootstrap methods are computationally too expensive in this setup.

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
$\sigma_{i\varepsilon}^2$	1.3616	0.0756	1.2200	1.5168	1.3603	0.0758	1.2118	1.5158
$\bar{\beta}_i$	1.2161	0.0270	1.1630	1.2680	1.2186	0.0270	1.1650	1.2710
ϕ_i	0.4233	0.0984	0.2241	0.6088	0.4210	0.0950	0.2191	0.6101
$\sigma_{i\eta}^2$	0.1621	0.0266	0.1107	0.2146	0.1627	0.0266	0.1101	0.2158
α_i	-	-	-	-	-3.9226	3.5507	-10.8900	3.0250

BMT is 146.9662, suggesting that Model (17) is misspecified. It is easy to find out that J_1 is 1.2179 (i.e., J_0=145.7483) which is less than the critical value of $\chi^2(1)$. Interestingly, using J_1 alone suggests that we cannot reject $\alpha_i = 0$ in Model (17). According BMT, the CAPM without intercept but with time-varying beta is rejected. Hence, a more appropriate CAPM specification is needed.

The batch means estimates of the long run variances are $\hat{\sigma}_{1n}^{2*} = 0.59$, $\hat{\sigma}_{2n}^{2*} = 5.51 \times 10^{-3}$ and $\hat{\sigma}_{Ln}^{2*} = 5.42 \times 10^{-3}$. Hence, the lower bound for the number of MCMC draws is $M_{BMT} = 460,400$ for the null model and $M_L = 11,792$ for the expanded model. In this example, we have used $M = 500,000$ and $M_L = 150,000$ which are large enough.

3.2.3 A stochastic volatility (SV) model

The dataset used here contains the daily returns on AUD/USD exchange rates from January 2005 to December 2012. The sample size is 2086. We first test the i.i.d. normal model with constant mean and constant variance given by

$$y_t = \alpha + \varepsilon_t,\ \varepsilon_t \overset{i.i.d.}{\sim} N\left(0,\sigma^2\right). \tag{18}$$

We first expand the conditional mean to the following AR(1) model

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t,\ \varepsilon_t \overset{i.i.d.}{\sim} N\left(0,\sigma^2\right). \tag{19}$$

The MCMC method is implemented to estimate the parameters with the following non-informative priors

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

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

BMT is 251.52, rejecting the i.i.d. normal model at the 1% level. This conclusion is not surprising as the volatility of stock returns is time-varying. 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 (19). This conclusion is also not surprising as the daily returns have very weak serial correlations. A large BMT value combined with a small J_1 value suggests that the conditional variance is incorrectly specified even when the conditional mean was expanded.

That is why in the next study we change the null model to the following basic SV model which differs from the i.i.d. normal model in the conditional variance specification,

$$\begin{aligned} y_t &= \alpha + \exp(h_t/2)\, u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1), \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \tau\nu_t,\ \ \nu_t \overset{i.i.d.}{\sim} N(0,1). \end{aligned} \tag{20}$$

The expanded model is as follows,

$$\begin{aligned} y_t &= \alpha + \beta_1 y_{t-1} + \exp(h_t/2)\, u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1). \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \tau\nu_t,\ \ \nu_t \overset{i.i.d.}{\sim} N(0,1). \end{aligned} \tag{21}$$

The following non-informative priors are used

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

To obtain BMT, for the null model we draw 30,000,000 MCMC samples from the posterior distribution and discard the first 1,000,000 as burn-in observations, and the remaining samples are stored as effective observations. For the expanded model, we draw

14,000,000 MCMC samples and discard the first 1,000,000 as burn-in observations.[3] The batch means estimates of the long run variances are $\hat{\sigma}_{1n}^{2*} = 0.2057$, $\hat{\sigma}_{2n}^{2*} = 0.1137$ and $\hat{\sigma}_{Ln}^{2*} = 0.0958$. The low bound for the number of MCMC draws is $M_{BMT} = 22,570,000$ and $M_L = 416,460$. In this example, we have used $M = 29,000,000$ and $M_L = 13,000,000$ which are large enough. Based on the MCMC draws, the posterior mean, standard deviation, 2.5% quantile, and 97.5% quantile of all the parameters are reported in Table 8.

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	-0.0005	0.0126	-0.0252	0.0242	-0.0004	0.0125	-0.0249	-0.0242
μ	-1.0174	0.1761	-1.3666	-0.6769	-1.0173	0.1761	-1.3666	-0.6765
ϕ	0.9761	0.0072	0.9603	0.9887	0.9761	0.0072	0.9603	0.9887
τ^2	0.0288	0.0067	0.0182	0.0444	0.0288	0.0067	0.0182	0.0444
β_1	-	-	-	-	0.045	0.0227	0.0007	0.899

All the first derivatives required by BMT are calculated based on particle filters.[4] The number of particles in each period is 1000. For the null model, the standard errors for the first order derivative with respect to α, μ, ϕ and τ^2 are 0.0865, 0.0151, 0.9717, 0.6512. For the expanded model, the standard errors for the first order derivative with respect to α, μ , ϕ, τ^2 and β are 0.0900 , 0.0143, 0.9048, 0.6009 and 0.0303. BMT=3.2714 which is less than 3.84, the critical value of $\chi^2(1)$ under 5% significant level, suggesting that the basic SV model is not misspecified at the 5% significant level.

4 Conclusions

In this paper, we have proposed two new specification test statistics based on MCMC output to check the validity of a model specification. The first one is the MCMC version of IOS_A test. We show that it is asymptotically normally distributed under the null hypothesis but has a complex asymptotic variance. While it does not require the alternative model be specified, a bootstrap method is needed to avoid calculating asymptotic variance. The second test, which is our main 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 first test. 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

[3]In this example, we have written C code to conduct MCMC analysis of the SV models. It takes about 2 hours to draw 30,000,000 MCMC draws using a common desktop PC with Intel(R) Core(TM) i7-7700k CPU @ 4.20GHz.

[4]The calculation details are given in Appendix 5. The approach of Chan and Lai (2013) is used to compute the standard errors of the first derivatives based on particle filters.

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 on the source of misspecification.

An important feature of the proposed tests is that they are based on 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 tests 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.

There is no reason why our proposed tests cannot be used in connection to other simulation-based methods. One example of simulation-based methods is the sequential Monte Carlo method of Chopin (2002). Moreover, 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.

5 Appendix

5.1 Appendix 1: Proof of Proposition 2.1

By using the first-order expansion, we can show that

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

where $\tilde{\boldsymbol{\theta}}_1$ lies between $\bar{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}$. Furthermore, note that

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

$$+\frac{1}{n}\sum_{t=1}^{n}\left[h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_1\right)\right] vech\left[\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}\right)\left(\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}'\right)'\right].$$

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 $\bar{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}=O_p\left(n^{-1}\right)$ from Remark 2.1. Hence, we can show that

$$\begin{aligned}\hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) &= \hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}\right)+O_p(1)O_p(n^{-1})+O_p(1)O_p(n^{-1})O_p(n^{-1})\\ &= \hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}\right)+O_p(n^{-1}).\end{aligned} \tag{22}$$

From Li, Yu and Zeng (2017), under Assumptions 1-12, we have $\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}_{ML}=O_p(n^{-1})$. Similar to (22), it can be shown that

$$\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}\right)=\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)+O_p(n^{-1}).$$

Based on Assumptions 6 and 10, we can get that $\hat{\mathbf{H}}_n(\hat{\boldsymbol{\theta}})=O_p(1)$. According to Remark 2.1, it is easy to show that

$$\begin{aligned}& V(\bar{\boldsymbol{\theta}})=E\left[\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}}\right)'|\mathbf{y}\right]\\ =\ & E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)'|\mathbf{y}\right]\\ =\ & E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)'|\mathbf{y}\right]+2E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)|\mathbf{y}\right]\left(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)'+\left(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)'\\ =\ & E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)'|\mathbf{y}\right]-\left(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}\right)'\\ =\ & -L_n^{-(2)}(\hat{\boldsymbol{\theta}})+O_p(n^{-2})\\ =\ & -\left[n\hat{\mathbf{H}}_n(\hat{\boldsymbol{\theta}})+\frac{\partial^2\log p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^{-1}+O_p(n^{-2})\\ =\ & -\frac{1}{n}\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})\left[\mathbf{I}_q+\frac{1}{n}\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})\frac{\partial^2\log p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^{-1}+O_p(n^{-2})\\ =\ & -\frac{1}{n}\hat{\mathbf{H}}_n(\hat{\boldsymbol{\theta}})\left[\mathbf{I}_q+O_p(n^{-1})\right]^{-1}+O_p(n^{-2})\\ =\ & -\frac{1}{n}\hat{\mathbf{H}}_n(\hat{\boldsymbol{\theta}})+O_p(n^{-2}),\end{aligned}$$

where $\mathbf{I}_q$ is q-dimensional identity matrix. Hence, can get that

$$V(\bar{\boldsymbol{\theta}})=E\left[(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}-\bar{\boldsymbol{\theta}})'|\mathbf{y}\right]=-\frac{1}{n}\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}})+O_p(n^{-2})=O_p(n^{-1}) \tag{23}$$

In addition, by using the Taylor expansion, similar to (22), we can further get that

$$\hat{\mathbf{H}}_n\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{n}\sum_{t=1}^{n}h_t\left(\hat{\boldsymbol{\theta}}\right)=\frac{1}{n}\sum_{t=1}^{n}h_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)+\frac{1}{n}\sum_{t=1}^{n}\triangledown l^{(3)}\left(\tilde{\boldsymbol{\theta}}_2\right)\left[I_q\otimes\left(\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}_{ML}\right)\right]$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n} h_t\left(\hat{\boldsymbol{\theta}}_{ML}\right) + O_p\left(n^{-1}\right) \\
&= \hat{\mathbf{H}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) + O_p\left(n^{-1}\right), \qquad (24)
\end{aligned}
$$

where $\bigtriangledown l^{(3)}\left(\tilde{\boldsymbol{\theta}}_2\right)$ is the third order derivative of $l_t\left(\boldsymbol{\theta}\right)$ evaluated at $\tilde{\boldsymbol{\theta}}_2$, and $\tilde{\boldsymbol{\theta}}_2$ lies between $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}_{ML}$.

From the definition of BIMT and (22)-(24), we get

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

Hence, the first part of Proposition 2.1 is proved.

Next, when the model is correctly specified, we derive the order of BIMT$-q$. According to White (1987), under H_0, it can be shown that, in White's IMT test, the elements of $\sqrt{n}\left[\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) + \hat{\mathbf{H}}_{\mathbf{n}}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\right]$ converge to the normal distribution so that

$$
\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) + \hat{\mathbf{H}}_{\mathbf{n}}\left(\hat{\boldsymbol{\theta}}_{ML}\right) = O_p(n^{-1/2}). \qquad (25)
$$

Based on (22)-(25), we can further show that

$$
\begin{aligned}
&-\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right) = -\hat{\mathbf{H}}_n\left(\hat{\boldsymbol{\theta}}\right) + O_p\left(n^{-1}\right) = -\mathbf{H}\left(\hat{\boldsymbol{\theta}}_{ML}\right) + O_p\left(n^{-1}\right) \\
&= \hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) + O_p\left(n^{-1/2}\right) = \hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}\right) + O_p\left(n^{-1/2}\right) = \hat{\mathbf{J}}_n\left(\bar{\boldsymbol{\theta}}\right) + O_p\left(n^{-1/2}\right). \qquad (26)
\end{aligned}
$$

From Li, Yu and Zeng (2017), under Assumptions 1-12, by the Laplace expansion,

$$
\mathbf{tr}\left(-n\hat{\mathbf{H}}_n\left(\bar{\boldsymbol{\theta}}\right)V\left(\bar{\boldsymbol{\theta}}\right)\right) = q + O_p\left(n^{-1}\right). \qquad (27)
$$

From (23), (26) and (27), we have

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

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

Hence, Proposition 2.1 is proved.

5.2 Appendix 2: Proof of Theorem 2.1

When the model is correctly specified, by Proposition 2.1, we can show that

$$
\begin{aligned}
J_0 &= \sqrt{n}(\mathrm{BIMT} - q)^2 = \sqrt{n}(\mathrm{BIMT} - q)(\mathrm{BIMT} - q) \\
&= \sqrt{n}O_p(n^{-1/2})O_p(n^{-1/2}) = O_p(n^{-1/2}) = o_p(1).
\end{aligned}
$$

Furthermore, according to Li, et al (2015), if $\boldsymbol{\theta}_E = 0$ in the expanded model, as $n \to \infty$, when the model is correctly specified, we have

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

Hence, we get

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

In the following, we derive the power of BMT. Similarly to the proof of Proposition 2.1, by using the Taylor expansion, we get

$$
\begin{aligned}
\hat{\mathbf{H}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) &= \frac{1}{n}\sum_{t=1}^{n} h_t\left(\hat{\boldsymbol{\theta}}_{ML}\right) = \frac{1}{n}\sum_{t=1}^{n} h_t\left(\boldsymbol{\theta}_0\right) + \frac{1}{n}\sum_{t=1}^{n} \triangledown l^{(3)}\left(\tilde{\boldsymbol{\theta}}_3\right)\left[I_q \otimes \left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right)\right] \\
&= \frac{1}{n}\sum_{t=1}^{n} h_t\left(\boldsymbol{\theta}_0\right) + O_p\left(n^{-1/2}\right) \\
&= \hat{\mathbf{H}}_n\left(\boldsymbol{\theta}_0\right) + O_p\left(n^{-1/2}\right), \qquad (28)
\end{aligned}
$$

where $\triangledown l^{(3)}\left(\tilde{\boldsymbol{\theta}}_3\right)$ is the third order derivative of $l_t\left(\boldsymbol{\theta}\right)$ evaluated at $\tilde{\boldsymbol{\theta}}_3$ and $\tilde{\boldsymbol{\theta}}_3$ lies between $\hat{\boldsymbol{\theta}}_{ML}$ and $\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0 = O_p\left(n^{-1/2}\right)$ by the quasi-ML theory given in Gallant and White (1988) and White (1982, 1987).

Furthermore, we can similarly get

$$
\begin{aligned}
\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right) &= \frac{1}{n}\sum_{t=1}^{n} s_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)s_t\left(\hat{\boldsymbol{\theta}}_{ML}\right)' \\
&= \frac{1}{n}\sum_{t=1}^{n}\left[s_t\left(\boldsymbol{\theta}_0\right) + h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right)\right]\left[s_t\left(\boldsymbol{\theta}_0\right) + h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right)\right]' \\
&= \frac{1}{n}\sum_{t=1}^{n} s_t\left(\boldsymbol{\theta}_0\right)s_t\left(\boldsymbol{\theta}_0\right)' + \frac{2}{n}\sum_{t=1}^{n} h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\right)s_t\left(\boldsymbol{\theta}_0\right)'
\end{aligned}
$$

$$+\frac{1}{n}\sum_{t=1}^{n} h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right)\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right)' h_t\left(\tilde{\boldsymbol{\theta}}_4\right)',$$

where $\tilde{\boldsymbol{\theta}}_4$ lies between $\hat{\boldsymbol{\theta}}_{ML}$ and $\boldsymbol{\theta}_0$. It can be rewritten as a vector form, that is,

$$\begin{aligned} vech\left(\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)\right) &= \frac{1}{n}\sum_{t=1}^{n} vech\left(s_t\left(\boldsymbol{\theta}_0\right)s_t\left(\boldsymbol{\theta}_0\right)'\right)+\frac{2}{n}\sum_{t=1}^{n}\left[s_t\left(\boldsymbol{\theta}_0\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\right] vech\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right) \\ &\quad +\frac{1}{n}\sum_{t=1}^{n}\left[h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\otimes h_t\left(\tilde{\boldsymbol{\theta}}_4\right)\right] vech\left(\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right)\left(\hat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}_0\right)'\right). \end{aligned}$$

Hence, similar to (28), we have

$$\hat{\mathbf{J}}_n\left(\hat{\boldsymbol{\theta}}_{ML}\right)=\hat{\mathbf{J}}_n\left(\boldsymbol{\theta}_0\right)+O_p(1)O_p\left(n^{-1/2}\right)+O_p(1)O_p\left(n^{-1}\right)=\hat{\mathbf{J}}_n\left(\boldsymbol{\theta}_0\right)+O_p\left(n^{-1/2}\right). \quad (29)$$

Based (28) and (29), by Assumption 7 and the central limit theorem, it follows that

$$\begin{aligned} &\mathrm{IOS}_A=\mathbf{tr}\left\{-\hat{\mathbf{J}}_n(\hat{\boldsymbol{\theta}}_{ML})\hat{\mathbf{H}}_n^{-1}(\hat{\boldsymbol{\theta}}_{ML})\right\} \\ =\;& \mathbf{tr}\left\{-\left[\hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)+O_p(n^{-1/2})\right]\left[\hat{\mathbf{H}}_{\mathbf{n}}^{-1}(\boldsymbol{\theta}_0)+O_p(n^{-1/2})\right]\right\} \\ =\;& \mathbf{tr}\left[-\hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)\hat{\mathbf{H}}_n^{-1}(\boldsymbol{\theta}_0)\right]-\mathbf{tr}\left[\hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)O_p(n^{-1/2})\right]-\mathbf{tr}\left[\hat{\mathbf{H}}_n^{-1}(\boldsymbol{\theta}_0)O_p(n^{-1/2})\right]+O_p(n^{-1}) \\ =\;& \mathbf{tr}\left[-\hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)\hat{\mathbf{H}}_n^{-1}(\boldsymbol{\theta}_0)\right]-O_p(1)O_p(n^{-1/2})-O_p(1)O_p(n^{-1/2})+O_p(n^{-1}) \\ =\;& \mathbf{tr}\left[-\hat{\mathbf{J}}_n(\boldsymbol{\theta}_0)\hat{\mathbf{H}}_n^{-1}(\boldsymbol{\theta}_0)\right]+O_p(n^{-1/2}) \\ =\;& \mathbf{tr}\left[-\left(\mathbf{J}(\boldsymbol{\theta}_0)+o_p(1)\right)\left(\mathbf{H}^{-1}(\boldsymbol{\theta}_0)+o_p(1)\right)\right]+O_p(n^{-1/2}) \\ =\;& \mathbf{tr}\left[-\mathbf{J}(\boldsymbol{\theta}_0)\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right]+o_p(1)+O_p(n^{-1/2}) \\ =\;& q^*+o_p(1)=O_p(1). \end{aligned}$$

By Proposition 3.1, whether the model is misspecified or not, we get

$$\mathrm{BIMT}=\mathrm{IOS}_A+O_p(n^{-1}).$$

Hence, we have

$$\begin{aligned} J_0 &= \sqrt{n}(\mathrm{BIMT}/q-1)^2=\sqrt{n}\left[\left(\mathrm{IOS}_A+O_p(n^{-1})\right)/q-1\right]^2 \\ &= \sqrt{n}\left[\mathrm{IOS}_A/q-1+O_p(n^{-1})\right]^2 \\ &= \sqrt{n}\left[\mathrm{IOS}_A/q-1\right]^2+\sqrt{n}\left[\mathrm{IOS}_A/q-1\right]O_p(n^{-1})+\sqrt{n}O_p(n^{-1})O_p(n^{-1}) \\ &= \sqrt{n}\left[\mathrm{IOS}_A/q-1\right]^2+\sqrt{n}O_p(1)O_p(n^{-1})+O_p(n^{-3/2}) \\ &= \sqrt{n}\left[\mathrm{IOS}_A/q-1\right]^2+O_p(n^{-1/2})+O_p(n^{-3/2}) \\ &= \sqrt{n}\left[\mathrm{IOS}_A/q-1\right]^2+O_p(n^{-1/2}) \\ &= \sqrt{n}\left[\left(q^*+o_p(1)\right)/q-1\right]^2+O_p(n^{-1/2}) \\ &= \sqrt{n}\left[q^*/q-1+o_p(1)\right]^2+O_p(n^{-1/2}) \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{n}\,[q^*/q-1]^2 + 2\sqrt{n}\,(q^*/q-1)\,o_p(1) + O_p(n^{-1/2}) \\
&= \sqrt{n}\,[q^*/q-1]^2 + 2\sqrt{n}\,(q^*/q-1)\,o_p(1) + O_p(n^{-1/2}).
\end{aligned}
$$

When the model is misspecified so that $q^* \neq q$, we have

$$J_0 = O_p(\sqrt{n}).$$

Since in J_1 is always large than zero, the order of the power of BMT is no less than $O_p(\sqrt{n})$.

5.3 Appendix 3: Computing BMT in Latent Variable Models

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

$$\begin{cases} y_t = F(z_t, u_t, \boldsymbol{\theta}) \\ z_t = G(z_{t-1}, v_t, \boldsymbol{\theta}) \end{cases}. \tag{30}$$

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

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}. \tag{31}$$

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 (31) 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 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 an 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 $\boldsymbol{\theta}$.

5.3.1 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

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = E_{\boldsymbol{\theta}^{(r)}}\{\mathcal{L}_c(\mathbf{y},\boldsymbol{z}|\boldsymbol{\theta})|\mathbf{y}, \theta^{(r)}\}, \tag{32}$$

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 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.

5.3.2 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.

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

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

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 $q_t(\cdot|\cdot)$, different particle filtering algorithms have been proposed in the literature, including the bootstrap particle filters of Gordon et al. (1993) and the auxiliary particle filters of Pitt and Shephard (1999). In this paper, we use the auxiliary particle 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.

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

Consider the state-space system

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

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

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

With the initial condition $x_{0|0}$ and $P_{0|0}$, the Kalman Filter algorithm is as follows:

$$\begin{aligned} x_{t|t-1} &= T x_{t-1|t-1}, \\ P_{t|t-1} &= T P_{t-1|t-1} T' + RQR', \end{aligned}$$

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 $t = 1, 2...n$.

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 5.1 *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:*

$$DF(X) = \frac{\partial vech(F(X))}{\partial (vech(X))'}.$$

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

Definition 5.2 *Let A be an $m \times n$ matrix. There exists a unique $mn \times mn$ permutation matrix K_{mn} which is defined as:*

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

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

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

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

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

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

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

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

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

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

where

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

and

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

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

$$
vech\left(P_{0|0}\right) = \left(I_{n_s^2} - T \otimes T\right)^{-1} vech\left(RQR'\right),
$$

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

5.5 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

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

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

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

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

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

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

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

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

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

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

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 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.

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.

Chan, H. P., and Lai, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *The Annals of Statistics*, **41(6)**, 2877-2904.

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.

Cheng, T., Gao, J., and Phillips, P. C. B. (2017). Bayesian estimation based on summary statistics: double asymptotics and practice. Working Paper.

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.

Chopin, N. (2002). A sequential particle filter for static models, *Biometrika*, **89(3)**, 539–551.

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.

Flegal, J. M., and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Annals of Statistics*, **38(2)**, 1034?70.

Gallant, A. R., and White, H. (1988). *A unified theory of estimation and inference for nonlinear dynamic models. Blackwell.*

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

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, **8(S1)**, S85-S118.

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. (1989). 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.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability Surveys*, **1(299-320)**, 5-1.

Jones, G. L., Haran, M., Caffo, B. S., & Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association,* **101(476)**, 1537-1547.

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., Yu, J. and Zeng, T.(2017). Integrated deviation information criterion for latent variable models. *Working paper*, Singapore Management University.

Lintner, J. (1965). The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **47**, 13-37.

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

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.

Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer.

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

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19**, 425-442.

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(4)**, 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.