



MONASH  
BUSINESS  
SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

**Bayesian Estimation Based on Summary Statistics:  
Double Asymptotics and Practice**

Tingting Cheng, Jiti Gao and Peter CB Phillips

March 2017

Working Paper 04/17

# Bayesian Estimation Based on Summary Statistics: Double Asymptotics and Practice

TINGTING CHENG, School of Finance, Nankai University

JITI GAO, Monash University

PETER CB PHILLIPS, Yale University, University of Auckland,  
University of Southampton, & Singapore Management University

## **Abstract:**

Ergodic theorem shows that ergodic averages of the posterior draws converge in probability to the posterior mean under the stationarity assumption. The literature also shows that the posterior distribution is asymptotically normal when the sample size of the original data considered goes to infinity. To the best of our knowledge, there is little discussion on the large sample behaviour of the posterior mean. In this paper, we aim to fill this gap. In particular, we extend the posterior mean idea to the conditional mean case, which is conditioning on a given vector of summary statistics of the original data. We establish a new asymptotic theory for the conditional mean estimator when both the sample size of the original data concerned and the number of Markov chain Monte Carlo iterations go to infinity. Simulation studies show that this conditional mean estimator has very good finite sample performance. In addition, we employ the conditional mean estimator to estimate a GARCH(1,1) model for S&P 500 stock returns and find that the conditional mean estimator performs better than quasi-maximum likelihood estimation in terms of out-of-sample forecasting.

**Keywords:** Bayesian average, conditional mean estimation, ergodic theorem, summary statistic

**JEL Classification:** C11, C15, C21

# 1 Introduction

In Bayesian analysis, posterior means are commonly used as point estimates for unknown parameters. However, little attention has been paid to the large sample properties of posterior means. In this paper, we aim to address this issue. Particularly, we focus on the investigation of large sample properties of a conditional mean estimator, which includes the posterior mean as a special case.

Suppose  $T_n$  is a vector of summary statistics of the original data  $(x_1, x_2, \dots, x_n)$  with sample size  $n$ . Let  $\theta$  be a  $d$ -dimensional vector of unknown parameters for a model which seeks to explain  $T_n$ . Our interest is learning about  $\theta$ . In Bayesian inference, inferences are normally based on the conditional density of  $\theta$  given the data denoted as  $f_n(\theta|T_n)$ . The conditional mean is then used as a point estimate of  $\theta$ , which is expressed as

$$\mathbb{E}[\theta|T_n] = \int \theta f_n(\theta|T_n) d\theta. \quad (1.1)$$

However, most of the time, (1.1) cannot be worked out analytically, especially in high dimensions. Therefore we evaluate (1.1) by doing simulation to obtain a sequence of draws  $\{\theta_{jn}\}$ ,  $j = 1, 2, \dots, m$ , from  $f_n(\theta|T_n)$ , where  $m$  denotes the number of iterations. For simplicity, we assume here that conditioning on  $T_n$ , we can obtain that  $\{\theta_{jn}\}$  is a sequence of stationary Markov chains. We then denote the simple average  $\frac{1}{m} \sum_{j=1}^m \theta_{jn}$  as  $\hat{\theta}_{mn}$  before we use  $\hat{\theta}_{mn}$  to approximate  $\mathbb{E}[\theta|T_n]$ . It can be checked that this approximation is valid because by the law of large numbers, as  $m \rightarrow \infty$ , we have  $\hat{\theta}_{mn} - \mathbb{E}[\theta|T_n] \rightarrow_P 0$  for any fixed  $n$ . Now, one natural question is: *What is the asymptotic behaviour of  $\hat{\theta}_{mn}$  and  $\mathbb{E}[\theta|T_n]$  when  $n \rightarrow \infty$ ?*

Admittedly, Bayesian inference is based on a given finite sample. However, with more and more data becoming available, it would be meaningful to investigate whether the performance of the conditional mean estimate  $\hat{\theta}_{mn}$  differs very much between small and large samples. For example, suppose we want to capture the movement of a stock market return by a GARCH (1,1) model given by

$$\begin{aligned} x_t &= \varepsilon_t \sigma_t, \quad \varepsilon_t \sim N(0, 1), \quad t = 1, 2, \dots, n, \\ \sigma_t^2 &= \theta_1 + \theta_2 y_{t-1}^2 + \theta_3 \sigma_{t-1}^2. \end{aligned}$$

Let  $\theta = (\theta_1, \theta_2, \theta_3)^\top$ . Will the behaviour of the conditional mean estimates of  $\theta$  change if the sample size varies from 500 to 10,000? In other words, what is the difference between  $\mathbb{E}[\theta|T_{500}]$  and  $\mathbb{E}[\theta|T_{10,000}]$ ?

To the best of our knowledge, there is little discussion available to answer this question. Until recently, most efforts were aimed at investigating the asymptotic behaviour of posterior distributions rather than that of posterior means. For example, [Walker \(1969\)](#) showed that

under suitable regularity conditions, as  $n \rightarrow \infty$ , the posterior distribution converges to a normal distribution. Based on Walker (1969), Chen (1985) further introduced three sets of conditions to establish an asymptotic posterior normality. Phillips (1996), and Phillips and Ploberger (1996) also developed Bayesian inference for stationary and nonstationary time series and provided the limiting form of the Bayesian data density for a general case of likelihoods and prior distributions. Kim (1998) also established an asymptotic posterior normality for a nonstationary time series setting. Chernozhukov and Hong (2003) developed a class of quasi-Bayesian estimators (QBEs), which are defined using a general econometric criterion function in place of the parametric likelihood function. They established  $\sqrt{n}$ -consistency and asymptotic normality for QBEs. The QBEs are computed using Markov chain Monte Carlo methods. Beaumont, Zhang and Balding (2002) proposed an approximate Bayesian computation (ABC) method to solve complex problems in population genetics, in which the principle of the ABC is that we make the best use of a vector of summary statistics rather than the whole sample  $(x_1, x_2, \dots, x_n)$ . Our work is related to the ABC idea. More recently, Gao and Hong (2014) established a link between the ABC method and the implementation of the so-called generalized method of moments (GMM).

The main contributions of this paper are summarized as follows.

- (i) We improve existing results for the posterior distribution with a fast rate of convergence, and also establish an asymptotic theory for the conditional mean estimator.
- (ii) We conduct several simulation studies to evaluate the finite sample performance of the conditional mean estimator, and we also employ the conditional mean estimator to estimate a GARCH(1,1) model for S&P 500 stock returns.

The rest of this paper is organized as follows. Section 2 proposes the conditional mean estimator. In Section 3, we establish new asymptotic properties for the conditional mean estimator. Section 4 presents Monte Carlo simulation studies to examine the finite sample performance of the conditional mean estimator. In Section 5, a GARCH(1,1) model for S&P 500 stock returns is used to illustrate the advantages of the conditional mean estimator. Section 6 concludes the paper. The proofs of the main theorems are given in an appendix.

## 2 Conditional mean estimation

Consider a random variable  $X$  with a density function  $f(x; \theta)$  characterized by  $\theta$ , a  $d$ -dimensional vector of unknown parameters. Let  $\theta_0$  be the true value of  $\theta$ . We aim to estimate  $\theta_0$  by the conditional mean  $\mathbb{E}[\theta|T_n]$ , where  $T_n$  is a vector of summary statistics of the original data  $(x_1, x_2, \dots, x_n)$  generated from  $f(x; \theta)$ .

We can construct the likelihood function given by

$$L_n(\theta) = L_n(\theta|T_n) = \prod_{i=1}^n f(x_i; \theta). \quad (2.1)$$

Define  $l_n(\theta) = \log L_n(\theta)$  as the log-likelihood function. Let  $\pi(\theta)$  denote a prior density of  $\theta$ . The conditional (posterior) density of  $\theta$  given  $T_n$  is

$$f_n(\theta|T_n) = \frac{e^{l_n(\theta)}\pi(\theta)}{\int e^{l_n(\theta)}\pi(\theta)d\theta}. \quad (2.2)$$

The conditional (posterior) mean of  $\theta$  given  $T_n$  can be expressed as

$$\mathbb{E}[\theta|T_n] = \int \theta f_n(\theta|T_n)d\theta. \quad (2.3)$$

This paper considers using the conditional distribution of  $\theta$  given  $T_n$  rather than  $X_n = (x_1, x_2, \dots, x_n)$ , mainly because using  $T_n$  is enough for the case where it is a vector of sufficient statistics. Meanwhile, it is both theoretically and conventionally more attractive in using  $T_n$  as a fixed-dimensional vector than using  $X_n$  as an  $n$ -dimensional vector of the full sample. Note that  $\hat{\theta}_n$  can be chosen as  $T_n$  in such cases where  $\hat{\theta}_n$  as the maximum likelihood estimator of  $\theta_0$ , the true value of  $\theta$ , is already a vector of sufficient summary statistics.

Due to the intractability of the likelihood function in some cases, equation (2.3) may not have any closed form. We thus evaluate (2.3) by simulating a sequence of draws from  $f_n(\theta|T_n)$ . If the conditional density  $f_n(\theta|T_n)$  is available for sampling, we can obtain a sequence of independent draws. If direct sampling is infeasible, we can use either an importance sampling or a cumulative distribution function transformation method. By such a method, we can obtain independent draws. If such methods are not applicable, we can instead use a Markov chain Monte Carlo (MCMC) algorithm, such as Metropolis-Hasting algorithm, one of the most popular MCMC methods.

Then, we approximate  $\mathbb{E}[\theta|T_n]$  by

$$\hat{\theta}_{mn} = \frac{1}{m} \sum_{j=1}^m \theta_{jn},$$

where  $\{\theta_{jn}, j = 1, 2, \dots, m\}$  denotes the  $j$ -th draw conditioning on  $T_n$  and  $m$  denotes the number of Markov chain Monte Carlo iterations. We refer to  $\hat{\theta}_{mn}$  as the conditional mean estimator (CME hereafter). In the next section, we establish some new asymptotic properties for both  $f_n(\theta|T_n)$  and  $\hat{\theta}_{mn}$ .

### 3 Asymptotic properties

Before we establish some asymptotic properties for  $f_n(\theta|T_n)$  and  $\hat{\theta}_{mn}$ , we introduce the following notation. Let  $\hat{\theta}_n$  denote the maximum likelihood estimator of  $\theta$ , and  $\theta_0$  be the true value of  $\theta$

involved in the model under consideration. Under Assumption 1(i) below, both the first and second derivatives,  $l_n^{(1)}(\cdot)$  and  $l_n^{(2)}(\cdot)$ , of  $l_n(\cdot)$ , do exist. Let  $\Delta_n(\widehat{\theta}_n) = -l_n^{(2)}(\widehat{\theta}_n)$ . For any given  $c > 0$ , denote  $C_n = \left\{ \theta : \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| < c \right\}$  and  $D_n = \left\{ \theta : \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \geq c \right\}$ .

In this section, we introduce the necessary conditions to establish our asymptotic results. Note that Assumptions 1–4 below may not be the weakest possible.

**Assumption 1.** (i) Suppose that  $f(\cdot; \theta)$  is twice differentiable with respect to  $\theta$ , and the second-order derivative,  $f^{(2)}(\cdot; \theta)$ , is continuous.

(ii) The density  $\pi(\theta)$  is twice differentiable and its second derivative is continuous.

**Assumption 2.** (i) Suppose that conditioning on  $T_n$ ,  $\{\theta_{jn}\}$  is a sequence of Harris and geometrically ergodic Markov chains with stationary distribution such that  $\mathbb{E}[\theta_{1n}|T_n] = \mathbb{E}[\theta|T_n]$ . Let  $\max_{n \geq 1} \mathbb{E}[|\theta_{1n}|^{2+\epsilon_0} | T_n] < \infty$  for some  $\epsilon_0 > 0$ .

(ii) Let  $\Delta_n^{-1}(\widehat{\theta}_n) \rightarrow_P 0$  and  $\frac{\Delta_n^{1/2}(\widehat{\theta}_n)\sigma_n}{\sqrt{m}} \rightarrow_P 0$  as  $n \rightarrow \infty$  and  $m \rightarrow \infty$ , where  $0 < \sigma_n^2 = \text{Var}[\theta_{1n}|T_n] + 2 \sum_{k=1}^{\infty} \gamma_n(k|T_n) < \infty$ , where  $\gamma_n(k|T_n) = \mathbb{E}[\theta_{1n}\theta_{1+k,n}|T_n] - \mathbb{E}[\theta_{1n}|T_n]\mathbb{E}[\theta_{1+k,n}|T_n]$ .

**Assumption 3.** (i) Let  $\int_{\{\theta: \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \geq \lambda\}} e^{l_n(\theta) - l_n(\widehat{\theta}_n)} \frac{\pi(\theta)}{\pi(\widehat{\theta}_n)} d\theta = o_P\left(\Delta_n^{-1/2}(\widehat{\theta}_n)\right)$  for  $n \geq 1$  and  $\lambda > 0$  both large enough.

(ii) Let  $\int_{\{\theta: \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \geq \lambda\}} f_n(\theta|T_n) d\theta = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n))$  and  $\int_{\{\theta: \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \geq \lambda\}} \theta f_n(\theta|T_n) d\theta = o_P(\Delta_n^{-1/2}(\widehat{\theta}_n))$  for  $n \geq 1$  and  $\lambda > 0$  both large enough.

**Assumption 4.** As  $n \rightarrow \infty$ ,  $\Delta_n^{1/2}(\widehat{\theta}_n) \left( \widehat{\theta}_n - \theta_0 \right) \rightarrow_D N(0, I_d)$ , where  $I_d$  is the  $d \times d$  dimensional identity matrix.

Assumptions 1 and 3 are similar to those used by Phillips (1996); Phillips and Ploberger (1996); and Kim (1998). Assumption 1(i) is standard to ensure that  $l_n(\theta)$  is twice differentiable with respect to  $\theta$ . It ensures that we can do Taylor expansion for  $l_n(\theta)$ . Assumption 1(ii) is used to make sure that we can do Taylor expansion for  $\pi(\theta)$ . Assumption 3 is another commonly used one. For example, if  $\frac{T_n - \theta}{\sigma_n} \rightarrow_D N(0, 1)$  and  $\sigma_n \rightarrow 0$  as  $n \rightarrow \infty$ , then Assumption 3 is automatically satisfied.

Assumption 2 basically requires that  $m$  going to infinity is faster than  $n$  going to infinity, and this simply reduces to requiring  $m \rightarrow \infty$  when  $\Delta_n(\widehat{\theta}_n)$  is proportional to  $n$  and  $\sigma_n^2$  is proportional to  $n^{-1}$  in many cases. It is pointed out that the geometric ergodicity assumption can be weakened to a polynomial ergodicity condition (see, for example, Corollaries 1 and 2 of Jones (2004)) without affecting the main results in Theorem 2 below. Assumption 4 imposes a general condition to allow that  $T_n$  is a vector of summary statistics of either independent, or stationary time series, or nonstationary time series. In either the independent or dependent

stationary time series case, an asymptotic normality can be the limiting distribution. As shown in Phillips (1996), Phillips and Ploberger (1996) and Kim (1998), Assumption 4 can be verified in the independent data as well as dependent stationary and nonstationary time series cases. In general, the asymptotic theory of the MLE is also available at Amemiya (1985), Fuller (1996) and Ling and Li (1998).

We now establish the main theorems of this paper.

**Theorem 1.** *Under Assumptions 1, 3(i) and 4, as  $n \rightarrow \infty$  and  $m \rightarrow \infty$ , we have*

$$\Delta_n(\widehat{\theta}_n) \left( \int_{C_n} f_n(\theta|T_n) d\theta - \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du \right) - \frac{\pi^{(2)}(\theta_0)}{2\pi(\theta_0)} \cdot \frac{1}{\sqrt{2\pi}} \int_{-c}^c u^2 e^{-\frac{1}{2}u^2} du \rightarrow_P 0 \quad (3.1)$$

whenever  $\pi(\theta_0) \neq 0$ , where  $C_n = \left\{ \theta : \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \leq c \right\}$  for any given  $c > 0$ , and  $\pi^{(2)}(\cdot)$  denotes the second-order derivative of  $\pi(\cdot)$ .

Theorem 1 strengthens and complements existing results (see, for example, Walker (1969), Chen (1985) and Kim (1998)) with a fast rate of convergence. In many conventional cases,  $\Delta_n(\widehat{\theta}_n)$  is proportional to  $n$ . In such cases, the rate of convergence of  $\int_{C_n} f_n(\theta|T_n) d\theta$  to  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$  can be as fast as  $n^{-1}$ .

Meanwhile, it should be pointed out that the constant  $c = c_n$  involved in Theorem 1 can also be varying and increasing with  $n$ , as long as  $\lambda_{\max} \left( \Delta_n^{-1/2}(\widehat{\theta}_n) \right) c_n \rightarrow_P 0$  as  $n \rightarrow \infty$  to ensure that a Taylor expansion is valid in a neighbourhood of  $\widehat{\theta}_n$ , where  $\lambda_{\max}(A)$  denotes the largest eigenvalue of matrix  $A$ . As a matter of fact, the larger the value of  $c$ , the faster of convergence of  $\Delta_n(\widehat{\theta}_n) \left( \int_{C_n} f_n(\theta|T_n) d\theta - \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du \right) - \frac{\pi^{(2)}(\theta_0)}{2\pi(\theta_0)} \cdot \frac{1}{\sqrt{2\pi}} \int_{-c}^c u^2 e^{-\frac{1}{2}u^2} du \rightarrow_P 0$ . Thus, in the finite-sample evaluation, it is probably more interesting to look at the case where  $c$  is not large, as  $c = 1.0$  in Section 4 below.

**Theorem 2.** (i) *Under Assumptions 1–4, as  $n \rightarrow \infty$  and  $m \rightarrow \infty$ , we have*

$$\Delta_n^{1/2}(\widehat{\theta}_n) \left( \widehat{\theta}_{mn} - \widehat{\theta}_n \right) = o_P(1). \quad (3.2)$$

Let Assumptions 1, 2, 3(i) and 4 hold. If we can strengthen Assumption 3(ii) to

$$\int_{\left\{ \theta : \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \geq \lambda \right\}} f_n(\theta|T_n) d\theta = o_P(\Delta_n^{-1}(\widehat{\theta}_n))$$

and

$$\int_{\left\{ \theta : \Delta_n^{1/2}(\widehat{\theta}_n) \|\widehat{\theta}_n - \theta\| \geq \lambda \right\}} \theta f_n(\theta|T_n) d\theta = o_P(\Delta_n^{-1}(\widehat{\theta}_n))$$

for large enough  $\lambda > 0$ , then equation (3.4) can be strengthened to

$$\Delta_n(\widehat{\theta}_n) \left( \widehat{\theta}_{mn} - \widehat{\theta}_n \right) \rightarrow_P \frac{\pi^{(1)}(\theta_0)}{\pi(\theta_0)} \quad (3.3)$$

when  $\pi(\theta_0) \neq 0$  and  $\frac{\Delta_n(\hat{\theta}_n)\sigma_n}{\sqrt{m}} \rightarrow_P 0$  as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ .

(ii) Under Assumptions 1-4, as  $n \rightarrow \infty$  and  $m \rightarrow \infty$ , we have

$$\frac{\sqrt{m}}{\sigma_n} \left( \hat{\theta}_{mn} - \mathbb{E}[\theta|T_n] \right) \rightarrow_D N(0, I_d), \quad (3.4)$$

$$\Delta_n^{1/2}(\hat{\theta}_n) \left( \hat{\theta}_{mn} - \theta_0 \right) \rightarrow_D N(0, I_d), \quad (3.5)$$

where  $\hat{\theta}_{mn} = \frac{1}{m} \sum_{j=1}^m \theta_{jn}$ .

Equation (3.2) of Theorem 2(i) also shows that the rate of convergence of  $\hat{\theta}_{mn} - \mathbb{E}[\theta|T_n]$  to zero can be as fast as  $\frac{1}{\sqrt{mn}}$  in the case where  $\sigma_n^2$  is proportional to  $n^{-1}$ . Equation (3.3) further shows that the rate of approximation between  $\hat{\theta}_{mn}$  and  $\hat{\theta}_n$  can be as fast as  $n^{-1}$ . Theorem 2(ii) shows that an asymptotic normality is achievable for the conditional mean estimator  $\hat{\theta}_{mn}$  with a rate of convergence of  $\Delta_n^{1/2}(\hat{\theta}_n)$ . Equation (3.5) shows not only that  $\hat{\theta}_{mn}$  is actually a consistent estimator of  $\theta_0$  even though  $n \rightarrow \infty$ , but also that we can now consider various estimation and testing issues about  $\theta_0$  for the case even when  $n \rightarrow \infty$ . In short, it is our knowledge that both Theorems 1 and 2 establish some considerably new findings about the asymptotic consistency of the posterior density function with a rate of convergence in Theorem 1 and an asymptotic normality for the conditional mean estimator,  $\hat{\theta}_{mn}$ , in Theorem 2.

Before we give the proofs of Theorems 1 and 2 in the Appendix, we will evaluate the finite sample properties of the estimation method and its resulting theory in Sections 4 and 5 below.

## 4 Simulation

In this simulation study, we have two objectives. First, we examine the finite sample performance of the proposed conditional mean estimator through six simulated examples. Second, we investigate the convergence of  $\int_{C_n} f_n(\theta|T_n)d\theta$  to  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$  graphically, where  $C_n = \left\{ \theta : \Delta_n^{1/2}(\hat{\theta}_n) \|\hat{\theta}_n - \theta\| \leq c \right\}$ .

The six simulated examples that we considered include a location model, a linear model, a random coefficient model, an AR(1) process, a GARCH model and a normal mixture density. For each simulated example, we consider sample sizes  $n = 200, 600$  and  $1200$ . For each sample size, we generate 1000 samples. Let  $\theta_0 = (\theta_{10}, \theta_{20}, \dots, \theta_{p0})^\top$  denote the true value of the unknown parameter vector  $\theta = (\theta_1, \theta_2, \dots, \theta_p)^\top$  and  $\hat{\theta}_{mn} = (\hat{\theta}_{1,mn}, \hat{\theta}_{2,mn}, \dots, \hat{\theta}_{p,mn})^\top$  denote the estimated parameter vector, where  $p$  denotes the dimension of the parameter vector.

Before we discuss how to evaluate the finite sample performance of  $\hat{\theta}_n$  and  $\hat{\theta}_{mn}$ , we show how  $\theta_0$  is related to the moments of the original data as follows:

- In Example 1 below,  $\theta_0$  is the unconditional mean of  $\mathbb{E}[x_1]$ .



- In Example 2 below,  $\theta_0 = (\alpha_0, \beta_0)^\top$ , where  $\beta_0 = \frac{\text{Cov}(x_1, y_1)}{\text{Var}(x_1)}$  and  $\alpha_0 = \mathbb{E}[y_1] - \frac{\text{Cov}(x_1, y_1)}{\text{Var}(x_1)} \mathbb{E}[x_1]$ . Similar definitions can be given for  $\beta_0$  involved in Example 3.
- In Example 4 below,  $\theta_0 \equiv \rho_0 = (\mathbb{E}[y_1^2])^{-1} \mathbb{E}[y_1 y_2]$ .
- In Example 5 below,  $\theta_0 = (\theta_{10}, \theta_{20}, \theta_{30})^\top$  satisfies  $\theta_{10} = (1 - \theta_{20} - \theta_{30}) \mathbb{E}[y_1^2]$ .
- In Example 6 below,  $\theta_0 = \frac{1}{2}\theta_{10} + \frac{1}{2}\theta_{20}$  is the population mean of a mixture normal distribution.

To examine the finite sample performance of the estimation method, we compute the absolute biases (abs.bias), standard deviations (std) and the mean squared errors (mse) of  $\widehat{\theta}_{k,mn}$ ,  $k = 1, 2, \dots, p$ , as follows:

$$\text{abs.bias}_k = \frac{1}{1000} \sum_{r=1}^{1000} \left| \widehat{\theta}_{k,mn}(r) - \theta_{k0} \right|, \quad \text{std}_k = \sqrt{\frac{1}{1000} \sum_{r=1}^{1000} (\widehat{\theta}_{k,mn}(r) - \overline{\widehat{\theta}_k})^2}, \quad (4.1)$$

$\text{mse}_k = \frac{1}{1000} \sum_{r=1}^{1000} (\widehat{\theta}_{k,mn}(r) - \theta_{k0})^2$ , in which  $\widehat{\theta}_{k,mn}(r)$  is the  $r$ -th replication of the  $k$ -th component,  $\widehat{\theta}_{k,mn}$ , of  $\widehat{\theta}_{mn}$ , and  $\overline{\widehat{\theta}_k} = \frac{1}{1000} \sum_{r=1}^{1000} \widehat{\theta}_{k,mn}(r)$ .

**Example 1.** Consider a location model given by

$$x_i = \theta + e_i, \quad \text{for } i = 1, 2, \dots, n, \quad (4.2)$$

where  $\theta$  follows a normal distribution  $N(\theta_0, 1)$  and  $e_i$  is assumed to be independent and identically distributed (i.i.d.) with the standard normal distribution. Here  $\theta$  and  $e_i$  are independent. For each replication, we generate  $\theta$  from  $N(\theta_0, 1)$  and generate  $e_i$  from  $N(0, 1)$  and then obtain  $x_i$  by  $x_i = \theta + e_i$ , for  $i = 1, 2, \dots, n$ .

In this example, we choose  $\theta_0 = 0.8$ . So the density of  $\theta$ ,  $\pi(\theta)$ , is given by

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\theta - 0.8)^2}{2}\right). \quad (4.3)$$

Given a sample  $X_n = (x_1, x_2, \dots, x_n)^\top$ , define  $T_n = \sum_{i=1}^n x_i$  and  $U_n = \sum_{i=1}^n x_i^2$ . Then the likelihood function can be written as

$$L(\theta|T_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{U_n - 2\theta T_n + n\theta^2}{2}\right). \quad (4.4)$$

By definition, the conditional density of  $\theta$  given  $T_n$  is

$$f_n(\theta|T_n) = \frac{\pi(\theta)L(\theta|T_n)}{\int \pi(\theta)L(\theta|T_n)d\theta}. \quad (4.5)$$

Substitute (4.3) and (4.4) into (4.5). Then we can get

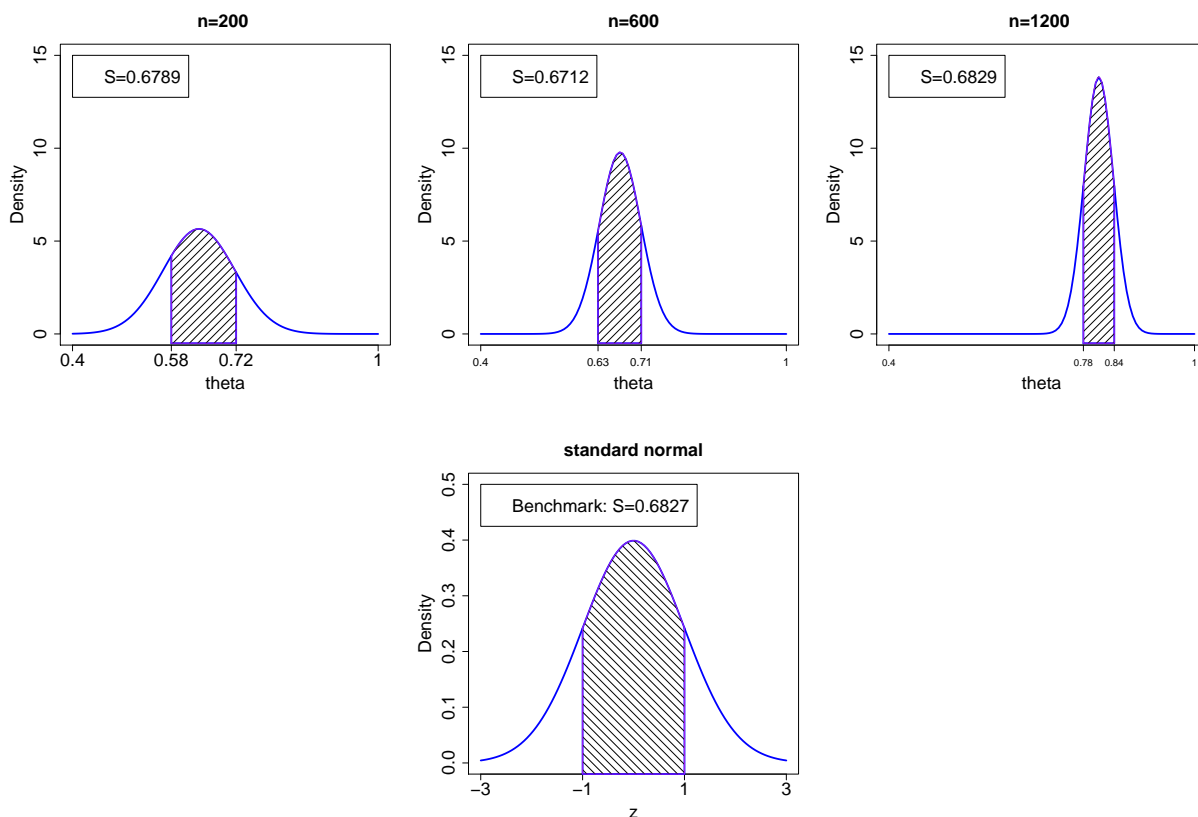
$$f_n(\theta|T_n) = \sqrt{\frac{n+1}{2\pi}} \exp\left(-\frac{(n+1)(\theta - \frac{T_n+0.8}{n+1})^2}{2}\right). \quad (4.6)$$

The conditional mean of  $\theta$  is then given by

$$\mathbb{E}[\theta|T_n] = \int \theta f_n(\theta|T_n) d\theta = \frac{T_n + 0.8}{n + 1}.$$

We compare  $\int_{C_n} f_n(\theta|T_n) d\theta$  with  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$  graphically for the case of  $c = 1.0$ , where  $C_n = \left\{ \theta : \Delta_n^{1/2}(\hat{\theta}_n) |\hat{\theta}_n - \theta| \leq c \right\}$ . Based on the first replication, we produce the plots for the shaded area of  $S = 0.6827$  and those for  $\int_{C_n} f_n(\theta|T_n) d\theta$  in Figure 1. Note that this is only based on one replication.

Figure 1: Plots of  $\int_{C_n} f_n(\theta|T_n) d\theta$  (the upper panel with sample sizes  $n=200, 600, 1200$ ) and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$ . In each plot,  $S$  denotes the shaded area.



We also considered the maximum likelihood estimator (MLE) of  $\theta_0$ . For the  $r$ -th replication,  $r = 1, 2, \dots, 1000$ , the maximum likelihood estimate of  $\theta$  can be computed by

$$\hat{\theta}_n(r) = \frac{1}{n} \sum_{i=1}^n x_{i,r}, \quad (4.7)$$

where  $x_{i,r}$ , for  $i = 1, 2, \dots, n$ , is the random sample from the  $r$ -th replication.

We compute the absolute bias, standard deviation and mean squared errors for both CME and MLE using (4.1). The results are presented in Table 1. From Table 1, we find that with the increase of sample size, the absolute bias, standard deviation and mean squared errors of the proposed conditional mean estimator decrease. This indicates that the proposed conditional mean estimator has very good finite sample performance. In addition, by comparing the results of CME and MLE, we can find that CME and MLE have similar performance in terms of the absolute bias, standard deviation and mean squared errors.

Table 1: Absolute bias, standard deviation and mean squared error of CME and MLE for  $\theta$  in the location model based on 1000 replications.

	$n$	$\mathbb{E}[\theta T_n]$	$\hat{\theta}_n$
abs.bias	200	0.0781	0.0785
	600	0.0469	0.0470
	1200	0.0324	0.0325
std	200	0.0984	0.0989
	600	0.0581	0.0582
	1200	0.0407	0.0408
mse	200	0.0097	0.0098
	600	0.0034	0.0034
	1200	0.0017	0.0017

From (4.6), it is easy to see that the conditional distribution of  $\theta$  given  $T_n$  is a normal distribution with mean  $\frac{T_n+0.8}{n+1}$  and variance  $\frac{1}{n+1}$ .

**Example 2.** Consider a linear regression model of the form

$$y_i = \alpha + \beta x_i + e_i, \quad \text{for } i = 1, 2, \dots, n, \quad (4.8)$$

where  $\alpha$  and  $\beta$  are random parameters, and the error term  $e_i$  is assumed to be i.i.d. with the standard normal distribution.

In this study,  $x_i$  is generated from a stationary AR(1) process  $x_t = 0.5x_{t-1} + u_t$  with  $u_t$  being generated from the standard normal distribution. We generate  $\alpha$  from  $N(1.0, 1.0)$  and  $\beta$  from  $N(2.5, 1.0)$ . So the densities of  $\alpha$  and  $\beta$  are given by

$$\pi(\alpha) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\alpha - 1)^2}{2}\right) \quad \text{and} \quad \pi(\beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\beta - 2.5)^2}{2}\right). \quad (4.9)$$

Given a sample  $X_n = (x_1, x_2, \dots, x_n)^\top$  and  $Y_n = (y_1, y_2, \dots, y_n)^\top$ .

Define  $T_n = (T_{x1}, T_{x2}, T_{y1}, T_{y2}, T_{xy})^\top$ , where  $T_{x1} = \sum_{i=1}^n x_i$ ,  $T_{x2} = \sum_{i=1}^n x_i^2$ ,  $T_{y1} = \sum_{i=1}^n y_i$ ,  $T_{y2} = \sum_{i=1}^n y_i^2$  and  $T_{xy} = \sum_{i=1}^n x_i y_i$ .

The likelihood function can be written as

$$L(\alpha, \beta | T_n) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left( -\frac{T_{y2} + \beta^2 T_{x2} - 2\beta T_{xy} + 2\alpha\beta T_{x1} - 2\alpha T_{y1} + n\alpha^2}{2} \right). \quad (4.10)$$

By definition, the conditional density of  $\theta = (\alpha, \beta)^\top$  given  $T_n$  is

$$\begin{aligned} f_n(\theta | T_n) &= \frac{\pi(\alpha) \pi(\beta) L(\alpha, \beta | T_n)}{\iint \pi(\alpha) \pi(\beta) L(\alpha, \beta | T_n) d\alpha d\beta} \\ &= \frac{1}{2\pi \sqrt{|\Sigma|}} \exp \left( -(\theta - \mu_\theta)^\top \Sigma^{-1} (\theta - \mu_\theta) \right), \end{aligned} \quad (4.11)$$

where  $1/|\Sigma| = (n+1)(1+T_{x2}) - T_{x1}^2$  and

$$\Sigma^{-1} = \begin{pmatrix} n+1 & T_{x1} \\ T_{x1} & 1+T_{x2} \end{pmatrix}, \quad \mu_\theta = \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} = \begin{pmatrix} \frac{1+T_{y1}-\mu_\beta T_{x1}}{n+1} \\ \frac{(T_{xy}+2.5)(n+1)-T_{x1}(T_{y1}+1)}{(1+T_{x2})(n+1)-T_{x1}^2} \end{pmatrix}. \quad (4.12)$$

The maximum likelihood estimators of  $\beta_0$  and  $\alpha_0$  are given by

$$\hat{\beta}_n = (X_n^\top X_n)^{-1} X_n^\top Y_n \quad \text{and} \quad \hat{\alpha}_n = \bar{y} - \hat{\beta}_n \bar{x}, \quad (4.13)$$

respectively, where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .

Then we compute the absolute bias, standard deviation and mean squared errors for CME and MLE using (4.1). The results are presented in Table 2. From Table 2, we find that with the increase of sample size, the absolute bias, standard deviation and mean squared errors of the proposed estimator decrease. This supports the theory that the proposed conditional mean estimator has some good finite sample performance. In addition, we find that the CME and MLE have similar performance.

From (4.11), we can see that the conditional distribution of  $\beta$  given  $T_n$  is a normal distribution with mean  $\mu_\beta$  and variance  $\frac{1}{1+T_{x2}}$  and the conditional distribution of  $\alpha$  given  $T_n$  is a normal distribution with mean  $\mu_\alpha$  with variance  $\frac{1}{n+1}$ .

Let  $c = 1.0$ . Based on the first replication, we produce the plots to visually see the change of  $\int_{C_{n\alpha}} f_n(\alpha | T_n) d\alpha$  and  $\int_{C_{n\beta}} f_n(\beta | T_n) d\beta$  with the increase of sample size in Figure 2, where  $C_{n\alpha} = \left\{ \alpha : \Delta_n^{1/2}(\hat{\alpha}_n) |\hat{\alpha}_n - \alpha| \leq c \right\}$  and  $C_{n\beta} = \left\{ \beta : \Delta_n^{1/2}(\hat{\beta}_n) |\hat{\beta}_n - \beta| \leq c \right\}$ .

Figure 2: Plots of  $\int_{C_{n\alpha}} f_n(\alpha|T_n)d\alpha$  (the upper panel with sample sizes  $n=200, 600, 1200$ ),  $\int_{C_{n\beta}} f_n(\beta|T_n)d\beta$  (the middle panel with sample sizes  $n=200, 600, 1200$ ) and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$ .

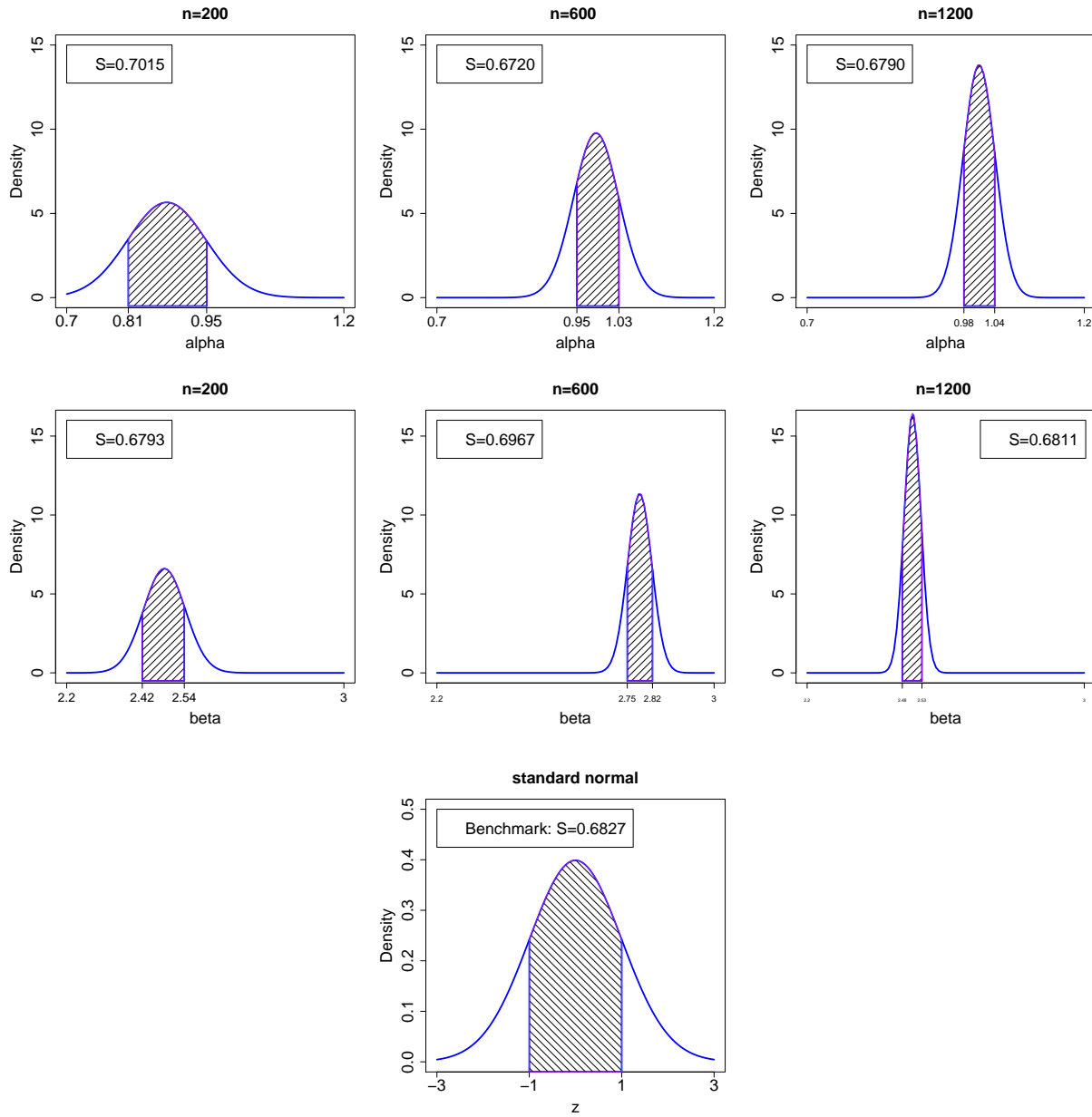


Table 2: Absolute bias, standard deviation and mean squared error of CME and MLE for parameters in a linear regression model based on 1000 replications.

		CME		MLE	
	$n$	$\mathbb{E}[\alpha T_n]$	$\mathbb{E}[\beta T_n]$	$\hat{\alpha}_n$	$\hat{\beta}_n$
abs.bias	200	0.1060	0.1121	0.1066	0.1125
	600	0.0586	0.0628	0.0587	0.0629
	1200	0.0404	0.0486	0.0404	0.0487
std	200	0.1326	0.1421	0.1333	0.1427
	600	0.0745	0.0774	0.0747	0.0775
	1200	0.0505	0.0606	0.0506	0.0607
mse	200	0.0176	0.0202	0.0178	0.0203
	600	0.0056	0.0060	0.0056	0.0060
	1200	0.0026	0.0037	0.0026	0.0037

**Example 3.** Consider a random coefficient model with one explanatory variable given by

$$y_i = \beta_i x_i + e_i, \quad \text{for } i = 1, 2, \dots, n, \quad (4.14)$$

where  $\beta_i = \beta + 0.5u_i$  with  $u_i$  being from the standard normal distribution, and  $e_i$  follows the standard normal distribution, in which  $\{u_i\}_{i=1}^n$  and  $\{e_i\}_{i=1}^n$  are assumed to be mutually independent.

Model (4.14) can be written as

$$y_i = \beta x_i + v_i, \quad (4.15)$$

where  $v_i = 0.5x_i u_i + e_i$ . It is easy to see that  $E[v_i] = 0$  and  $E[v_i^2] = 0.25x_i^2 + 1$ . In this example, we generate  $\beta$  from a normal distribution  $N(1.5, 1)$ . So the density of  $\beta$ ,  $\pi(\beta)$ , is given by

$$\pi(\beta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\beta - 1.5)^2}{2}\right). \quad (4.16)$$

Let  $Y_n = (y_1, y_2, \dots, y_n)^\top$  and  $X_n = (x_1, x_2, \dots, x_n)^\top$ . Let  $T_n$  be the generalised least squares (GLS) estimator of  $\beta$ , which is given by

$$T_n = (X_n^\top D^{-1} X_n)^{-1} X_n^\top D^{-1} Y_n, \quad (4.17)$$

where  $D = \text{diag}\{E[v_1^2], E[v_2^2], \dots, E[v_n^2]\}$ .

We approximate the conditional density of  $T_n$  given  $\beta$  by its limiting distribution  $f(T_n|\beta)$ , which is a normal distribution with mean  $\beta$  and variance  $\sigma_*^2 = (\sum_{i=1}^n x_i^2 (0.25x_i^2 + 1))^{-1}$ .

By definition, the conditional density of  $\beta$  given  $T_n$  is

$$f_n(\beta|T_n) = \frac{f(T_n|\beta)\pi(\beta)}{\int f(T_n|\beta)\pi(\beta)d\beta}.$$

It is easy to show that

$$f_n(\beta|T_n) = \sqrt{\frac{1 + \sigma_*^2}{2\pi\sigma_*^2}} \exp\left(-\frac{(1 + \sigma_*^2)(\beta - \frac{T_n + 1.5\sigma_*^2}{1 + \sigma_*^2})^2}{2\sigma_*^2}\right). \quad (4.18)$$

The conditional mean of  $\beta$  is then given by

$$\mathbb{E}[\beta|T_n] = \int \beta f_n(\beta|T_n) d\beta = \frac{T_n + 1.5\sigma_*^2}{1 + \sigma_*^2}.$$

The maximum likelihood estimator of  $\beta_0$  can be obtained by maximizing the following likelihood function

$$L(\beta|Y_n, X_n) = \frac{1}{\prod_{i=1}^n \sqrt{2\pi(0.25x_i^2 + 1)}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - x_i\beta)^2}{0.25x_i^2 + 1}\right). \quad (4.19)$$

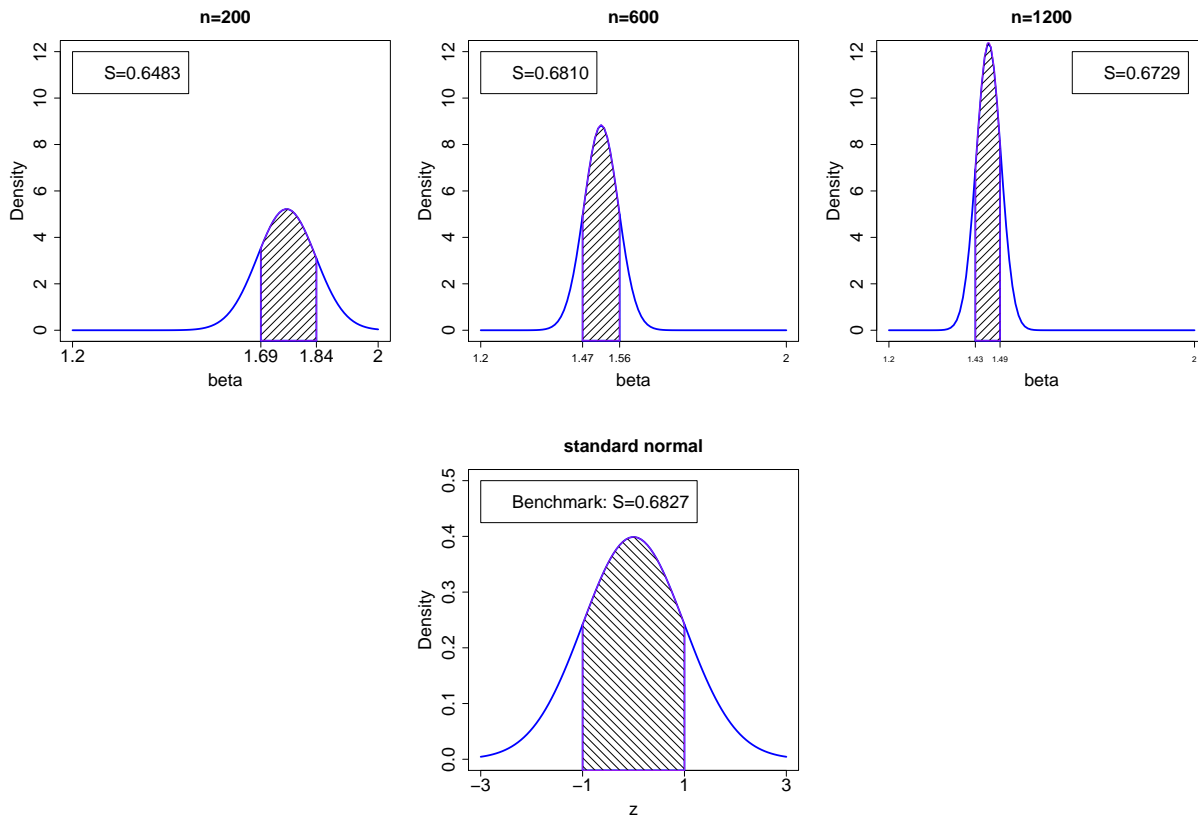
The resulting maximum likelihood estimate is denoted as  $\hat{\beta}_n$ . The results of the absolute bias, standard deviation and mean squared errors of CME and MLE are presented in Table 3. From Table 3, we find that with the increase of sample size, the absolute bias, standard deviation and mean squared errors of the proposed estimator decrease. This indicates that the proposed conditional mean estimator has very good finite sample performance. In addition, by comparing the results of CME and MLE, we can find that CME and MLE have similar performance in terms of the absolute bias, standard deviation and mean squared errors.

Table 3: Absolute bias, standard deviation and mean squared error of CME and MLE for  $\beta$  in the random coefficient regression model based on 1000 replications.

		CME	MLE	GLS
$n$		$\mathbb{E}[\beta T_n]$	$\hat{\beta}_n$	$T_n$
abs.bias	200	0.098238	0.098890	0.098889
	600	0.057195	0.057339	0.057321
	1200	0.040029	0.040074	0.040073
std	200	0.124469	0.125299	0.125294
	600	0.071303	0.071474	0.071460
	1200	0.049653	0.049720	0.049708
mse	200	0.015463	0.015670	0.015669
	600	0.005104	0.005128	0.005126
	1200	0.002471	0.002478	0.002477

From (4.18), it is clear that the conditional distribution of  $\beta$  given  $T_n$  is a normal distribution with mean  $\frac{T_n+1.5\sigma_*^2}{1+\sigma_*^2}$  and variance  $\frac{\sigma_*^2}{1+\sigma_*^2}$ .

Figure 3: Plots of  $\int_{C_n} f_n(\beta|T_n)d\beta$  (the upper panel with sample sizes  $n=200, 600, 1200$ ) and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$ .



Unlike Examples 1–3, we cannot obtain any closed forms for  $f_n(\theta|T_n)$  and the conditional cumulative distribution function  $F_n(\theta|T_n)$  in Examples 4 and 5 below. In those examples,  $F_n(\theta|T_n)$  is approximated by

$$\hat{F}_n(\theta|T_n) = \frac{1}{m} \sum_{j=1}^m \Phi \left( \frac{\hat{\theta}_{jn} - \theta}{h} \right), \quad (4.20)$$

where  $\Phi(\cdot)$  denotes the distribution function of the standard normal distribution. The bandwidth  $h$  is obtained by using the normal reference rule  $h = 1.06 \sigma_\theta m^{-1/5}$ , where  $\sigma_\theta$  denotes the standard deviation of simulated draws  $\{\hat{\theta}_{jn}, j = 1, 2, \dots, m\}$ .

The conditional density function  $f_n(\theta|T_n)$  is approximated by a kernel density estimator  $\hat{f}_n(\theta|T_n)$  with the form of

$$\hat{f}_n(\theta|T_n) = \frac{1}{mh} \sum_{j=1}^m \phi \left( \frac{\hat{\theta}_{jn} - \theta}{h} \right),$$



where  $h$  is bandwidth and  $\phi(\cdot)$  is the density function of the standard normal distribution. The choice of  $h$  is same as that in (4.20).

Based on the first replication, Figure 3 shows the finite sample behaviour of  $\int_{C_n} \widehat{f}_n(\beta|T_n)d\beta$  for the case of  $c = 1$ , where  $C_n = \left\{ \beta : \Delta_n^{1/2}(\widehat{\beta}_n)|\widehat{\beta}_n - \beta| \leq c \right\}$ .

**Example 4.** Consider an autoregressive model of the form

$$y_t = \rho y_{t-1} + e_t, \quad (4.21)$$

where  $\rho$  is random with expectation  $\rho_0 = 1$ ,  $y_0 = O_P(1)$ , and  $\{e_t\}$  is a sequence of independent and identically distributed random errors with  $E[e_1] = 0$  and  $E[e_1^2] = 1$ .

We generate data from (4.21) with  $\rho$  being generated from an uniform distribution  $U(0.9, 1.1)$ , which implies that  $y_t$  is nearly nonstationary when  $0.9 < \rho < 1.0$  and nonstationary when  $\rho \geq 1$ . We generate  $e_t$  from the standard normal distribution. We set  $y_0 = 1$ .

Given a sample  $Y_n = (y_1, y_2, \dots, y_n)^\top$ . Define  $T_n = (T_{n1}, T_{n2}, T_{n3})^\top$ , where  $T_{n1} = \sum_{t=1}^n y_t^2$ ,  $T_{n2} = \sum_{t=1}^n y_t y_{t-1}$  and  $T_{n3} = \sum_{t=1}^n y_{t-1}^2$ .

The likelihood function can be written as

$$L(\rho|T_n) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left( -\frac{T_{n1} - 2\rho T_{n2} + \rho^2 T_{n3}}{2} \right).$$

By definition, the conditional density of  $\rho$  given  $T_n$  is

$$f_n(\rho|T_n) = \frac{\pi(\rho) L(\rho|T_n)}{\int \pi(\rho) L(\rho|T_n) d\rho}.$$

The conditional mean of  $\rho$  is then given by  $\mathbb{E}[\rho|T_n] = \int \rho f_n(\rho|T_n) d\rho$ .

For the  $r$ -th replication,  $r = 1, 2, \dots, 1000$ , we use the random-walk Metropolis algorithm to sample  $\rho$  from  $f_n(\rho|T_n)$ . We record  $\rho$  at each iteration after the burn-in period to obtain the sequence  $\{\rho_{jn}(r)\}$ , for  $j = 1, 2, \dots, 10,000$ . Then the conditional mean estimate will be the ergodic mean of each recorded chain given by

$$\widehat{\rho}_{mn}(r) = \frac{1}{m} \sum_{j=1}^m \rho_{jn}(r),$$

where  $m = 10,000$ .

Then we compute the absolute bias, standard deviation and mean squared errors using (4.1). We also compare the finite sample performance of our estimate with the MLE of the form  $\widehat{\rho}_n = \left( \sum_{t=1}^n y_{t-1}^2 \right)^{-1} \sum_{t=1}^n y_{t-1} y_t$ .

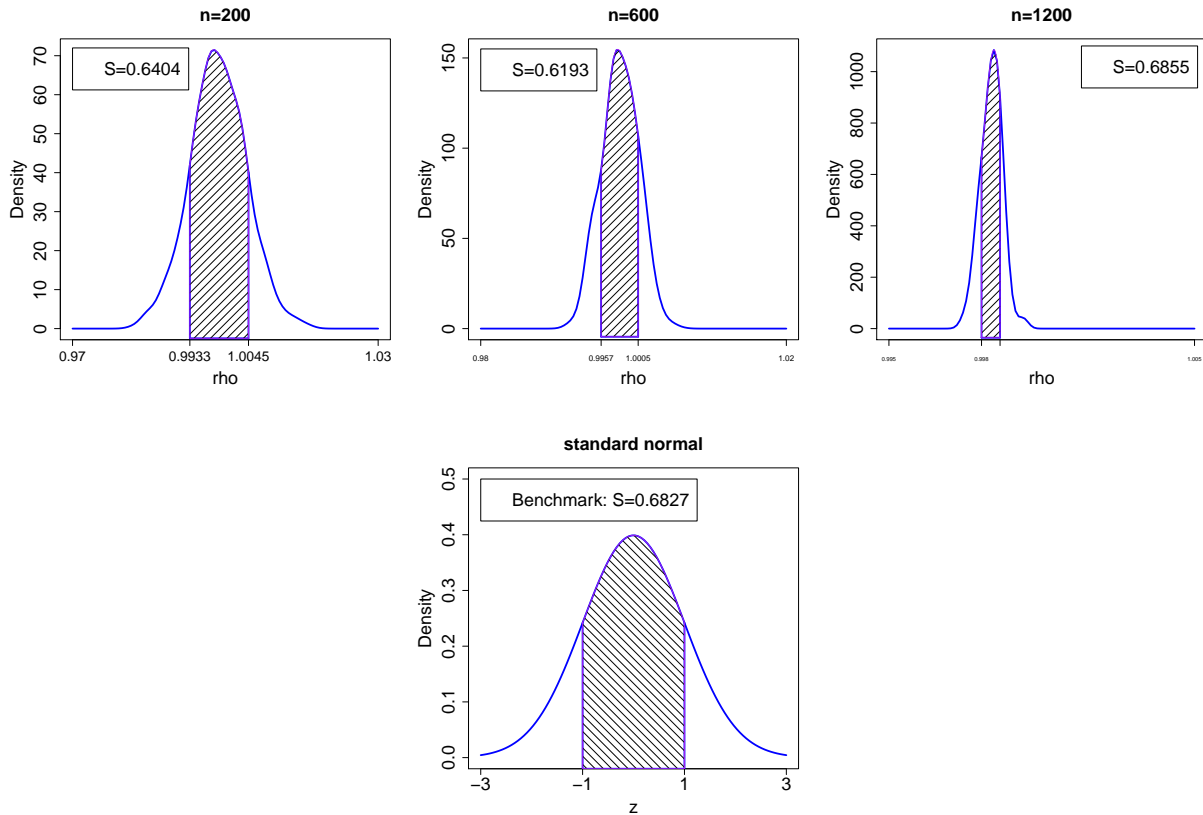
The results are presented in Table 4. From Table 4, we find that with the increase of sample size, the absolute bias, standard deviation and mean squared errors of the proposed estimator decrease. This indicates that the proposed conditional mean estimator has very good finite

sample performance. In addition, the performance of estimates of  $\rho$  by MLE and CME are very close.

Table 4: Absolute bias, standard deviation and mean squared error of CME and MLE for the parameter  $\rho$  in AR(1) model based on 1000 replications.

	$n$	$\hat{\rho}_{mn}$	$\hat{\rho}_n$
abs.bias	200	0.01264	0.01275
	600	0.00532	0.00544
	1200	0.00377	0.00372
std	200	0.01631	0.01661
	600	0.00583	0.00605
	1200	0.00351	0.00347
mse	200	0.00036	0.00038
	600	0.00005	0.00006
	1200	0.00002	0.00002

Figure 4: Plots of  $\int_{C_n} \hat{f}_n(\rho|T_n) d\rho$  (the upper panel with sample sizes  $n=200, 600, 1200$ ) and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$ .



Based on the first replication, we produce the plots to visually see the change of  $\int_{C_n} \widehat{f}_n(\rho|T_n)d\rho$  for the case of  $c = 1$ , where  $C_n = \left\{ \rho : \Delta_n^{1/2}(\widehat{\rho}_n)|\widehat{\rho}_n - \rho| \leq c \right\}$ . The finite sample closeness can be seen from Figure 4.

**Example 5.** A GARCH (1,1) model with Gaussian innovations can be written as

$$y_t = e_t \sigma_t, \quad e_t \sim N(0, 1), \quad t = 1, 2, \dots, n,$$

$$\sigma_t^2 = \theta_1 + \theta_2 y_{t-1}^2 + \theta_3 \sigma_{t-1}^2,$$

where  $\theta_1 > 0$ ,  $\theta_2, \theta_3 \geq 0$ ,  $\theta_2 + \theta_3 < 1$ . Note that the restrictions on the GARCH parameters  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  guarantee positivity of the conditional variance  $\sigma_t^2$  and make sure that the series  $y_t$  is covariance stationary. In this model, we assume  $\theta = (\theta_1, \theta_2, \theta_3)^\top$  is random.

We generate data from the GARCH model with  $\theta_1$  being generated from an uniform distribution  $U(0.01, 1)$ ,  $\theta_3$  from  $U(0, 1)$  and  $\theta_2$  from  $U(0, 1 - \theta_3)$ , respectively.

For this GARCH model, we choose the summary statistics  $T_n$  to be the maximum likelihood estimates of  $\theta$ . As the exact closed form of the conditional density of  $T_n$  given  $\theta$  is unknown, we use its limiting distribution  $f(T_n|\theta)$ , which is a normal distribution with mean  $\theta$  and covariance matrix  $\Sigma_\theta$ , in which  $\Sigma_\theta$  can be computed using the method proposed by Ma (2008).

By definition, the conditional density of  $\theta$  given  $T_n$  is

$$f_n(\theta|T_n) = \frac{\pi(\theta_1) \pi(\theta_2) \pi(\theta_3) f(T_n|\theta)}{\iiint \pi(\theta_1) \pi(\theta_2) \pi(\theta_3) f(T_n|\theta) d\theta_1 d\theta_2 d\theta_3}.$$

The conditional mean of  $\theta$  is then given by  $\mathbb{E}[\theta|T_n] = \int \theta f_n(\theta|T_n) d\theta$ .

For the  $r$ -th replication,  $r = 1, 2, \dots, 1000$ , we use the random-walk Metropolis algorithm to sample  $(\theta_1, \theta_2, \theta_3)^\top$  from  $f_n(\theta|T_n)$ . We record  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  at each iteration after the burn-in period to obtain sequences  $\{\theta_{1,jn}(r)\}$ ,  $\{\theta_{2,jn}(r)\}$  and  $\{\theta_{3,jn}(r)\}$ , for  $j = 1, 2, \dots, m$ . In this study,  $m = 10,000$ . Then the conditional mean estimates will be the ergodic mean of each recorded chain given by

$$\widehat{\theta}_{1,mn}(r) = \frac{1}{m} \sum_{j=1}^m \theta_{1,jn}(r), \quad \widehat{\theta}_{2,mn}(r) = \frac{1}{m} \sum_{j=1}^m \theta_{2,jn}(r) \quad \text{and} \quad \widehat{\theta}_{3,mn}(r) = \frac{1}{m} \sum_{j=1}^m \theta_{3,jn}(r).$$

Then we compute the absolute bias, standard deviation and mean squared errors using (4.1).

We also compute the overall absolute bias and standard deviation for  $\theta_2$  and  $\theta_3$  together as follows. Define  $\theta_{23} = (\theta_2, \theta_3)^\top$  and  $\widehat{\theta}_{23} = (\widehat{\theta}_{2,mn}, \widehat{\theta}_{3,mn})^\top$ .

$$\text{abs.bias}_{\theta_{23}} = \frac{1}{1000} \sum_{r=1}^{1000} \left| \widehat{\theta}_{2,mn}(r) - \theta_{20} \right| + \frac{1}{1000} \sum_{r=1}^{1000} \left| \widehat{\theta}_{3,mn}(r) - \theta_{30} \right|,$$

$$\text{std}_{\theta_{23}} = \sqrt{\sum_{i,j} \sigma_{ij}^2}, \quad \text{where } \sigma_{ij}^2 = \text{cov}(\theta_i, \theta_j), \quad \text{for } i, j = 2, 3,$$

$$\text{mse}_{\theta_{23}} = \frac{1}{1000} \sum_{r=1}^{1000} \left( \hat{\theta}_{2,mn}(r) - \theta_{20} \right)^2 + \frac{1}{1000} \sum_{r=1}^{1000} \left( \hat{\theta}_{3,mn}(r) - \theta_{30} \right)^2.$$

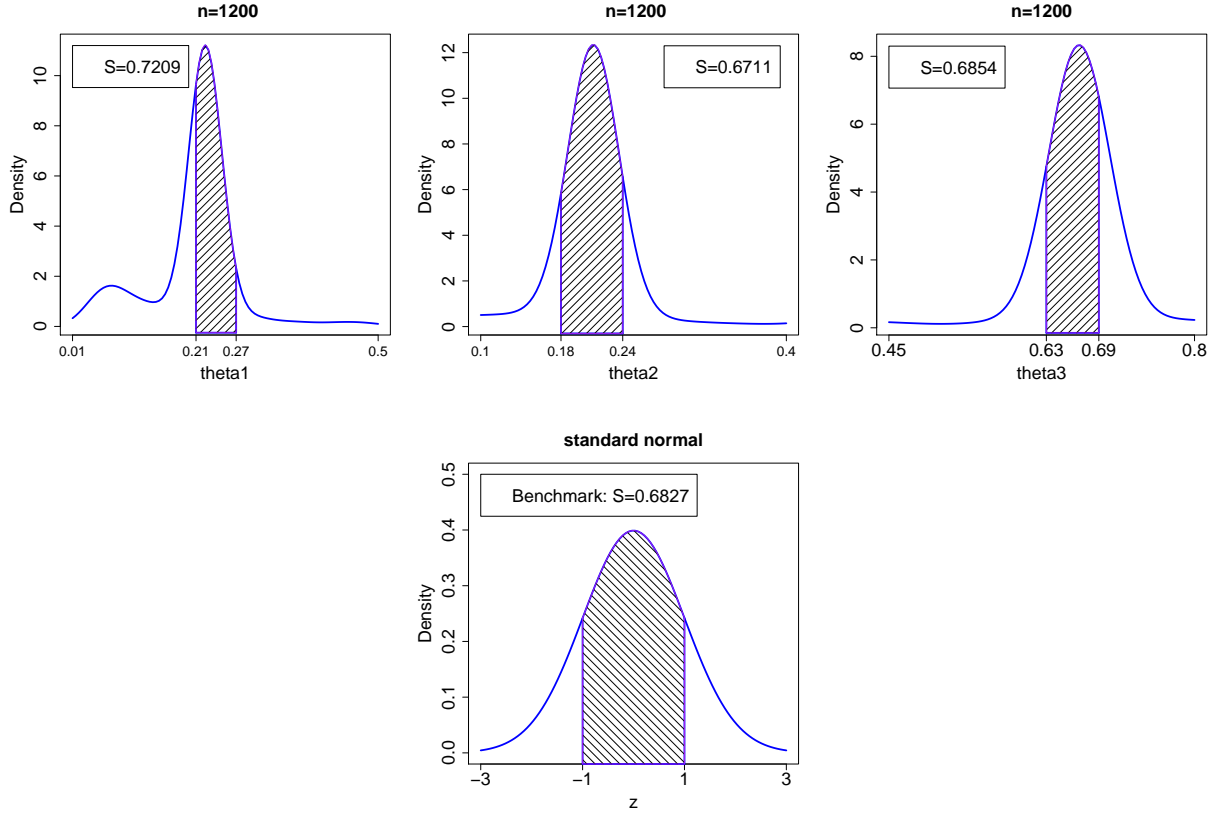
We also compare the performance of CME with that of maximum likelihood estimator of  $\theta$ . The results of parameter estimates are presented in Table 5. From Table 5, we find that with the increase of sample size, the absolute bias, standard deviation and mean squared errors of the proposed estimator decrease. This indicates that the proposed conditional mean estimator has some good finite sample performance. In addition, we find that CME has slightly better performance than MLE in terms of the absolute bias, standard deviation and mean squared errors.

Table 5: Absolute bias, standard deviation and mean squared error of CME and MLE for parameters in GARCH (1,1) model based on 1000 replications.

		CME				MLE			
	$n$	$\hat{\theta}_{1,mn}$	$\hat{\theta}_{2,mn}$	$\hat{\theta}_{3,mn}$	$\hat{\theta}_{23,mn}$	$\hat{\theta}_{1,n}$	$\hat{\theta}_{2,n}$	$\hat{\theta}_{3,n}$	$\hat{\theta}_{23,n}$
abs.bias	200	0.1698	0.0893	0.1501	0.2394	0.2456	0.1171	0.2030	0.3202
	600	0.1064	0.0614	0.0875	0.1489	0.1476	0.0717	0.1180	0.1896
	1200	0.0915	0.0414	0.0710	0.1123	0.1000	0.0584	0.0899	0.1483
std	200	0.1906	0.0899	0.1739	0.1240	0.3045	0.1402	0.2262	0.1861
	600	0.1239	0.0783	0.1191	0.0890	0.1891	0.0726	0.1312	0.1108
	1200	0.1116	0.0497	0.0870	0.0735	0.1263	0.0537	0.0910	0.0729
mse	200	0.0451	0.0109	0.0330	0.0439	0.1035	0.0226	0.0640	0.0866
	600	0.0176	0.0061	0.0141	0.0202	0.0399	0.0082	0.0248	0.0330
	1200	0.0133	0.0025	0.0077	0.0102	0.0180	0.0052	0.0131	0.0184

Based on the first replication, we produce the plots for  $\int_{C_{n1}} \hat{f}_n(\theta_1|T_n)d\theta_1$ ,  $\int_{C_{n2}} \hat{f}_n(\theta_2|T_n)d\theta_2$  and  $\int_{C_{n3}} \hat{f}_n(\theta_3|T_n)d\theta_3$  with sample size 1200 in Figure 5, where  $C_{ni} = \left\{ \theta_i : n^{1/2}|\hat{\theta}_{i,n} - \theta_i| \leq c \right\}$ , for  $i = 1, 2, 3$ . We can see that when sample size is large,  $\int_{C_{ni}} \hat{f}_n(\theta_i|T_n)d\theta_i$  is quite close to  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$  for the case of  $c = 1$ .

Figure 5: Plots of  $\int_{C_{ni}} \widehat{f}_n(\theta_i|T_n)d\theta_i$  (from left to right in the upper panel,  $i=1,2,3$ , respectively) and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$ .



We then consider the final simulated example, in which we consider the case where it is difficult to obtain an easily computable MLE, while it is easy to compute  $\widehat{\theta}_{mn}$ .

**Example 6.** Consider a mixture of two normal distributions with the density function given by  $f(x|\theta) = w_1\phi(x|\mu_1, \sigma_1^2) + w_2\phi(x|\mu_2, \sigma_2^2)$ . Let  $w_1 = w_2 = 0.5$  and  $\sigma_1^2 = \sigma_2^2 = 1$  for the two normal components, which means that we generate  $x$  from  $N(\mu_1, 1)$  with probability 0.5 and generate  $x$  from  $N(\mu_2, 1)$  with probability 0.5. So in one sample, we have  $n_1$  observations  $\{x_{1i}\}_{i=1}^{n_1}$  from  $N(\mu_1, 1)$  and  $n_2$  observations  $\{x_{2i}\}_{i=n_1+1}^{n_1+n_2}$  from  $N(\mu_2, 1)$ . Let  $n = n_1 + n_2$ . We generate  $\mu_1$  from  $N(-3, 1)$  and  $\mu_2$  from  $N(3, 1)$ . So the densities of  $\mu_1$  and  $\mu_2$  are given by

$$\pi(\mu_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mu_1 + 3)^2}{2}\right) \quad \text{and} \quad \pi(\mu_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mu_2 - 3)^2}{2}\right). \quad (4.22)$$

Denote  $\theta = (\mu_1, \mu_2)^\top$ . Define  $T_n = (T_{n1}, T_{n2})^\top$ , where  $T_{n1} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i}$  and  $T_{n2} = \frac{1}{n_2} \sum_{i=n_1+1}^n x_{2i}$ . Define  $U_n = (U_{n1}, U_{n2})^\top$ , where  $U_{n1} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i}^2$  and  $U_{n2} = \frac{1}{n_2} \sum_{i=n_1+1}^n x_{2i}^2$ . Then the likelihood function can be written as

$$L(\theta|T_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{n_1 U_{n1} + n_2 U_{n2}}{2}\right) \exp\left(-\frac{n_1 \mu_1^2 - 2\mu_1 n_1 T_{n1}}{2}\right) \exp\left(-\frac{n_2 \mu_2^2 - 2\mu_2 n_2 T_{n2}}{2}\right). \quad (4.23)$$

By definition, the conditional density of  $\theta$  given  $T_n$  is

$$f_n(\theta|T_n) = \frac{\pi(\theta)L(\theta|T_n)}{\int \pi(\theta)L(\theta|T_n)d\theta}. \quad (4.24)$$

Substituting (4.22) and (4.23) into (4.24), then we can obtain the conditional distribution of  $(\mu_1, \mu_2)$  given  $T_n$  as follows.

$$\mu_1|T_n \sim N\left(\frac{n_1T_{n1} - 3}{n_1 + 1}, \frac{1}{n_1 + 1}\right) \quad \text{and} \quad \mu_2|T_n \sim N\left(\frac{n_2T_{n2} + 3}{n_2 + 1}, \frac{1}{n_2 + 1}\right).$$

The conditional mean of  $\mu_1$  and  $\mu_2$  is then given by

$$\mathbb{E}[\mu_1|T_n] = \frac{n_1T_{n1} - 3}{n_1 + 1} \quad \text{and} \quad \mathbb{E}[\mu_2|T_n] = \frac{n_2T_{n2} + 3}{n_2 + 1}.$$

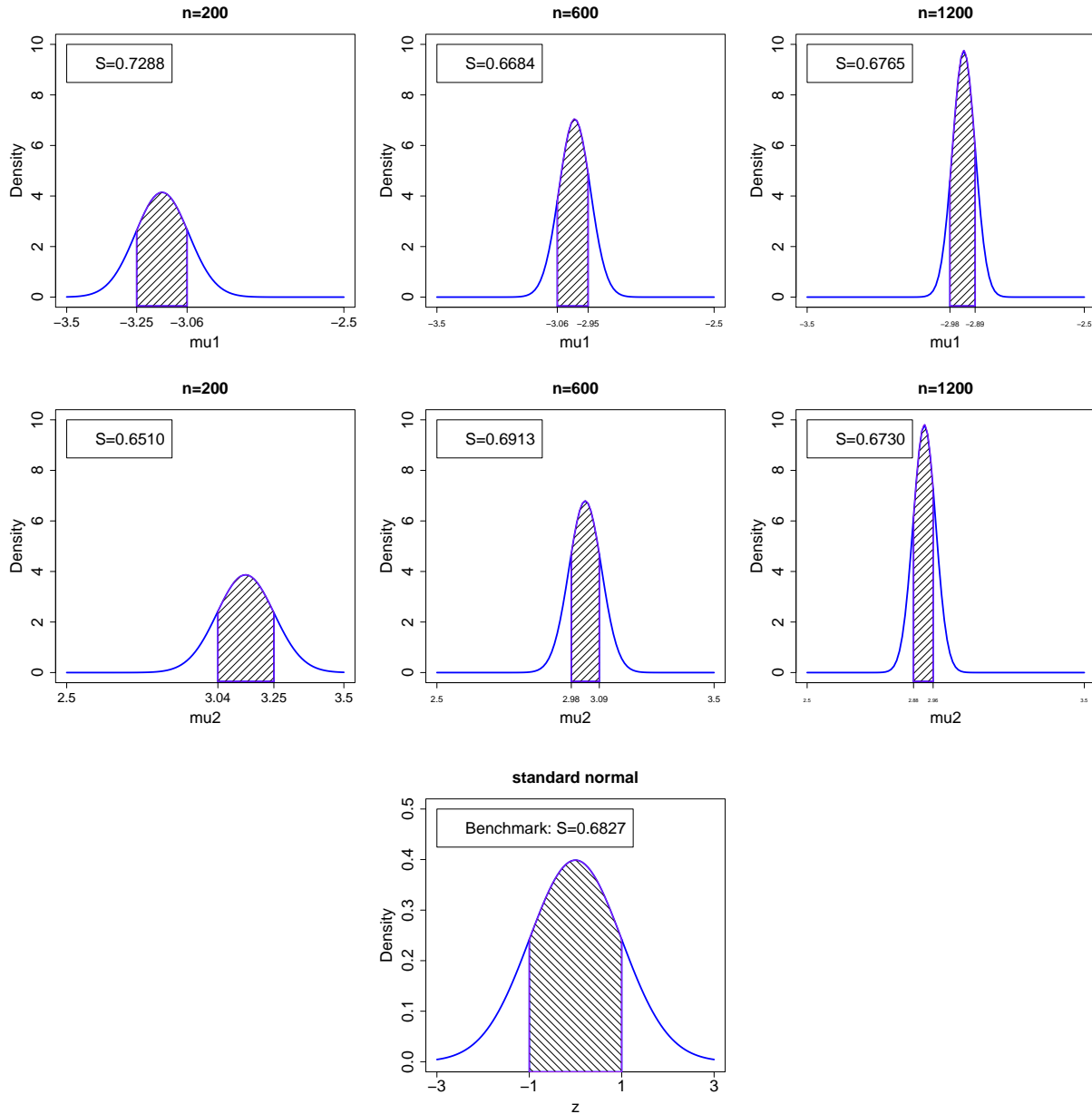
The maximum likelihood estimators of  $\mu_1$  and  $\mu_2$  are given by  $\hat{\mu}_{1n} = T_{n1}$  and  $\hat{\mu}_{2n} = T_{n2}$ . We then compute the absolute bias, standard deviation and mean squared errors for CME and MLE using (4.1). The results are presented in Table 6. From Table 6, we find that with the increase of sample size, the absolute bias, standard deviation and mean squared errors of the proposed estimator decrease. This indicates that the proposed conditional mean estimator has good finite sample performance. We also find that the CME and MLE have similar performance.

Table 6: Absolute bias, standard deviation and mean squared error of CME and MLE for parameters in a normal mixture density based on 1000 replications.

		CME		MLE	
$n$		$\mathbb{E}[\mu_1 T_n]$	$\mathbb{E}[\mu_2 T_n]$	$\hat{\mu}_{1n}$	$\hat{\mu}_{2n}$
abs.bias	200	0.1109	0.1043	0.1120	0.1053
	600	0.0625	0.0672	0.0627	0.0674
	1200	0.0442	0.0457	0.0443	0.0458
std	200	0.1386	0.1305	0.1400	0.1318
	600	0.0786	0.0836	0.0789	0.0839
	1200	0.0566	0.0567	0.0567	0.0568
mse	200	0.0192	0.0171	0.0196	0.0174
	600	0.0062	0.0070	0.0062	0.0071
	1200	0.0032	0.0032	0.0032	0.0032

For  $c = 1.0$  and  $i = 1, 2$ , recall  $C_{n\mu_i} = \left\{ \mu_i : \Delta_n^{1/2}(\hat{\mu}_{in})|\hat{\mu}_{in} - \mu_i| \leq c \right\}$ . Based on the first replication, we produce the plots to compare  $\int_{C_{n\mu_i}} f_n(\mu_i|T_n)d\mu_i$  with and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$  graphically in Figure 6.

Figure 6: Plots of  $\int_{C_{n\mu_1}} f_n(\mu_1|T_n)d\mu_1$  (the upper panel with sample sizes  $n=200, 600, 1200$ ),  $\int_{C_{n\mu_2}} f_n(\mu_2|T_n)d\mu_2$  (the middle panel with sample sizes  $n=200, 600, 1200$ ) and  $\frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{1}{2}u^2} du$ .



## 5 Application

In this section, we use our conditional mean estimator to estimate a GARCH (1,1) model for S&P 500 daily stock return. We downloaded the S&P 500 daily closing prices,  $p_t$ , during the period from 2 January 2002 to the 9 December 2009 from <http://finance.yahoo.com>. The date  $t$  return is calculated as  $r_t = \log(p_t/p_{t-1})$ , and there are  $T = 2000$  observations of the return. The time series plot of the return series is presented in Figure 7.

We fit a GARCH (1,1) model for the demeaned log-returns  $\{y_t\}$  given as follows, where  $y_t = r_t - \frac{1}{T} \sum_{t=1}^T r_t$ ,  $y_t = e_t \sigma_t$  and  $\sigma_t^2 = \theta_1 + \theta_2 y_{t-1}^2 + \theta_3 \sigma_{t-1}^2$ , in which  $\theta_1 > 0$ ,  $\theta_2, \theta_3 \geq 0$ ,  $\theta_2 + \theta_3 < 1$ . Note that the restrictions on the GARCH parameters  $\theta_1, \theta_2$  and  $\theta_3$  guarantee the positivity of the conditional variance and that the log-return series is covariance stationary.

Let  $\theta = (\theta_1, \theta_2, \theta_3)^\top$ . By using the quasi-maximum likelihood estimation method, we obtain the estimates of  $\theta$ ,  $\hat{\theta}_n = (0.0000, 0.0651, 0.9282)^\top$ . Use the estimation procedure outlined in Example 5 in the simulation studies, we obtain the conditional mean estimates of  $\theta$ ,  $\hat{\theta}_{mn} = (0.0166, 0.069, 0.7733)^\top$ . As a by-product of the conditional mean estimation, we produce the kernel density estimates of the conditional density of  $\theta_1, \theta_2$  and  $\theta_3$  in Figure 8.

Figure 7: Time series plot of S&P 500 daily returns.

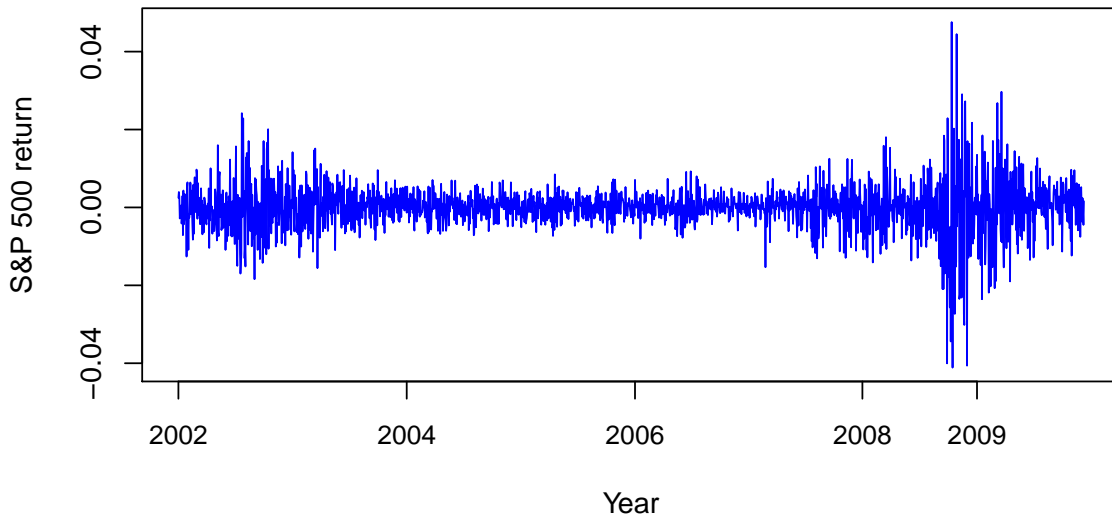




Figure 8: Kernel density estimates of the conditional density function of  $\theta$  given  $T_n$ . From left to right, we have  $f(\theta_1|T_n)$ ,  $f(\theta_2|T_n)$  and  $f(\theta_3|T_n)$ .

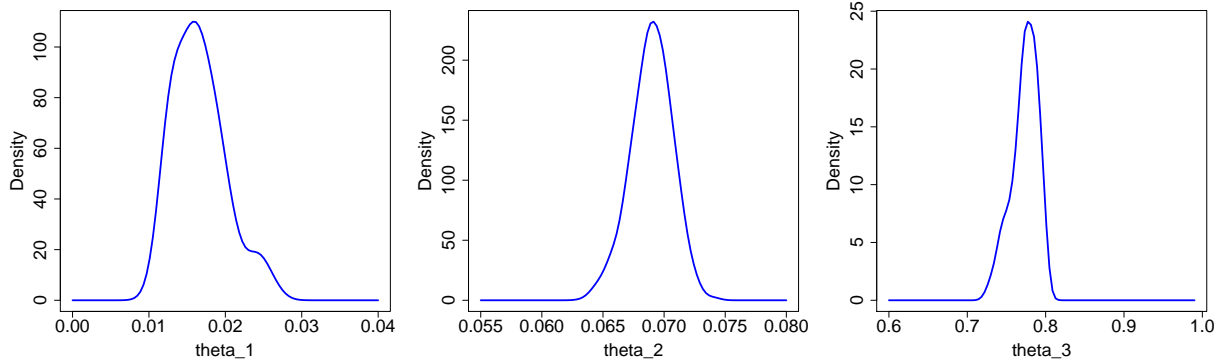
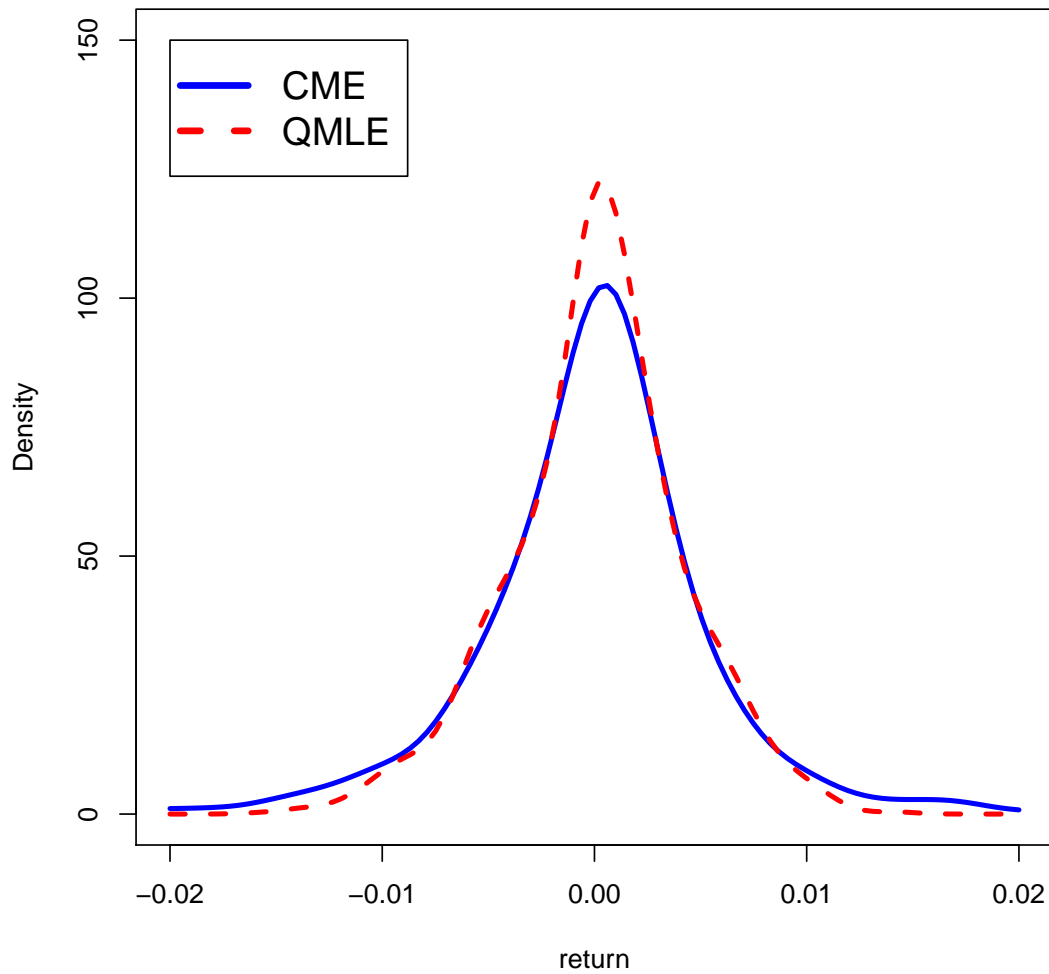


Figure 9: Kernel density estimates of return on 10 December 2009 with CME and QMLE.



Based on the parameter estimates  $\widehat{\theta}_{mn}$  and  $\widehat{\theta}_n$ , we estimate the density of the return series on 10 December 2009 using kernel estimation method:

$$\begin{aligned}\widehat{f}(y|\widehat{\theta}_{mn}) &= \frac{1}{Th\widehat{\sigma}_{T+1}(\widehat{\theta}_{mn})} \sum_{t=1}^T K\left(\frac{y/\widehat{\sigma}_{T+1}(\widehat{\theta}_{mn}) - y_t/\widehat{\sigma}_t(\widehat{\theta}_{mn})}{h}\right) \\ \widehat{f}(y|\widehat{\theta}_n) &= \frac{1}{Th\widehat{\sigma}_{T+1}(\widehat{\theta}_n)} \sum_{t=1}^T K\left(\frac{y/\widehat{\sigma}_{T+1}(\widehat{\theta}_n) - y_t/\widehat{\sigma}_t(\widehat{\theta}_n)}{h}\right),\end{aligned}$$

where  $K(\cdot)$  is chosen to be the Gaussian kernel function and  $h$  is selected by normal reference rule. The density estimates are displayed in Figure 9, from which we can find that  $\widehat{f}(y|\widehat{\theta}_{mn})$  and  $\widehat{f}(y|\widehat{\theta}_n)$  have different characteristics at the peak and tail areas.

As in practice, the true data generating process is unknown, we cannot tell which method is better just based on these estimation results. In the following discussion, we evaluate our conditional mean method and the quasi-maximum likelihood estimation method based on the performance of out-of-sample density forecasting and squared return forecasting.

For density forecasting, we employ the scoring rule introduced by [Amisano and Giacomini \(2007\)](#). Using this scoring rule, we are able to decide which method performs better. We conducted a rolling-sample procedure to evaluate the performance of each estimation method. Let  $T$  denote the number of the observations, and let  $y_t$  denote return at day  $t$ , for  $t = 1, 2, \dots, T$ . The first sample contains the first  $n$  observations,  $y_1, y_2, \dots, y_n$ , and is used to obtain the parameter estimates, based on which, we can obtain the estimated variance at time  $n + 1$ ,  $\widehat{\sigma}_{n+1}$ . As the error distribution is unknown, here we forecast the density of  $y_{n+1}$  by the kernel density estimation method:

$$\widehat{f}(y_{n+1}|\widehat{\theta}) = \frac{1}{nh\widehat{\sigma}_{n+1}(\widehat{\theta})} \sum_{t=1}^n K\left(\frac{y_{n+1}/\widehat{\sigma}_{n+1}(\widehat{\theta}) - y_t/\widehat{\sigma}_t(\widehat{\theta})}{h}\right),$$

where  $K(\cdot)$  is chosen to be the Gaussian kernel function and  $h$  is selected by normal reference rule.

The second sample contains  $y_2, y_3, \dots, y_{n+1}$ , which are obtained by rolling the first sample forward for one step. Using this sample, we repeat what was done based on the previous sample and forecast the density of  $y_{n+2}$ . This rolling procedure continues until the density of  $y_T$  is forecasted.

We calculated the average likelihood scores over the out-of-sample period:

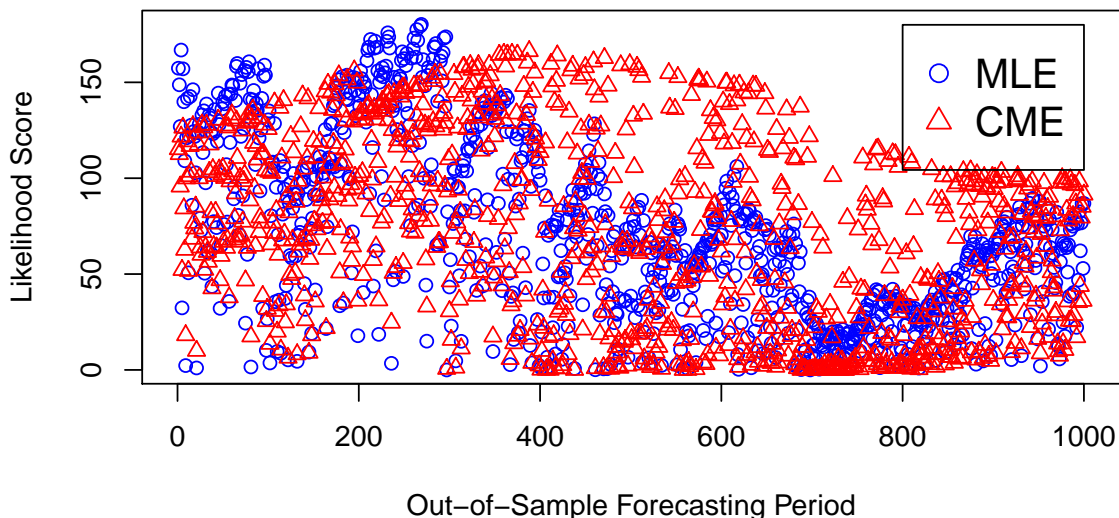
$$S_n = \frac{1}{T-n} \sum_{r=1}^{T-n} \widehat{f}(y_{n+r}|\widehat{\theta}_n^{(r)}), \quad S_{mn} = \frac{1}{T-n} \sum_{r=1}^{T-n} \widehat{f}(y_{n+r}|\widehat{\theta}_{mn}^{(r)}),$$

where  $\widehat{\theta}_n^{(r)} = \left(\widehat{\theta}_{1,n}^{(r)}, \widehat{\theta}_{2,n}^{(r)}, \widehat{\theta}_{3,n}^{(r)}\right)^\top$  and  $\widehat{\theta}_{mn}^{(r)} = \left(\widehat{\theta}_{1,mn}^{(r)}, \widehat{\theta}_{2,mn}^{(r)}, \widehat{\theta}_{3,mn}^{(r)}\right)^\top$  are the parameter estimates

based on the  $r$ -th rolling sample. In terms of the average likelihood score, the larger it is, the better the corresponding estimation method performs.

The number of the observations is  $T = 2000$ , and the size of a rolling sample is  $n = 1000$ , which implies that we have 1000 rolling samples in total. By 1000 rolling samples, we obtained  $S_n = 69.98$  and  $S_{mn} = 71.33$ , which means that the forecast with the use of our conditional mean estimator leads to a slightly higher score. To see this more clearly, we plot the likelihood score ( $\hat{f}(y_{n+r}|\hat{\theta}^{(r)})$ ) in different out-of-sample period ( $r$ ) with both methods in Figure 10. From this plot, we can see that most of the time, the likelihood score with CME is higher than that with MLE. This implies that the use of our conditional mean estimator outperforms that of the maximum likelihood estimation method in terms of density forecasting.

Figure 10: Out-of-sample likelihood score with MLE and CME.



Using the same rolling-sample procedure, we also evaluate the performance of squared return forecasting. We compute the mean absolute error (MAE) and mean squared error (MSE) for 1-step ahead squared return using conditional mean estimator as follows:

$$\text{MAE}(\hat{\theta}_{mn}) = \frac{1}{T-n} \sum_{t=1}^{T-n} \left| y_{n+r}^2 - \hat{\sigma}_{n+r}^2(\hat{\theta}_{mn}) \right|,$$

$$\text{MSE}(\hat{\theta}_{mn}) = \frac{1}{T-n} \sum_{t=1}^{T-n} \left[ y_{n+r}^2 - \hat{\sigma}_{n+r}^2(\hat{\theta}_{mn}) \right]^2.$$

Similarly, we define MAE and MSE with parameter estimated by quasi-maximum likelihood

estimation method:

$$\begin{aligned} \text{MAE}(\hat{\theta}_n) &= \frac{1}{T-n} \sum_{t=1}^{T-n} \left| y_{n+r}^2 - \hat{\sigma}_{n+r}^2(\hat{\theta}_n) \right|, \\ \text{MSE}(\hat{\theta}_n) &= \frac{1}{T-n} \sum_{t=1}^{T-n} \left[ y_{n+r}^2 - \hat{\sigma}_{n+r}^2(\hat{\theta}_n) \right]^2. \end{aligned}$$

We obtain that

$$\text{MAE}(\hat{\theta}_{mn}) = 1.8366 \times 10^{-5}, \quad \text{MSE}(\hat{\theta}_{mn}) = 1.4409 \times 10^{-8}$$

and

$$\text{MAE}(\hat{\theta}_n) = 5.3789 \times 10^{-5}, \quad \text{MSE}(\hat{\theta}_n) = 2.1873 \times 10^{-8}.$$

Given the smaller MAE and MSE results for the conditional mean estimator, we can conclude that conditional mean has better performance in squared return forecasting. Therefore, we can conclude that the conditional mean estimator performs better than the quasi-maximum likelihood estimator in terms of the out-of-sample density forecasting and squared return forecasting.

## 6 Conclusions

This paper has established some new asymptotic properties for Bayesian estimators. The simulation studies have shown that our proposed conditional mean estimators have very good finite sample performance. We have used the conditional mean estimation method to estimate a GARCH(1,1) model for S&P 500 daily returns and we found that compared with quasi-maximum likelihood estimation method, our conditional mean estimator has better out-of-sample forecasting performance.

## 7 Acknowledgements

The second author would like to specially thank Han Hong for various constructive discussions about this topic. Acknowledgements also go to many participants of several seminars and conferences in the last few years, particularly to Rong Chen, Jianqing Fan, David Frazier, Robert Kohn, Lung-Fei Lee, Oliver Linton, Gael Martin, Adrian Pagan, Anastasios Panagiotelis, Scott Sisson, Yundong Tu, Aman Ullah and Xibin Zhang, on earlier versions of this paper. The first and the second authors would also like to acknowledge financial support by the Australian Research Council Discovery Grants Program under Grant Numbers: DP150101012 & DP170104421.

# Appendix

This appendix gives the full proofs of Theorems 1 and 2. For notational simplicity, we only consider the case where the dimensionality of  $\theta$  is  $d = 1$ .

## Proof of Theorem 1

Let  $p_n(T_n) = \int e^{l_n(\theta)} \pi(\theta) d\theta$ . Let us first deal with  $p_n(T_n)$ . Observe that

$$\begin{aligned} p_n(T_n) &= \int e^{l_n(\theta)} \pi(\theta) d\theta = e^{l_n(\hat{\theta}_n)} \int e^{l_n(\theta) - l_n(\hat{\theta}_n)} \pi(\theta) d\theta \\ &= e^{l_n(\hat{\theta}_n)} \int_{C_n} e^{l_n(\theta) - l_n(\hat{\theta}_n)} \pi(\theta) d\theta + e^{l_n(\hat{\theta}_n)} \int_{D_n} e^{l_n(\theta) - l_n(\hat{\theta}_n)} \pi(\theta) d\theta \\ &\equiv p_{1n}(T_n) + p_{2n}(T_n), \end{aligned} \tag{A.1}$$

where  $C_n = \{\theta : \Delta_n^{1/2}(\hat{\theta}_n) \|\theta - \hat{\theta}_n\| \leq c\}$  and  $D_n = \{\theta : \Delta_n^{1/2}(\hat{\theta}_n) \|\theta - \hat{\theta}_n\| > c\}$  for any given  $c > 0$ .

For  $\theta \in C_n$ , we can have the following Taylor expansion:

$$l_n(\theta) - l_n(\hat{\theta}_n) = l_n^{(1)}(\hat{\theta}_n) (\theta - \hat{\theta}_n) + \frac{1}{2} l_n^{(2)}(\hat{\theta}_n^*) (\theta - \hat{\theta}_n)^2 = -\frac{1}{2} \Delta_n(\hat{\theta}_n^*) (\theta - \hat{\theta}_n)^2, \tag{A.2}$$

using the fact that  $l_n^{(1)}(\hat{\theta}_n) = 0$ , where  $\hat{\theta}_n^*$  is between  $\theta$  and  $\hat{\theta}_n$ ,  $l_n^{(i)}(\cdot)$ , for  $i = 1, 2$ , denote the first and second derivatives of  $l_n(\cdot)$ , and for each given  $n$ , both exist and  $\Delta_n(\theta)$  is continuous in  $\theta$  under Assumption 1.

For  $p_{1n}(T_n)$ , in view of (A.2), using Assumptions 1 and 3(i), we have

$$\begin{aligned} p_{1n}(T_n) &= e^{l_n(\hat{\theta}_n)} \int_{C_n} e^{l_n(\theta) - l_n(\hat{\theta}_n)} \pi(\theta) d\theta = (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \int_{C_n} e^{-\frac{1}{2}(\theta - \hat{\theta}_n)^\top \Delta_n(\hat{\theta}_n^*) (\theta - \hat{\theta}_n)} \pi(\theta) d\theta \\ &= (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \int_{|y| \leq c} e^{-\frac{y^2}{2}} \pi\left(\hat{\theta}_n + \Delta_n^{-1/2}(\hat{\theta}_n) y\right) dy \\ &= (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \pi\left(\hat{\theta}_n\right) \int_{|y| \leq c} e^{-\frac{y^2}{2}} dy \\ &\quad + \frac{1}{2} (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-3/2}(\hat{\theta}_n) \pi^{(2)}\left(\hat{\theta}_n\right) \int_{|y| \leq c} y^2 e^{-\frac{y^2}{2}} dy. \end{aligned} \tag{A.3}$$

Let  $E_n = \{\theta : \Delta_n^{1/2}(\hat{\theta}_n) \|\theta - \hat{\theta}_n\| \leq \lambda\}$  and  $F_n = \{\theta : \Delta_n^{1/2}(\hat{\theta}_n) \|\theta - \hat{\theta}_n\| > \lambda\}$  for large enough  $\lambda > 0$ . Under Assumption 3(i), similarly to the derivations in (A.1), we have

$$\begin{aligned} p_n(T_n) &= e^{l_n(\hat{\theta}_n)} \int_{E_n} e^{l_n(\theta) - l_n(\hat{\theta}_n)} \pi(\theta) d\theta + e^{l_n(\hat{\theta}_n)} \int_{F_n} e^{l_n(\theta) - l_n(\hat{\theta}_n)} \pi(\theta) d\theta \\ &= (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \pi\left(\hat{\theta}_n\right) \int_{|y| \leq \lambda} e^{-\frac{y^2}{2}} dy \\ &= (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \pi\left(\hat{\theta}_n\right) \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \\ &= (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \pi\left(\hat{\theta}_n\right) \sqrt{2\pi} \end{aligned} \tag{A.4}$$

by letting  $\lambda \rightarrow \infty$ .

Equations (A.3) and (A.4) then imply

$$\begin{aligned}
& \Delta_n(\widehat{\theta}_n) \left( \int_{C_n} f_n(\theta|T_n)d\theta - \frac{1}{\sqrt{2\pi}} \int_{|y|\leq c} e^{-\frac{y^2}{2}} dy \right) \\
& - \frac{\pi^{(2)}(\widehat{\theta}_n) (1 + o_P(1))}{2\pi(\widehat{\theta}_n)} \frac{1}{\sqrt{2\pi}} \int_{|y|\leq c} y^2 e^{-\frac{y^2}{2}} dy \\
& = \Delta_n(\widehat{\theta}_n) \left( \int_{C_n} f_n(\theta|T_n)d\theta - \frac{1}{\sqrt{2\pi}} \int_{|y|\leq c} e^{-\frac{y^2}{2}} dy \right) \\
& - (1 + o_P(1)) \frac{\pi^{(2)}(\theta_0)}{2\pi(\theta_0)} \frac{1}{\sqrt{2\pi}} \int_{|y|\leq c} y^2 e^{-\frac{y^2}{2}} dy \rightarrow_P 0,
\end{aligned} \tag{A.5}$$

which completes the proof of Theorem 1.

### Proof of Theorem 2

Recall that we define  $\widehat{\theta}_{mn} = \frac{1}{m} \sum_{j=1}^m \theta_{jn}$ , in which  $\theta_{jn}$  denotes the  $j$ th posterior draw and  $m$  denotes the number of Markov chain Monte Carlo (MCMC) iterations. It is assumed here that conditioning on  $T_n$ ,  $\{\theta_{jn}\}$  is a sequence of Harris ergodic Markov chains.

In order to show the properties of  $\widehat{\theta}_{mn}$ , we proceed with the following three steps.

1. First, by Assumption 1(ii) and using an existing result (see, for example, Corollary 2 of Jones (2004)), we have  $\frac{1}{\sqrt{m}\sigma_n} \sum_{j=1}^m (\theta_{jn} - \mathbb{E}[\theta|T_n]) \rightarrow_D N(0, 1)$  as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ , where  $\sigma_n^2$  is the same as defined in Assumption 2.
2. Second, we show that  $\Delta_n^{1/2}(\widehat{\theta}_n) (\mathbb{E}[\theta|T_n] - \theta_0) \rightarrow_D N(0, 1)$ , as  $n \rightarrow \infty$ .
3. Third, we combine the previous two steps to show that  $\Delta_n^{1/2}(\widehat{\theta}_n) \left( \frac{1}{m} \sum_{j=1}^m \theta_{jn} - \theta_0 \right) \rightarrow_D N(0, 1)$ , as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ .
4. Lastly, we show Theorem 2.1(i) as a consequence.

By definition, we have that

$$\mathbb{E}[\theta|T_n] = \int \theta f_n(\theta|T_n)d\theta = \frac{\int \theta e^{l_n(\theta)} \pi(\theta)d\theta}{\int e^{l_n(\theta)} \pi(\theta)d\theta} = \frac{q_n(T_n)}{p_n(T_n)},$$

where  $q_n(T_n) = \int \theta e^{l_n(\theta)} \pi(\theta)d\theta$  and  $p_n(T_n) = \int e^{l_n(\theta)} \pi(\theta)d\theta$ .

Define  $q_n(T_n) = \int \theta e^{l_n(\theta)} \pi(\theta)d\theta$ . Similarly to equation (A.4), we can write

$$\begin{aligned}
q_n(T_n) &= \int_{E_n} \theta e^{l_n(\theta)} \pi(\theta)d\theta + \int_{F_n} \theta e^{l_n(\theta)} \pi(\theta)d\theta \\
&= (1 + o_P(1)) \int_{E_n} \theta e^{l_n(\widehat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\widehat{\theta}_n)(\theta - \widehat{\theta}_n)^2} \pi(\theta)d\theta + \int_{F_n} \theta e^{l_n(\theta)} \pi(\theta)d\theta \\
&= (1 + o_P(1)) q_{1n}(T_n) + q_{2n}(T_n),
\end{aligned} \tag{A.6}$$

where  $q_{1n}(T_n) = \int_{E_n} \theta e^{l_n(\hat{\theta}_n)} e^{-\frac{1}{2}\Delta_n(\hat{\theta}_n)(\theta - \hat{\theta}_n)^2} \pi(\theta) d\theta$  and  $q_{2n}(T_n) = \int_{E_n} \theta e^{l_n(\theta)} \pi(\theta) d\theta$ .

In a similar way to equation (A.3), we also have

$$\begin{aligned} q_{1n}(T_n) &= e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \int_{|y| \leq \lambda} \left( \hat{\theta}_n + \Delta_n^{-1/2}(\hat{\theta}_n) y \right) \pi \left( \hat{\theta}_n + \Delta_n^{-1/2}(\hat{\theta}_n) y \right) e^{-y^2/2} dy \\ &= (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-1/2}(\hat{\theta}_n) \hat{\theta}_n \pi \left( \hat{\theta}_n \right) \int_{|y| \leq \lambda} e^{-y^2/2} dy \\ &\quad + (1 + o_P(1)) e^{l_n(\hat{\theta}_n)} \Delta_n^{-3/2}(\hat{\theta}_n) \pi^{(1)} \left( \hat{\theta}_n \right) \int_{|y| \leq \lambda} y^2 e^{-y^2/2} dy, \end{aligned} \quad (\text{A.7})$$

which, along with equation (A.4), implies that as  $n \rightarrow \infty$

$$\begin{aligned} \Delta_n^{1/2}(\hat{\theta}_n) \left( \mathbb{E}[\theta|T_n] - \hat{\theta}_n \right) &= \Delta_n^{1/2}(\hat{\theta}_n) \left( \int_{E_n} \theta f_n(\theta|T_n) d\theta - \hat{\theta}_n \int_{E_n} f_n(\theta|T_n) d\theta \right) \\ &\quad - \Delta_n^{1/2}(\hat{\theta}_n) \hat{\theta}_n \int_{F_n} f_n(\theta|T_n) d\theta + \Delta_n^{1/2}(\hat{\theta}_n) \int_{F_n} \theta f_n(\theta|T_n) d\theta \\ &= (1 + o_P(1)) \Delta_n^{-1/2}(\hat{\theta}_n) \frac{\pi^{(1)}(\hat{\theta}_n)}{\pi(\hat{\theta}_n)} \cdot \frac{1}{\sqrt{2\pi}} \int_{|y| \leq \lambda} y^2 e^{-y^2/2} dy \\ &\quad - \Delta_n^{1/2}(\hat{\theta}_n) \hat{\theta}_n \int_{F_n} f_n(\theta|T_n) d\theta + \Delta_n^{1/2}(\hat{\theta}_n) \int_{F_n} \theta f_n(\theta|T_n) d\theta \rightarrow_P 0 \end{aligned} \quad (\text{A.8})$$

by Assumption 3. Furthermore, we have as  $\lambda \rightarrow \infty$

$$\Delta_n(\hat{\theta}_n) \left( \mathbb{E}[\theta|T_n] - \hat{\theta}_n \right) = (1 + o_P(1)) \frac{\pi^{(1)}(\theta_0)}{\pi(\theta_0)} \cdot \frac{1}{\sqrt{2\pi}} \int_{|y| \leq \lambda} y^2 e^{-y^2/2} dy \rightarrow_P \frac{\pi^{(1)}(\theta_0)}{\pi(\theta_0)} \quad (\text{A.9})$$

whenever  $\pi(\theta_0) \neq 0$ , if both  $\Delta_n(\hat{\theta}_n) \int_{F_n} f_n(\theta|T_n) d\theta = o_P(1)$  and  $\Delta_n(\hat{\theta}_n) \int_{F_n} \theta f_n(\theta|T_n) d\theta = o_P(1)$ .

Equation (A.8), along with Assumption 4, implies

$$\begin{aligned} \Delta_n^{1/2}(\hat{\theta}_n) \left( \mathbb{E}[\theta|T_n] - \theta_0 \right) &= \Delta_n^{1/2}(\hat{\theta}_n) \left( \mathbb{E}[\theta|T_n] - \hat{\theta}_n \right) + \Delta_n^{1/2}(\hat{\theta}_n) \left( \hat{\theta}_n - \theta_0 \right) \\ &= \Delta_n^{1/2}(\hat{\theta}_n) \left( \hat{\theta}_n - \theta_0 \right) + o_P(1) \rightarrow_D N(0, 1). \end{aligned} \quad (\text{A.10})$$

Meanwhile, Assumption 2, along with an application of an existing central limit theorem for stationary Markov chains (see, for example, Corollary 2 of Jones (2004)), implies that as  $m \rightarrow \infty$

$$\frac{1}{\sqrt{m} \sigma_n} \sum_{j=1}^m (\theta_{jn} - \mathbb{E}[\theta|T_n]) \rightarrow_D N(0, 1), \quad (\text{A.11})$$

which completes the proof of the first part of Theorem 2(ii).

Therefore, under Assumptions 1–4, as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ , we have that

$$\begin{aligned} \Delta_n^{1/2}(\hat{\theta}_n) \left( \frac{1}{m} \sum_{j=1}^m \theta_{jn} - \theta_0 \right) &= \Delta_n^{1/2}(\hat{\theta}_n) \left( \frac{1}{m} \sum_{j=1}^m \theta_{jn} - \mathbb{E}[\theta|T_n] + \mathbb{E}[\theta|T_n] - \theta_0 \right) \\ &= \Delta_n^{1/2}(\hat{\theta}_n) \left( \frac{1}{m} \sum_{j=1}^m \theta_{jn} - \mathbb{E}[\theta|T_n] \right) + \Delta_n^{1/2}(\hat{\theta}_n) \left( \mathbb{E}[\theta|T_n] - \theta_0 \right) \end{aligned}$$

$$= \frac{\Delta_n^{1/2}(\hat{\theta}_n) \sigma_n}{\sqrt{m}} \left( \frac{1}{\sqrt{m} \sigma_n} \sum_{j=1}^m (\theta_{jn} - \mathbb{E}[\theta|T_n]) \right) + \Delta_n^{1/2}(\hat{\theta}_n) (\mathbb{E}[\theta|T_n] - \theta_0) \rightarrow_D N(0, 1),$$

which completes the proof of equation (3.3) of Theorem 2(ii).

In addition, equations (A.11) and (A.8) also show that as  $m \rightarrow \infty$  and  $n \rightarrow \infty$

$$\begin{aligned} \Delta_n^{1/2}(\hat{\theta}_n) (\hat{\theta}_{mn} - \hat{\theta}_n) &= \Delta_n^{1/2}(\hat{\theta}_n) (\hat{\theta}_{mn} - \mathbb{E}[\theta|T_n]) + \Delta_n^{1/2}(\hat{\theta}_n) (\mathbb{E}[\theta|T_n] - \hat{\theta}_n) \\ &= \frac{\Delta_n^{1/2}(\hat{\theta}_n) \sigma_n}{\sqrt{m}} \left( \frac{1}{\sqrt{m} \sigma_n} \sum_{j=1}^m (\theta_{jn} - \mathbb{E}[\theta|T_n]) \right) + \Delta_n^{1/2}(\hat{\theta}_n) (\mathbb{E}[\theta|T_n] - \hat{\theta}_n) \\ &\rightarrow_P 0, \end{aligned} \tag{A.12}$$

which completes the proof of the first part of Theorem 2(i).

Using equation (A.9), we have as  $m \rightarrow \infty$  and  $n \rightarrow \infty$

$$\begin{aligned} \Delta_n(\hat{\theta}_n) (\hat{\theta}_{mn} - \hat{\theta}_n) &= \Delta_n(\hat{\theta}_n) (\hat{\theta}_{mn} - \mathbb{E}[\theta|T_n]) + \Delta_n(\hat{\theta}_n) (\mathbb{E}[\theta|T_n] - \hat{\theta}_n) \\ &= \frac{\Delta_n(\hat{\theta}_n) \sigma_n}{\sqrt{m}} \left( \frac{1}{\sqrt{m} \sigma_n} \sum_{j=1}^m (\theta_{jn} - \mathbb{E}[\theta|T_n]) \right) + \Delta_n(\hat{\theta}_n) (\mathbb{E}[\theta|T_n] - \hat{\theta}_n) \\ &= (1 + o_P(1)) \frac{\pi^{(1)}(\theta_0)}{\pi(\theta_0)} \cdot \frac{1}{\sqrt{2\pi}} \int_{|y| \leq \lambda} y^2 e^{-y^2/2} dy \rightarrow_P \frac{\pi^{(1)}(\theta_0)}{\pi(\theta_0)} \end{aligned} \tag{A.13}$$

whenever  $\pi(\theta_0) \neq 0$  and  $\frac{\Delta_n(\hat{\theta}_n) \sigma_n}{\sqrt{m}} \rightarrow_P 0$  and as  $\lambda \rightarrow \infty$ . Equation (A.13) completes the proof of the second part of Theorem 2(i).

## References

- Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Boston.
- Amisano, G. and Giacomini, R. (2007), ‘Comparing density forecasts via weighted likelihood ratio tests’, *Journal of Business & Economic Statistics* **25**(2), 177–190.
- Beaumont, M. A., Zhang, W. and Balding, D. J. (2002), ‘Approximate Bayesian computation in population genetics’, *Genetics* **162**(4), 2025–2035.
- Chen, C.-F. (1985), ‘On asymptotic normality of limiting density functions with Bayesian implications’, *Journal of the Royal Statistical Society. Series B (Methodological)* **47**(3), 540–546.
- Chernozhukov, V. and Hong, H. (2003), ‘An MCMC approach to classical estimation’, *Journal of Econometrics* **115**(2), 293–346.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series*, John Wiley, New York.
- Gao, J. and Hong, H. (2014), A Computational Implementation of GMM, Working paper, Monash University and Stanford University.
- URL:** <http://ssrn.com/abstract=2503199>



- Jones, G. L. (2004), ‘On the Markov chain central limit theorem’, *Probability Surveys* **1**, 299–320.
- Kim, J.-Y. (1998), ‘Large sample properties of posterior densities, Bayesian information criterion and the likelihood principle in nonstationary time series models’, *Econometrica* **66**(2), 359–380.
- Ling, S. and Li, W. K. (1998), ‘Limiting distributions of maximum likelihood estimators for unstable arma time series with general arch errors’, *Annals of Statistics* **26**, 84–125.
- Ma, J. (2008), A closed–form asymptotic variance–covariance matrix for the quasi–maximum likelihood estimator of the GARCH(1,1) model, Working paper, University of Alabama.  
**URL:** <http://ssrn.com/abstract=889461>
- Phillips, P. C. B. (1996), ‘Econometric model determination’, *Econometrica* **64**(3), 763–812.
- Phillips, P. C. B. and Ploberger, W. (1996), ‘An asymptotic theory of Bayesian inference for time series’, *Econometrica* **64**(2), 381–412.
- Walker, A. (1969), ‘On the asymptotic behaviour of posterior distributions’, *Journal of the Royal Statistical Society. Series B (Methodological)* **31**(1), 80–88.