ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Nonparametric Regression Approach to Bayesian Estimation

Jiti Gao and Han Hong

November 2014

Working Paper 25/14

Nonparametric Regression Approach to Bayesian Estimation [1]

Jiti Gao and Han Hong

Monash University and Stanford University

Abstract

Estimation of unknown parameters and functions involved in complex nonlinear econometric models is a very important issue. Existing estimation methods include generalised method of moments (GMM) by Hansen (1982) and others, efficient method of moments (EMM) by Gallant and Tauchen (1997), Markov chain Monte Carlo (MCMC) method by Chernozhukov and Hong (2003), and nonparametric simulated maximum likelihood estimation (NSMLE) method by Creel and Kristensen (2011), and Kristensen and Shin (2012). Except the NSMLE method, other existing methods do not provide closed–form solutions. This paper proposes non– and semi–parametric based closed–form approximations to the estimation and computation of posterior means involved in complex nonlinear econometric models. We first consider the case where the samples can be independently drawn from both the likelihood function and the prior density. The samples and observations are then used to nonparametrically estimate posterior mean functions. The estimation method is also applied to estimate the posterior mean of the parameter–of–interest on a summary statistic. Both the asymptotic theory and the finite sample study show that the nonparametric estimate of this posterior mean is superior to existing estimates, including the conventional sample mean.

This paper then proposes some non– and semi–parametric dimension reductions methods to deal with the case where the dimensionality of either the regressors or the summary statistics is large. Meanwhile, the paper develops a nonparametric estimation method for the case where the samples are obtained from using a resampling algorithm. The asymptotic theory shows that in each case the rate of convergence of the nonparametric estimate based on the resamples is faster than that of the conventional nonparametric estimation method by an order of the number of the resamples. The proposed models and estimation methods are evaluated through using simulated and empirical examples. Both the simulated and empirical examples show that the proposed nonparametric estimation based on resamples outperforms existing estimation methods.

Key words: Bayesian method; double asymptotics; Markov chain and Monte Carlo; parametric regression; nonparametric regression; stationary time series data.

JEL Classification: C12, C14, C22.

Abbreviated Title: Nonparametric Estimation of Bayesian Means.

[1]The authors acknowledge constructive comments and suggestions from several seminar participants. Thanks also go to Tingting Cheng and Jiying Yin for their excellent computing assistance. The first author was supported by an Australian Research Council Professorial Fellowship Award: DP1096374 and an Australian Research Council Discovery Projects Scheme under Grant number: DP130104229. The second author acknowledges financial support by the National Science Foundation (SES 1164589) and SIEPR.

1 Introduction

Beyesian estimation and computation is a complicated issue, particularly when estimation issues involve computational complexity. The literature basically shows that there are three stages of the developments. The first stage of the developments is due to the fact that empirical Bayesian approach has been used to provide some closed–form solutions to various Bayesian estimation problems. One useful class of models covers a class of exponential families, in which the Bayes estimate is a ratio of the first–order derivative of the marginal density and the marginal density itself. Then, a nonparametric kernel density estimation method may be employed to consistently estimate the posterior mean. Similar approaches have also been done for a class of uniform families. The literature is summarised and discussed in Carlin and Louis (1996), Efron (1996), and some other studies.

Mainly due to the fact that most posterior means do not have closed–form relationships with the marginal density and its functionals, computation of posterior means involves dealing with some non–tractable integrals and possible high dimensionality and therefore becomes a very difficult issue. This comes to the second stage of the developments that importance sampling, the Gibbs sampler and other MCMC tools become available and effective for implementing Bayesian estimation and computation. There is a huge literature about such developments. We refer the reader to Liu (2001), Geweke (2005), and Brooks *et al* (2011). Since Bayesian inference basically relies on the full posterior density function and the dimensionality of such posterior density is usually large, both computation and simulation involve all sorts of difficulties. To partially address such computational issues, the third stage of the developments is based on the proposal of the so–called "Approximate Bayesian Computation" (ABC). Recent studies include Beaumont, Zhang and Balding (2002), Blum (2010), Fearnhead and Prangle (2012), and Blum *et al* (2013).

This paper proposes some general non– and semi–parametric regression approaches to the estimation and computation of posterior means involved in complex nonlinear econometric models. The proposed estimation method provides a simple and useful alternative to existing estimation methods, such as MCMC (Chernozhukov and Hong 2003), GMM (Hansen 1982), EMM (Gallant and Tauchen 1997), and NSMLE method proposed recently by Creel and Kristensen (2011), and Kristensen and Shin (2012). More recently, Gao and Hong (2014) considerably explore the ABC idea and the NSMLE method for nonparametric implementation of GMM in practice. As we discuss in Section 4 of this paper, based on direct sampling, the proposed nonparametric approach makes it possible to provide a closed–form estimate for a general conditional moment of the form $E[\psi(\theta)|T_n]$, where θ is the parameter of interest, $\psi(\cdot)$ is of a known functional form and T_n is a summary statistic, such as the sample mean of $X_1, X_2, \cdots, X_n$. As proposed in Section 4 below, moreover, a nonparametric estimation method based on resamples results in asymptotically normal estimates for unknown conditional moments with rates of convergence faster than those for existing estimates. Such theoretical findings are evaluated in Sections 6 and 7 through both simulated and real data examples.

In summary, this paper proposes non– and semi–parametric methods for the establishment of closed–form estimates for conditional moments. We believe that the newly proposed estimation method reveals some important findings and has the following theoretical and computational advantages:

a) it results in closed–form expressions for estimates of unknown parameters and functions involved in non– and semi–parametric models;

b) it avoids involving numerical approximations to intractable integrals involved in the computation of Bayesian estimates;

c) it directly and naturally addresses various high–dimensional issues involved in non– and semi–parametric approximation and estimation;

d) it facilitates both the implementation and the application of Bayesian estimation and computation for economic and financial models; and

e) it provides a simple and useful alternative to estimating unknown parameters and functions involved in classes of complex nonlinear econometric models.

The organisation of this paper is given as follows. Section 2 gives some examples and models to link and motivate the discussion of this paper with the relevant literature before a nonparametric kernel estimation method is proposed to estimate the posterior mean function. Section 2 then establishes an asymptotic theory for the estimation method proposed in this section. Using a resampling algorithm, Section 3 significantly improves the rate of convergence of a nonparametric kernel estimator based on the resamples and its resulting theory is then established in the end of Section 3. Section 4 proposes to estimate a general posterior mean of the form $E[\theta|T_n]$ before giving a comparison with an existing estimation method. Estimation problems involving dependent data are discussed respectively in Sections 3 and 4. Section 5 extends the discussion in Sections 2–4 to the case where there are nuisance parameters involved and then considers a nonparametric estimation issue where the nuisance parameters involved are consistently estimated. This large sample theory is supported by the small and finite sample evaluation given in Sections 6 and 7. Section 6 gives some numerical evidence to support the proposed models and estimation methods. An empirical example discussing parameter estimation of unknown parameters involved in a GARCH model is given in Section 7. Some concluding comments are given in Section 8 before the mathematical technicalities are given in Section 9.

2 Models and Estimation Methods

2.1 Examples and motivation

Before we propose our models and estimation methods, we use some examples to motivate our discussion.

Example 2.1: Consider a general distributional model of the form

$$X_t \sim F_t(x;\theta), \quad t = 1, 2, \cdots, n, \tag{2.1}$$

where each $F_t(\cdot;\theta)$ is a parametric distributional function indexed by θ, a vector of unknown parameters. Note that $\{X_t\}$ can be either independent, stationary or nonstationary time series.

For model (2.1), the vector of unknown parameters, θ, can be consistently estimated by classical estimates, such as, the conventional sample moment and MLE. Section 6 below shows that if we move one–step further by combining simulated samples with a nonparametric estimation method based on the simulated samples, a nonparametric kernel estimator for a conditional moment of the form $E[\psi(\theta)|T_n]$ is more efficient than such classical estimates, where T_n is a summary statistic, and $\psi(\cdot)$ is a known function available for computation.

Extensions of model (2.1) are needed to deal with the general conditional mean case discussed in Chen (2007), in which $F_t(\cdot;\cdot)$ is allowed to be semiparametric.

Example 2.2 (GARCH model): Consider a GARCH (1,1) model of the form:

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t, \; t = 1, 2, \cdots, n, \\ \sigma_t^2 &= b_0 + b_1 y_{t-1}^2 + b_2 \sigma_{t-1}^2, \end{aligned} \tag{2.2}$$

where $\{\varepsilon_t\}$ is a sequence of errors and $\theta = (b_0, b_1, b_2)'$ denotes a vector of unknown parameters.

Our study in Section 7 below discusses model (2.2) and evaluates the applicability and practical relevance of the proposed estimation method to be discussed in Sections 3 and 4 below to show that a nonparametric estimator for $g(T_n) = E[\theta|T_n]$ is more efficient than T_n itself when T_n is the MLE of θ.

2.2 Estimation based on simulation

Let $f(x|\theta)$ be the conditional density of x given θ and $\pi(\cdot)$ be the prior density. The Bayesian estimate of θ given x is defined by

$$g(x) = E[\theta|x] = \int \theta f(\theta|x) d\theta = \frac{\int \theta f(x|\theta)\pi(\theta) d\theta}{\int f(x|\theta)\pi(\theta) d\theta} \equiv \frac{q(x)}{p(x)}, \tag{2.3}$$

where $\theta = (\theta_1, \cdots, \theta_d)^\tau$ is a vector of unknown parameters, $p(x) = \int f(x|\theta)\pi(\theta)d\theta$ and $q(x) = \int \theta f(x|\theta)\pi(\theta)d\theta$.

Throughout the rest of this paper, we discuss the case where the model is exactly identified. To present the main idea in this section, we focus on the case of $d = 1$.

Assume that the functional form of $f(x|\theta)$ is available for computation. Suppose that θ_i is drawn from a proper probability density $\lambda(\theta)$ and both the forms of $\pi(\theta)$ and $\lambda(\theta)$ are available for computation. In this case, we may estimate $g(x)$ by

$$\widehat{g}_m(x) = \frac{\sum_{j=1}^m \theta_j f(x|\theta_j) \frac{\pi(\theta_j)}{\lambda(\theta_j)}}{\sum_{j=1}^m f(x|\theta_j) \frac{\pi(\theta_j)}{\lambda(\theta_j)}}, \tag{2.4}$$

and for the case where T_n is a summary statistic, we have

$$\widehat{g}_m(T_n) = \frac{\sum_{j=1}^m \theta_j f(T_n|\theta_j) \frac{\pi(\theta_j)}{\lambda(\theta_j)}}{\sum_{j=1}^m f(T_n|\theta_j) \frac{\pi(\theta_j)}{\lambda(\theta_j)}}. \tag{2.5}$$

Such discussions may be found from Geweke (1989), Gelfand and Smith (1990), and Geweke (2005) for examples. Note that there is no need to draw samples from $\pi(\theta)$ as long as it is possible to either draw $\{X_i\}$ from $p(\cdot)$ or to have the data $\{X_i\}$ available for use.

This section proposes to directly estimate the posterior mean by the nonparametric kernel method. In the rest of this section, we assume that we may draw (x_j, θ_j) jointly from $f(x|\theta)\lambda(\theta)$ when $\lambda(\theta)$ is a proper probability density. For notational simplicity, in the discussion of the rest of this section and Section 3, we assume that $\pi(\theta)$ is already a proper prior density available for sampling, and thus choose $\lambda(\cdot) = \pi(\cdot)$. This is consistent with the sampling approach adopted in the ABC literature. In Section 4 below, we consider the case where $\lambda(\cdot)$ is the only proper prior density available for sampling and computation. Section 5 proposes a nonparametric estimation method that is based on MCMC samples.

Equation (2.3) implies that we can introduce a regression model of the form

$$\theta = E[\theta|x] + (\theta - E[\theta|x]) \equiv g(x) + e, \tag{2.6}$$

where $e = \theta - E[\theta|x]$ satisfies $E[e|x] = 0$.

Note that the functional of $g(x)$ may not be feasibly available for computation even though the functional forms of $f(x|\theta)$ and $\pi(\theta)$ may be assumed to be either parametrically or semiparametrically known for sampling and computation. Thus, we propose to estimate $g(x)$ directly using the samples $\{(x_j, \theta_j)\}$ readily drawn from $f(x|\theta)\pi(\theta)$.

The first objective is to estimate $g(x)$. Suppose that we may simulate $(x_j, \theta_j : j = 1, 2, \cdots, m)$ directly from $f(x|\theta)\pi(\theta)$ and then define

$$\theta_j = g(x_j) + e_j, \quad j = 1, 2, \cdots, m, \tag{2.7}$$

where $\{e_j\}$ is a sequence of independent errors with mean zero and finite variance $\sigma^2 = E[e_1^2]$.

We then estimate $g(\cdot)$ by

$$g_m(x) = \sum_{j=1}^m K_{mj}(x)\theta_j, \tag{2.8}$$

where $K_{mj}(x) = \frac{K\left(\frac{x_j - x}{h}\right)}{\sum_{k=1}^{m} K\left(\frac{x_k - x}{h}\right)}$, in which $K(\cdot)$ is a kernel function and h is a bandwidth.

In order to incorporate the data $\{X_i : 1 \leq i \leq n\}$ into the estimation procedure, we simulate $\{\theta_{mi} : 1 \leq i \leq n < m\}$ from a regression model of the form

$$\theta_{mi} = g_m(X_i) + \varepsilon_{mi}, \tag{2.9}$$

where $\{\varepsilon_{mi} : 1 \leq i \leq n\}$ is available for sampling as a sequence of conditionally independent random errors given $\{(x_j, \theta_j) : 1 \leq j \leq m\}$, and is independent of $\{X_i\}$ satisfying

$$E[\varepsilon_{mi}|(x_1, \cdots, x_m; \theta_1, \cdots, \theta_m)] = 0 \quad \text{and} \quad E[\varepsilon_{mi}^2|(x_1, \cdots, x_m; \theta_1, \cdots, \theta_m)] = \sigma_{mx}^2 < \infty. \tag{2.10}$$

In practice, ε_{mi} can be simulated from $\varepsilon_{mi} = \lambda_{mi}(e_{m1}, \cdots, e_{mm})\ \xi_i$, in which $e_{mj} = \theta_j - g_m(x_j)$, $\{\xi_i\}$ is a sequence of independent and identically distributed (i.i.d.) random variables, with $E[\xi_1] = 0$ and $E\left[\xi_1^2\right] = 1$, generated from a pre–specified probability distribution, such as, either the standard normal distribution–$N(0,1)$ or

$$P\left(\eta_1 = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} \quad \text{and} \quad P\left(\eta_1 = \frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}, \tag{2.11}$$

and $\lambda_{mi}(\cdots)$ is a sequence of measurable functions. There are many cases one may use in practice:

- **Case I**: $\lambda_{mi}(e_{m1}, \cdots, e_{mm}) = \sigma_m$, where $\sigma_m^2 = \frac{1}{m}\sum_{j=1}^{m} e_{mj}^2$ with $e_{mj} = \theta_j - g_m(x_j)$;
- **Case II**: $\lambda_{mi}(e_{m1}, \cdots, e_{mm}) = e_{mi} + \frac{1}{\sqrt{m-n}}\sum_{j=n+1}^{m-n} e_{mj}$; and
- **case III**: $\lambda_{mi}(e_{m1}, \cdots, e_{mm}) = \sum_{j=1}^{m} \alpha_{ji} e_{mj}$, where $\{\alpha_{ji}\}$ is a sequence of real numbers chosen such that $\alpha_{ji} \geq 0$ and $\sum_{j=1}^{m} \alpha_{ji} = 1$.

The construction of equation (2.9) involves some kind of bootstrap idea through compressing the information already available from $\{(x_j, \theta_j) : j = 1, 2, \cdots, m\}$ and then equation (2.8). We finally estimate $g(x)$ by

$$g_{mn}(x) = \sum_{i=1}^{n} L_{ni}(x)\theta_{mi}, \tag{2.12}$$

where $L_{ni}(x) = \frac{L\left(\frac{X_i - x}{b}\right)}{\sum_{k=1}^{n} L\left(\frac{X_k - x}{b}\right)}$, in which $L(\cdot)$ is a kernel function and b is a bandwidth.

Our experience shows that the choice of $\{\varepsilon_{mi}\}$ does not affect asymptotic consistency of $g_{mn}(\cdot)$. Before asymptotic properties for $g_m(x)$ and $g_{mn}(x)$ are established in Sections 2.3 and 2.4 below, we summarise the estimation procedure as follows:

- **Step 1**: Simulate $\{(x_j, \theta_j) : j = 1, 2, \cdots, m\}$ from $f(x|\theta)\pi(\theta)$;

- **Step 2**: Estimate $g(x)$ by $g_m(x) = \sum_{j=1}^{m} K_{mj}(x)\theta_j$;
- **Step 3**: Simulate $\{\theta_{mi} : 1 \le i \le n\}$ from $\theta_{mi} = g_m(X_i) + \varepsilon_{mi}$; and
- **Step 4**: Re–estimate $g(x)$ by $g_{mn}(x) = \sum_{i=1}^{n} L_{ni}(x)\theta_{mi}$.

We will establish an asymptotic theory for the univariate case in Section 2.3 and then the multivariate case in Section 2.4 below.

2.3 Univariate Case

In this section, we assume that the dimensionality of θ is $d = 1$. Let $x = (x_1, \cdots, x_r)^\tau$ be the r–dimensional vector. To establish an asymptotic theory for $g_m(x)$ and $g_{mn}(x)$, we now introduce the following assumptions.

Assumption 2.1: (i) Let the product of $f(x|\theta)$ and $\pi(\theta)$ be a proper probability density function.

(ii) Let $f(x|\theta)$ be three times differentiable with respect to x and $f_x^{(i)}(x|\theta)$ be the i–th partial derivative of $f(x|\theta)$ with respect to x such that $\int |\theta| \left|\left|f_x^{(i)}(x|\theta)\right|\right| \pi(\theta)d\theta < \infty$ and $\int \left|\left|f_x^{(i)}(x|\theta)\right|\right| \pi(\theta)d\theta < \infty$ for any given x and $i = 0, \cdots, 3$, where $|| \cdot ||$ denotes the conventional Euclidean norm.

(iii) Suppose that both $p_2(x) = \int f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ and $q_2(x) = \int \theta f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ are continuous in x.

(iv) Suppose that $\{\theta_j : j = 1, 2, \cdots, m\}$ is a sequence of i.i.d. random variables drawn from $\pi(\theta)$ and that $\{(x_j, \theta_j) : j = 1, 2, \cdots, m\}$ is a vector of i.i.d. random vectors drawn from $f(x|\theta)\pi(\theta)$. Let $f(x)$ be the marginal density of $\{x_j\}$.

Assumption 2.2: (i) Suppose that there is a data set $\{X_i : i = 1, 2, \cdots, n\}$ that is available as an i.i.d. random variables with $p(x)$ being the density function.

(ii) Suppose that $\{X_i : i = 1, 2, \cdots, n\}$ is independent of $\{(x_j, \theta_j) : j = 1, 2, \cdots, m\}$. Let $\{\varepsilon_{mi}\}$ satisfy equations (2.9) and (2.10).

Assumption 2.3: (i) Let $K(\cdot)$ be the probability kernel function satisfying $\int uK(u)du = 0$, $0 < \int ||u||^2 K(u)du < \infty$ and $0 < \int K^2(u)du < \infty$. Let the bandwidth h satisfy $h \to 0$, $mh^r \to \infty$ and $mh^{r+4} \to c(r)$ for some $0 < c(r) < \infty$.

(ii) Let $L(\cdot)$ be a probability kernel function satisfying $\int vL(v)dv = 0$, $0 < \int ||v||^2 L(v)dv < \infty$, $0 < \int L^2(v)dv < \infty$, $\int ||v||^3 L(v)dv < \infty$ and $\int ||v||^4 L(v)dv < \infty$. Let the bandwidth b satisfy $b \to 0$ and $nb^r \to \infty$.

(iii) Let $\frac{h}{b} = o(1)$, $\frac{n}{m} = o(1)$, $nb^r h^4 = O(1)$, $nb^{r+4} = O(1)$ and $\frac{nb^r}{mh^r} = o(1)$ as $(m, n) \to (\infty, \infty)$.

Assumption 2.1(i) assumes the existence of proper density functions. Assumption 2.1(ii)(iii) is assumed to ensure that $g(x)$ is twice differentiable and that the second–order derivative is continuous. In the usual regression setting, such smoothness conditions are imposed directly on the conditional mean function $g(x)$. Assumption 3.1(iv) implies that (x_j, θ_j) and $e_j = \theta_j - g(x_j)$ are i.i.d. random variables. As discussed below, Assumption 2.1(iv) may be relaxed to the stationary and nonstationary time series case.

Assumption 2.2 imposes that there is a set of data $\{X_i : i = 1, 2, \cdots, n\}$ such that $\{X_i\}$ has a density function $p(x)$. Assumption 2.3 is a set of standard regularity conditions. Such conditions are therefore easily verifiable. Assumption 2.3(iii) basically imposes the rate of convergence on (h, b). When $h = C_1 \cdot m^{-\frac{1}{4+r}}$ and $b = C_2 \cdot n^{-\frac{1}{4+r}}$, Assumption 2.3(iii) reduces to just $\frac{n}{m} \to 0$.

We now establish the following theorems; their proofs are given in Section 9.1 below.

Theorem 2.1: *Let Assumptions 2.1 and 2.3(i) hold. Then as $m \to \infty$*

$$\sqrt{\frac{\sum_{j=1}^{m} K\left(\frac{x_j - x}{h}\right)}{\widehat{\sigma}_m^2}} \left(g_m(x) - g(x) - \sum_{j=1}^{r} B_j(x)\ h^2\right) \to_D N(0, \sigma^2(K)), \tag{2.13}$$

where $\sigma^2(K) = \int K^2(u)du$, $B_j(x) = \frac{\int ||u||^2 K(u)du}{2} \cdot \left(2f^{(j)}(x)g^{(j)}(x) + f(x)g^{(jj)}(x)\right)$ and $\widehat{\sigma}_m^2 = \frac{1}{m}\sum_{j=1}^{m}\left(\theta_j - g_m(x_j)\right)^2$, in which $r^{(j)}(x)$ and $r^{(jj)}(x)$ are the first and second order derivatives of $r(x) = g(x)$ or $f(x)$, respectively, and $f(x)$ denotes the marginal density function of x_j.

Theorem 2.2: *Let Assumptions 2.1–2.3 hold. Then as $(m, n) \to (\infty, \infty)$*

$$\sqrt{\frac{\sum_{i=1}^{n} L\left(\frac{X_i - x}{b}\right)}{\widehat{\sigma}_{mn}^2}} \left(g_{mn}(x) - g(x) - \sum_{j=1}^{r} B_j(x)\ h^2\right) \to_D N(0, \sigma^2(L)), \tag{2.14}$$

where $\sigma^2(L) = \int L^2(u)du$ and $\widehat{\sigma}_{mn}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\theta_{mi} - g_{mn}(X_i)\right)^2$.

Theorem 2.2 shows that one may use the data set $\{X_i : 1 \leq i \leq n\}$ to re–estimate $g(\cdot)$ and obtain asymptotic consistency. This is mainly because of the following reasoning:

$$\begin{aligned}
\sqrt{nb^r}&\left(g_{mn}(x) - g(x) - \sum_{j=1}^{r} B_j(x)\ h^2\right) = \sqrt{nb^r}\left(g_{mn}(x) - g_m(x)\right) \\
&+ \frac{\sqrt{nb^r}}{\sqrt{mh^r}} \cdot \sqrt{mh^r}\left(g_m(x) - g(x) - \sum_{j=1}^{r} B_j(x)\ h^2\right) \\
&= \sqrt{nb^r}\left(g_{mn}(x) - g_m(x) - \sum_{j=1}^{r} B_j(x)\ h^2\right) + o_P(1) \to_D N\left(0, \sigma^2(x)\right),
\end{aligned} \tag{2.15}$$

where $\sigma^2(x) > 0$ is a variance function.

2.4 Multivariate Case

In this section, we assume that the dimensionality of θ is $d > 1$. Let $x = (x_{(1)}, \cdots, x_{(r)})^\tau$ be the r–dimensional vector. We impose the following assumptions for the multivariate case.

Assumption 2.4: (i) Assumption 2.1(i) holds.

(ii) Let $f(x|\theta)$ be twice differentiable with respect to x such that $\int ||\theta|| \left|\left| f_x^{(i)}(x|\theta)\right|\right| \pi(\theta)d\theta < \infty$ and $\int \left|\left| f_x^{(i)}(x|\theta)\right|\right| \pi(\theta)d\theta < \infty$ for any given x and $i = 0, 1, 2$, where $f_x^{(i)}(x|\theta)$ denotes the i–th partial derivative of $f(x|\theta)$ with respect to x.

(iii) Assumption 2.1(iii) holds

(iv) Suppose that $\{\theta_j : j = 1, 2, \cdots, m\}$ is a vector of i.i.d. random variables drawn from $\pi(\theta)$ and that $\{(x_j, \theta_j) : j = 1, 2, \cdots, m\}$ is a vector of i.i.d. random variables drawn from $f(x|\theta)\pi(\theta)$.

As θ is now a vector, the conditions corresponding to Assumptions 2.1–2.3 are being changed. We now establish the following theorems; their proofs are given in Section 9.2 below.

Theorem 2.3: *Let Assumptions 2.2, 2.3(i) and 2.4 hold. Then as* $(m, n) \to (\infty, \infty)$

$$\sqrt{\sum_{j=1}^m K\left(\frac{x^j - x}{h}\right)} \cdot \widehat{\Sigma}_m^{-1} \left(g_m(x) - g(x) - \sum_{j=1}^r B_{jm}(x)\ h^2\right) \to_D N\left(0, \sigma^2(K) \cdot I_d\right), \quad (2.16)$$

where $\sigma^2(K) = \int K^2(u)du$, $\widehat{\Sigma}_m^2 = \frac{1}{m}\sum_{j=1}^m (\theta_j - g_m(x_j))(\theta_j - g_m(x_j))^\tau$, I_d *is the* $d \times d$ *identity matrix and* $B_{jm}(x)$ *is defined in the same way as for* $B_j(x)$.

Theorem 2.4: *Let Assumptions 2.2–2.4 hold. Then as* $(m, n) \to (\infty, \infty)$ *and* $\frac{nb^r}{mh^r} \to 0$

$$\sqrt{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} \cdot \widehat{\Sigma}_{mn}^{-1} \left(g_{mn}(x) - g(x) - \sum_{j=1}^r B_{jm}(x)\ h^2\right) \to_D N\left(0, \sigma^2(L) \cdot I_d\right), \quad (2.17)$$

where $\sigma^2(L) = \int L^2(u)du$ *and* $\widehat{\Sigma}_{mn}^2 = \frac{1}{n}\sum_{i=1}^n (\theta_{mi} - g_{mn}(X_i))(\theta_{mi} - g_{mn}(X_i))^\tau$.

It is pointed out that when r, the dimensionality of x, is large, one should use a dimensional–reduction method, such as, either an additive model or a single–index model, as discussed in Chapter 2 of Gao (2007), to approximate $g(x)$ and $g(T_n)$ as considered in Section 4 below. We leave such discussion to future research.

As discussed in Section 3 below, the rate of convergence of $g_{mn}(x)$ can be made faster than the standard rate when a resampling method is used for generating new samples.

3 Estimation based on Resampling

3.1 Resampling for stationary data

This section considers the case where $\{X_i : i = 1, 2, \cdots, n\}$ is available as a data set and $\{X_i\}$ is stationary time series having the same marginal density as $p(x)$. Let $f(\theta|x)$ be the conditional density of θ given $X_i = x$. Using the Metropolis–Hastings algorithm (see, for example, Chib and Greenberg 1995), we generate a stationary sequence $\theta_{i1}, \cdots, \theta_{im}$ from a proposal density such that as $j \to \infty$, the limiting density of θ_{ij} is $f(\theta|X_i)$ (see, for example, Theorem 3 of Tierney 1994). Note that we need not require $\{\theta_{ij}; 1 \le j \le m\}$ to be stationary, although they may be conditionally stationary given X_i.

Recall $g(x) = E[\theta|x] = \int \theta f(\theta|x) d\theta$. The main objective of this section is to estimate $g(x)$ based on $\{(\theta_{ij}, X_i) : 1 \le i \le n; 1 \le j \le m\}$. Let $e_{ij} = \theta_{ij} - g(X_i)$, $\theta_{mi} = \frac{1}{m}\sum_{j=1}^m \theta_{ij}$ and $e_{mi} = \frac{1}{m}\sum_{j=1}^m e_{ij}$. Then, we have

$$\begin{aligned} \theta_{ij} &= g(X_i) + e_{ij}, \quad i = 1, 2, \cdots, n; \; j = 1, 2, \cdots, m; \\ \theta_{mi} &= g(X_i) + e_{mi}, \quad i = 1, 2, \cdots, n, \end{aligned} \tag{3.1}$$

where $\{e_{ij} : 1 \le i \le n; 1 \le j \le m\}$ is assumed to be a stationary sequence in Assumption 3.1 below.

As in Section 2 above, we estimate $g(\cdot)$ by

$$g_{mn}(x) = \sum_{i=1}^n L_{ni}(x)\theta_{mi}, \tag{3.2}$$

where $L_{ni}(x) = \frac{L\left(\frac{X_i - x}{b}\right)}{\sum_{l=1}^n L\left(\frac{X_l - x}{b}\right)}$, in which $L(\cdot)$ is a probability kernel function and b is a bandwidth parameter.

In order to establish an asymptotic theory for $g_{mn}(x)$, we need to introduce the following assumptions.

Assumption 3.1: (i) Suppose that $\{X_i\}$ is a vector of stationary time series data that are available for generating $\{\theta_{ij}\}$. Let $e_{ij} = \theta_{ij} - E[\theta_{ij}|X_i]$ and $e_i = (e_{i1}, \cdots, e_{im})^\tau$. Suppose that $\{(e_i, X_i)\}$ is a vector of stationary time series satisfying $0 < E[e_{ij}^2|X_i = x] = \sigma^2(x) < \infty$ and $E\left[e_{ij}^4|X_i = x\right] = \mu_4(x) < \infty$, where $\sigma^2(x)$ is continuous at x.

(ii) Let $\gamma_j(x) = E[e_{1,1+j}e_{11}|X_1 = x]$ satisfy $\sum_{j=1}^\infty |\gamma_j(x)| < \infty$ and $\lambda(x) \equiv \sigma^2(x) + 2\sum_{j=1}^\infty \gamma_j(x) > 0$ for each given x. Suppose that $\{(e_i, X_i)\}$ is ρ–mixing with mixing coefficient $\rho(\cdot)$ satisfying $\sum_{k=1}^\infty k^u \sqrt{\rho(k)} < \infty$ for some $u > \frac{1}{2}$. In addition, the conditional density of (X_1, X_j) given (e_{m1}, e_{mj}) is bounded by a positive constant independent of $j > 1$.

Assumption 3.2: (i) Let $f(x|\theta)$ be twice differentiable with respect to x such that

$$\int |\theta| \left|\left|f_x^{(i)}(x|\theta)\right|\right| \pi(\theta) d\theta < \infty \quad \text{and} \quad \int \left|\left|f_x^{(i)}(x|\theta)\right|\right| \pi(\theta) d\theta < \infty$$

for any given x and $i = 0, 1, 2$, where $f_x^{(i)}(x|\theta)$ denotes the i–th partial derivative of $f(x|\theta)$ with respect to x, and $||\cdot||$ denotes the conventional Euclidean norm.

(ii) Suppose that both $p_2(x) = \int f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ and $q_2(x) = \int \theta f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ are continuous in x.

Assumption 3.3: (i) Let $L(\cdot)$ be a bounded probability kernel function satisfying $\int vL(v)dv = 0$, $0 < \int ||v||^2 L(v)dv < \infty$ and $0 < \int L^2(v)dv < \infty$.

(ii) Let the bandwidth b satisfy $b \to 0$ and $nb^{r(1+\frac{2}{2u+1})} = O(n^c)$ for some $c > 0$, where r is the dimensionality of X_i. In addition, $mnb^r \to \infty$ and $mnb^{r+4} = O(1)$ as $(m, n) \to (\infty, \infty)$.

There case where $\{X_i\}$ is a sequence of i.i.d. random variables is covered in Assumption 3.1. The verification of Assumptions 3.1–3.3 may be done similarly to what has been done for Assumptions 2.1–2.3. While the assumptions may not be the weakest ones, they are easily verifiable. The stationarity assumption is based on the nature of the MCMC algorithm. The mixing condition is also standard while the fourth moment condition on $E[e_{ij}^4] < \infty$ may be weakened to $E\left[|e_{ij}|^{2+c(e)}\right] < \infty$ for some $c(e) > 0$. Assumption 3.2 is needed to ensure that the second–order derivative of $g(x)$, $g^{(2)}(x)$, is continuous. The bandwidth conditions assumed in Assumption 3.3(ii) are also quite standard.

We now establish the following theorem; its proof is given in Section 9.3 below.

Theorem 3.1: *Suppose that Assumptions 3.1–3.3 are satisfied. Then, we have as* $(m, n) \to (\infty, \infty)$

$$\sqrt{m \cdot \sum_{i=1}^{n} L\left(\frac{X_i - x}{b}\right)} \left(g_{mn}(x) - g(x) - \sum_{j=1}^{r} B_j(x)\, b^2\right) \to_D N\left(0, \Sigma(x)\right), \tag{3.3}$$

where $B_j(x)$ *is the same as defined in Theorem 2.1 and* $\Sigma(x) = \lambda(x) \cdot \int L^2(v)dv$, *in which* $\lambda(x) = \sigma^2(x) + 2\sum_{j=1}^{\infty} \gamma_j(x)$.

Theorem 3.1 shows that one can achieve a fast rate of convergence of an order of the form $\left(\sqrt{mnb^r}\right)^{-1} = m^{-\frac{1}{2}} \cdot \left(\sqrt{nb^r}\right)^{-1} = o\left(\left(\sqrt{nb^r}\right)^{-1}\right)$ as $m \to \infty$, because our estimation method makes the best use of the availability of the sample $(X_1, \cdots, X_n)$. The finite sample evaluation given in Section 6 below supports this fast rate of convergence. In the following subsection, we consider the case where a summary statistic is available for resampling.

3.2 Resampling for nonstationary data

Since $\{X_i : i = 1, 2, \cdots, n\}$ is available as a nonstationary time series in many practical situations, this section considers the nonstationary case. Using the Metropolis–Hastings algorithm (see, for example, Chib and Greenberg 1995) again, we generate an array of random variables, $\theta_{i1}, \cdots, \theta_{im}$, from a proposal density for each given X_i. Once again, we need not require $\{\theta_{ij}; 1 \leq j \leq m\}$ to be stationary. Consider the case of $r = 1$ in this subsection.

We then assume that there are an array of martingale differences $\{\varepsilon_{ij}\}$ and a suitable function $g(\cdot)$ such that

$$\begin{aligned} \theta_{ij} &= g(X_i) + \varepsilon_{ij}, \quad i = 1, 2, \cdots, n; \; j = 1, 2, \cdots, m; \\ \theta_{mi} &= g(X_i) + \varepsilon_{mi}, \quad i = 1, 2, \cdots, n, \end{aligned} \tag{3.4}$$

where $\theta_{mi} = \frac{1}{m}\sum_{j=1}^{m} \theta_{ij}$ and $\varepsilon_{mi} = \frac{1}{m}\sum_{j=1}^{m} \varepsilon_{ij}$.

We then estimate $g(\cdot)$ by

$$g_{mn}(x) = \sum_{i=1}^{n} L_{ni}(x)\theta_{mi}, \tag{3.5}$$

where $L_{ni}(x) = \frac{L\left(\frac{X_i - x}{b}\right)}{\sum_{l=1}^{n} L\left(\frac{X_l - x}{b}\right)}$, in which $L(\cdot)$ is a probability kernel function and b is a bandwidth parameter.

For the case where X_i is nonstationary and $U_i = X_i - X_{i-1}$ reduces to be stationary, we modify Assumptions 3.1–3.3 as follows.

Assumption 3.4: (i) Suppose that $\{\varepsilon_{mi}\}$ and U_i are independent of each other. Suppose also that there is a stochastic process $B(r)$ such that $\sup_{0 \leq r \leq 1} \left| \frac{X_{[nr]}}{\sqrt{n}} - B(r) \right| = o_P(1)$ as $n \to \infty$.

(ii) Let $\{\varepsilon_{ij}, \mathcal{F}_{mi} : 1 \leq i \leq n\}$ be an array of martingale differences with $E\left[\varepsilon_{ij} | \mathcal{F}_{m,i-1}\right] = 0$ and $\max_{1 \leq j \leq m} E\left[\varepsilon_{ij}^4 | \mathcal{F}_{m,i-1}\right] < \infty$ almost surely (a.s.). Moreover, there is some $0 < \sigma_\varepsilon^2 < \infty$ such that $\frac{1}{m}\sum_{j=1}^{m} E\left[\varepsilon_{ij}^2 | \mathcal{F}_{m,i-1}\right] \to_{a.s.} \sigma_\varepsilon^2$ and $\frac{1}{m}\sum_{j_1=2}^{m} \sum_{j_2=1}^{j_1-1} E\left[\varepsilon_{ij_1}\varepsilon_{ij_2} | \mathcal{F}_{m,i-1}\right] \to_{a.s.} 0$ as $m \to \infty$.

Assumption 3.5: (i) Let $f(x|\theta)$ be twice differentiable with respect to x such that

$$\int |\theta| \left|\left| f_x^{(i)}(x|\theta) \right|\right| \pi(\theta) d\theta < \infty \text{ and } \int \left|\left| f_x^{(i)}(x|\theta) \right|\right| \pi(\theta) d\theta < \infty$$

for any given x and $i = 0, 1, 2$, where $f_x^{(i)}(x|\theta)$ denotes the i–th partial derivative of $f(x|\theta)$ with respect to x, and $||\cdot||$ denotes the conventional Euclidean norm.

(ii) Suppose that both $p_2(x) = \int f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ and $q_2(x) = \int \theta f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ are continuous in x.

Assumption 3.6: (i) Let $L(\cdot)$ be a bounded probability kernel function satisfying $\int vL(v)dv = 0$, $0 < \int ||v||^2 L(v)dv < \infty$ and $0 < \int L^2(v)dv < \infty$.

(ii) Let the bandwidth b satisfy $b \to 0$, $m\sqrt{n}b \to \infty$ and $m\sqrt{n}b^5 \to c(0)$ for some $0 < c(0) < \infty$.

The verification of Assumptions 3.4–3.6 may be done in a similar way to those of Theorem 3.2 of Gao and Phillips (2013). We then establish the following theorem; its proof is given in Section 9.4 below.

Theorem 3.2: *Suppose that Assumptions 3.4–3.6 are satisfied. Then, we have as* $(m,n) \to (\infty, \infty)$

$$\sqrt{m \sum_{i=1}^{n} L\left(\frac{X_i - x}{b}\right)} \left(g_{mn}(x) - g(x) - B(x)\ b^2\right) \to_D N\left(0, \Sigma_1^2\right), \tag{3.6}$$

where $\Sigma_1^2 = \sigma_\varepsilon^2 \cdot \int L^2(v)dv$ *and* $B(x) = \frac{\int u^2 L(u) du}{2} \cdot \left(2f^{(1)}(x)g^{(1)}(x) + f(x)g^2(x)\right)$. *Note that we also have* $\frac{1}{\sqrt{nb}} \sum_{i=1}^{n} L\left(\frac{X_i - x}{b}\right) \to_D L_W(1,0)$, *in which* $L_W(1,0)$ *is a local–time random variable driven by a standard Brownian process* $W(r)$.

In comparison with Theorem 3.2 of Gao and Phillips (2013), Theorem 3.2 establishes a much faster rate of $(m\sqrt{n}b)^{-\frac{1}{2}}$ than $(\sqrt{n}b)^{-\frac{1}{2}}$ when a sampling method is used. Meanwhile, the multivariate case may be done similarly to Chapter 2 of Gao (2007), and Gao and Phillips (2013) when a semiparametric reduction method is used. When there is a type of endogeneity, bias corrections may be done similarly to Phillips and Hansen (1990).

4 Estimation Based on Summary Statistics

We consider the case where we may use a summary statistic based on direct sampling, importance sampling and resampling in Sections 4.1–4.3, respectively.

4.1 Estimation based on direct sampling

In econometric estimation problems, the parameter–of–interest, θ, is often involved in a complex model, such as, a structural model of the form $\psi(X;\theta) = 0$. Instead of estimating a conditional mean of the form $E[\theta|X]$, we may make the best use of the availability of some summary statistics. In this case, we may just be interested in estimating the conditional mean $g(T_n) = E\left[\theta|T_n\right]$, where T_n is a one–dimensional summary statistic, such as the sample mean $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, in which $X_1, X_2, \cdots, X_n$ is a sequence of i.i.d. random variables.

Suppose that we may sample $\{\theta_j : 1 \leq j \leq m\}$ from $\pi(\theta)$ and then (θ_j, T_{nj}) from $f(T_n|\theta)\pi(\theta)$. We then estimate $g(T_n)$ by

$$g_{km}(T_n) = \frac{\sum_{j=1}^{m} K\left(\frac{T_{nj} - T_n}{h}\right)\theta_j}{\sum_{j=1}^{m} K\left(\frac{T_{nj} - T_n}{h}\right)}, \tag{4.1}$$

where $K(\cdot)$ is a probability kernel function and h is a bandwidth parameter.

In some situations, we may estimate $g(T_n)$ by

$$g_{am}(T_n) = \frac{\sum_{j=1}^{m} f(T_n|\theta_j)\theta_j}{\sum_{j=1}^{m} f(T_n|\theta_j)} \tag{4.2}$$

when $f(T_n|\theta)$ is available for feasible computation.

We will show that $g_{km}(T_n)$ is more efficient than T_n and $g_{am}(T_n)$ in terms of their standard deviations. In order to establish an asymptotic theory for $g_{km}(T_n)$, we introduce the following assumptions.

Assumption 4.1: (i) Let the product of $f(T_n|\theta)$ and $\pi(\theta)$ be a proper probability density function for each given $n \geq 1$.

(ii) Suppose that (T_{nj}, θ_j) is a vector of i.i.d. random variables sampled from $f(T_n|\theta)\pi(\theta)$, and that T_{nj} and T_n are independent of each other and have the same distribution. For each given $n \geq 1$, let $e_{nj} = \theta_j - g(T_{nj})$ be a sequence of i.i.d. errors independent of $\{T_{nj}\}$ with $E[e_{n1}] = 0$, $0 < E[e_{n1}^2] < \infty$ and $E[e_{n1}^4] < \infty$.

(iii) Let $f(x|\theta)$ be twice differentiable with respect to x such that

$$\sup_{n\geq 1} \int |\theta| \left|\left| f_1^{(i)}(T_n|\theta) \right|\right| \pi(\theta) d\theta < \infty \text{ and } \sup_{n\geq 1} \int \left|\left| f_1^{(i)}(T_n|\theta) \right|\right| \pi(\theta) d\theta < \infty$$

for $i = 0, 1, 2$, where $f_1^{(i)}(T_n|\theta)$ denotes the i–th partial derivative of $f(T_n|\theta)$ with respect to T_n, and $||\cdot||$ denotes the conventional Euclidean norm.

(iii) Suppose that both $p_2(x) = \int f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ and $q_2(x) = \int \theta f_x^{(2)}(x|\theta)\pi(\theta)d\theta$ are continuous in x.

Assumption 4.2: (i) Let $K(\cdot)$ be the probability kernel function satisfying $\int uK(u)du = 0$, $0 < \int u^2K(u)du < \infty$, and $0 < \int K^2(u)du < \infty$ and $0 < \int u^2K^2(u)du < \infty$.

(ii) Let the bandwidth h satisfy $h \to 0$, $mh \to \infty$, $\frac{mh}{\lambda_n^2} \to \infty$, $\frac{\sqrt{mh^5}}{\lambda_n}\gamma(T_n) \to C_\lambda$ for some $0 < C_\lambda < \infty$ and $\frac{h}{\lambda_n} \cdot \gamma(T_n) \to 0$ as $(m,n) \to (\infty,\infty)$, where $\lambda_n^2 = \text{Var}[\theta|T_n]$ and $\gamma(T_n) = \frac{1}{2}g^{(2)}(T_n) + \frac{p^{(1)}(T_n)}{p(T_n)}g^{(1)}(T_n) + g^{(1)}(T_n)$, in which $p(T_n) = \int f(T_n|\theta)\pi(\theta)d\theta$.

Assumption 4.1 corresponds to Assumption 2.1. Assumption 4.2(i) is similar to Assumption 2.3(i). Assumption 4.2(ii) imposes a set of additional conditions on the relationship among (m, h, λ_n) for the establishment of an asymptotic normality for $g_{km}(T_n)$.

We then introduce the following assumption for the establishment of an asymptotic normality for $g_{am}(T_n)$.

Assumption 4.3: (i) Suppose that $\{\theta_j\}$ is a sequence of i.i.d. random variables drawn from $\pi(\theta)$.

(ii) For given T_n, suppose that $\sigma^2(T_n) = \int \left(\theta f(T_n|\theta) - \frac{q(T_n)}{p(T_n)}f(T_n|\theta)\right)^2 \pi(\theta)d\theta$ satisfies $\inf_{n\geq 1}\sigma^2(T_n) > 0$ and $\frac{m\ p^2(T_n)}{\sigma^2(T_n)} \to \infty$ as $(n,m) \to (\infty,\infty)$, where $q(T_n) = \int \theta f(T_n|\theta)\pi(\theta)d\theta$.

(iii) For given T_n, suppose that $\sup_{n\geq 1}\int \left|\theta f(T_n|\theta) - \frac{q(T_n)}{p(T_n)}f(T_n|\theta)\right|^4 \pi(\theta)d\theta < \infty$.

Assumption 4.3(ii)(iii) can be easily verifiable in many cases. For example, we consider case where $T_n = \frac{1}{n}\sum_{j=1}^n X_i \sim N\left(\theta, \frac{1}{n}\right)$ when $X_i \sim N(\theta, 1)$ and $\theta \sim N(0,1)$. Note that in this case, we have $f(T_n|\theta) = \frac{1}{\sqrt{2\pi}\sigma_n}e^{-\frac{(T_n-\theta)^2}{2\sigma_n^2}}$, $p(T_n) = \frac{1}{\sqrt{2\pi(1+\sigma_n^2)}}e^{-\frac{T_n^2}{2(1+\sigma_n^2)}}$, $q(T_n) = p(T_n) \cdot \frac{T_n}{1+\sigma_n^2}$ and $\sigma^2(T_n) = \frac{1}{2\pi} \cdot \frac{1}{2+\sigma_n^2}\exp\left(-\frac{T_n^2}{2+\sigma_n^2}\right) \cdot \left(\frac{1}{2+\sigma_n^2} + \frac{4T_n^2}{(2+\sigma_n^2)^2}\right)$, where $\sigma_n^2 = \frac{1}{n}$.

With this setting, Assumption 4.2(ii) requires that (m, n, h) satisfies $m\ n\ h \to \infty$, $nh^2 \to 0$, $m\ n\ h^5 = O(1)$ and $m\ n\ h^7 \to 0$. Such conditions are satisfiable when (m, n) is suitably chosen. In the case $n = [m^c]$ for some $0 < c < 1$, one may just need to choose c such that $m^{1+c}h^5 = O(1)$ and $m^c h^2 \to 0$. As shown in Theorem 4.1 below, with the possibility to sample (θ_j, T_{nj}) from $f(T_n|\theta)\pi(\theta)$, the asymptotic variance of $g_{km}(\cdot)$ is of an order of the form $\frac{1}{mnh}$, which is smaller than the conventional order of $\frac{1}{nh}$, while the bias term remains the same.

Assumption 4.1(i)(ii) just imposes the i.i.d. structure on (θ_j, T_{nj}). Assumptions 4.1(iii) and 4.3 then impose some moment conditions on $f(T_n|\theta)$. In other words, there is no need to require assuming an explicit or implicit distributional structure or even asymptotically distributional structure on T_n.

We establish two important asymptotic distributions in Theorems 4.1 and 4.2 below; their proofs are given in Sections 9.5 and 9.6, respectively.

Theorem 4.1: *Let Assumptions 4.1 and 4.2 hold. Then as* $(m, n) \to (\infty, \infty)$

$$\sqrt{\frac{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)}{\lambda_n^2}} \left(g_{km}(T_n) - g(T_n) - B(T_n)\ h^2\right) \to_D N\left(0, \sigma^2(K)\right), \tag{4.3}$$

where $\sigma^2(K) = \int K^2(u)du$, *and* $B(T_n) = \frac{\int u^2 K(u)du}{2} \cdot \left(2p^{(1)}(T_n)g^{(1)}(T_n) + p(T_n)g^2(T_n)\right)$.

Theorem 4.2: *Let Assumption 4.3 hold. Then as* $(m, n) \to (\infty, \infty)$

$$\sqrt{\frac{m\ p^2(T_n)}{\sigma^2(T_n)}} \left(g_{am}(T_n) - g(T_n)\right) \to_D N(0, 1), \tag{4.4}$$

where $\sigma^2(T_n)$ *is as defined in Assumption 4.3(ii).*

Let us now compare the rates of convergence in (4.3) and (4.4). Note that the rate of convergence in (4.3) of an order of $\sqrt{m} \cdot \sqrt{nh}$ is faster than the rate of convergence of $\sqrt{m}$ involved in (4.4) when $nh \to \infty$, $n\ \lambda_n^2 \to C_1 > 0$ and $\frac{p(T_n)}{\sigma(T_n)} \to_P C_2 > 0$.

In general, we conclude that $g_{km}(\cdot)$ has a faster rate of convergence than that for g_{am} as long as $\frac{\sqrt{h}\ \sigma(T_n)}{\lambda_n\ p(T_n)} \to \infty$, which is easily verified when $nh \to \infty$, $n\ \lambda_n^2 \to C_1 > 0$ and $\frac{p(T_n)}{\sigma(T_n)} \to_P C_2 > 0$. Such an asymptotic behaviour is verified by the finite–sample evaluation in Sections 6 and 7 below.

4.2 Estimation based on importance sampling

In the above discussions, we assume that $\pi(\theta)$ is a proper probability density function and it is feasible to sample $\{\theta_j\}$ from $\pi(\theta)$. Let $\lambda(\theta)$ be the importance distribution and θ_j^* be sampled from $\lambda(\theta)$. We also assume that the ratio $\frac{\pi(\theta)}{\lambda(\theta)}$ is available for computation.

Let $\rho(\theta) = \frac{\pi(\theta)}{\lambda(\theta)}$. Suppose that we can sample $\{\theta_j^*\}$ from $\lambda(\theta)$ and (θ_j^*, T_{nj}^*) from $f(T_n|\theta)\lambda(\theta)$. Then, we can replace $g_{km}(\cdot)$ and $g_{am}(\cdot)$ by

$$g_{km}^*(T_n) = \frac{\sum_{j=1}^m K\left(\frac{T_{nj}^* - T_n}{h}\right)\rho(\theta_j^*)\theta_j^*}{\sum_{j=1}^m K\left(\frac{T_{nj}^* - T_n}{h}\right)\rho(\theta_j^*)}, \tag{4.5}$$

$$g_{am}^*(T_n) = \frac{\sum_{j=1}^m f(T_n|\theta_j^*)\rho(\theta_j^*)\theta_j^*}{\sum_{j=1}^m f(T_n|\theta_j^*)\rho(\theta_j^*)}. \tag{4.6}$$

In this case, Assumptions 4.1–4.3 may be replaced by Assumptions 4.1*–4.3* below.

Assumption 4.1*: (i) Let the product of $f(T_n|\theta)$ and $\lambda(\theta)$ be a proper probability density function for each given $n \geq 1$.

(ii) Suppose that (T_{nj}^*, θ_j^*) is a vector of i.i.d. random variables sampled from $f(T_n|\theta)\lambda(\theta)$. Let $e_j^* = \theta_j^* - g\left(T_{nj}^*\right)$ be a sequence of i.i.d. errors with $E\left[e_j^*|T_{nj}^*\right] = 0$, $0 < E\left[e_j^{*2}|T_{nj}^*\right] < \infty$ and $E\left[e_j^{*4}|T_{nj}^*\right] < \infty$.

(iii) Let $f(x|\theta)$ be twice differentiable with respect to x such that

$$\sup_{n\geq 1} \int |\theta| \left|\left|f_1^{(i)}(T_n|\theta)\right|\right| \rho(\theta)\lambda(\theta)d\theta < \infty \text{ and } \sup_{n\geq 1} \int \left|\left|f_1^{(i)}(T_n|\theta)\right|\right| \rho(\theta)\lambda(\theta)d\theta < \infty$$

for $i = 0, 1, 2$, where $f_1^{(i)}(T_n|\theta)$ denotes the i–th partial derivative of $f(T_n|\theta)$ with respect to T_n, and $||\cdot||$ denotes the conventional Euclidean norm.

(iii) Suppose that both $p_2(x) = \int f_x^{(2)}(x|\theta)\rho(\theta)\lambda(\theta)d\theta$ and $q_2(x) = \int \theta f_x^{(2)}(x|\theta)\rho(\theta)\lambda(\theta)d\theta$ are continuous in x.

Assumption 4.2*: Let Assumption 4.2 hold with $\sigma^2(T_n)$ being replaced by $\sigma^{*2}(T_n) = \mathrm{Var}[\theta^*|T_n]$.

Assumption 4.3*: (i) Suppose that $\{\theta_j^*\}$ is a sequence of independent and identically distributed (i.i.d.) random variables drawn from $\lambda(\theta)$.

(ii) For given T_n, suppose that $\sigma^{2*}(T_n) = \int \left(\theta f(T_n|\theta) - \frac{q^*(T_n)}{p^*(T_n)} f(T_n|\theta)\right)^2 \rho^2(\theta)\lambda(\theta)d\theta$ satisfies $\inf_{n\geq 1} \sigma^{2*}(T_n) > 0$ and $\frac{m\ p^{*\ 2}(T_n)}{\sigma^2(T_n)} \to \infty$, where $p^*(T_n) = \int f(T_n|\theta)\rho(\theta)\lambda(\theta)d\theta$ and $q^*(T_n) = \int \theta f(T_n|\theta)\rho(\theta)\lambda(\theta)d\theta$.

(iii) For given T_n, suppose that $\sup_{n\geq 1} \int \left|\theta f(T_n|\theta) - \frac{q(T_n)}{p(T_n)} f(T_n|\theta)\right|^4 \rho^4(\theta)\lambda(\theta)d\theta < \infty$.

We then have the following theorems.

Theorem 4.1*: *Let Assumptions 4.1* and 4.2* hold. Then, as* $(m, n) \to (\infty, \infty)$

$$\sqrt{\frac{\sum_{j=1}^m K\left(\frac{T_{nj}^* - T_n}{h}\right)\rho(\theta_j^*)}{\lambda_n^2}}\ \left(g_{km}^*(T_n) - g(T_n) - B(T_n)\ h^2\right) \to_D N\left(0, \sigma^{*2}(K)\right), \tag{4.7}$$

whenever $\frac{\int \rho^2(\theta)\lambda(\theta)d\theta}{\int \rho(\theta)\lambda(\theta)d\theta} > 0$, *where* $\sigma^{*2}(K) = \int K^2(u)du \cdot \frac{\int \rho^2(\theta)\lambda(\theta)d\theta}{\int \rho(\theta)\lambda(\theta)d\theta}$ *and* $B(T_n)$ *is the same as defined in Theorem 4.1.*

Theorem 4.2*: *Let Assumption 4.3* hold. Then, as* $(m, n) \to (\infty, \infty)$

$$\sqrt{\frac{m\ p^{*2}(T_n)}{\sigma^{*2}(T_n)}}\,(g_{am}^*(T_n) - g(T_n)) \to_D N(0,1), \tag{4.8}$$

where $\sigma^{*2}(T_n)$ *is as defined in Assumption 4.3*(ii).*

The proofs and implications of Theorem 4.1* and 4.2* are almost the same as those of Theorems 4.1 and 4.2, and are therefore omitted.

4.3 Estimation based on resampling of summary statistics

Let T_n be a summary statistic and denote its density function by $p_n(\cdot)$ and $f_n(\theta|T_n)$ be the conditional density of θ given T_n, where n is the number of observations involved in T_n. Suppose that one can sample a stationary sequence $T_{n1}, \cdots, T_{nN}$ from the distribution of T_n and then an array of random variables $\{\theta_{nij} : 1 \le j \le m\}$ from a proposal density such that as $m \to \infty$, the limiting density of θ_{nij} is $f_n(\theta|T_{ni})$ for each fixed i.

Similarly to equation (3.1), we may write

$$\begin{aligned} \theta_{nij} &= g(T_{ni}) + \varepsilon_{nij},\ i = 1, 2, \cdots, N;\ j = 1, 2, \cdots, m, \\ \theta_{ni} &= g(T_{ni}) + \varepsilon_{ni},\ i = 1, 2, \cdots, N, \end{aligned} \tag{4.9}$$

where $\theta_{ni} = \frac{1}{m}\sum_{j=1}^m \theta_{nij}$ and $\varepsilon_{ni} = \frac{1}{m}\sum_{j=1}^m \varepsilon_{nij}$.

In the same way as in equation (4.1), we estimate $g(\cdot)$ by

$$g_{mnN}(T_n) = \frac{\sum_{i=1}^N K\left(\frac{T_{ni}-T_n}{h}\right)\theta_{ni}}{\sum_{i=1}^N K\left(\frac{T_{ni}-T_n}{h}\right)}, \tag{4.10}$$

where $K(\cdot)$ is a probability kernel function and h is a bandwidth parameter.

Combining the establishments and the proofs of Theorems 3.1 and 4.1, we have as $(m, n, N) \to (\infty, \infty, \infty)$

$$\sqrt{\frac{m\sum_{i=1}^N K\left(\frac{T_{ni}-T_n}{h}\right)}{\lambda_n^2}}\ \left(g_{mnN}(T_n) - g(T_n) - B(T_n)\ h^2\right) \to_D N\left(0, \Sigma^2(K)\right), \tag{4.11}$$

where $\Sigma^2(K)$ is defined in the same way as for $\Sigma(x)$ involved in Theorem 3.1, and λ_n^2 and $B(T_n)$ are defined in the same way as in Theorem 4.1.

Note that the rate of convergence involved in (4.11) can be as fast as $m^{-\frac{1}{2}} \cdot \left(\sqrt{n\ Nh}\right)^{-\frac{1}{2}}$. In comparison with Theorem 4.1, the rate of convergence of $g_{mnN}(T_n)$ can be improved by an order of $m^{-\frac{1}{2}}$ when a resampling algorithm is used. The estimation theory is applied to a stationary GARCH model in Sections 6 and 7 below. The case involving summary statistics of nonstationary data may be discussed analogously.

5 Estimation with Hyperparameters

In the discussion so far, we assume that the prior density functions $\pi(\theta)$ and $\lambda(\theta)$ do not involve any hyperparameters. Thus, there is no parameter involved in $g(x)$ and $g(T_n)$. Both functions are estimated nonparametrically by the kernel method even though the functional forms of $f(x|\theta)$, $f(T_n|\theta)$ and $\pi(\theta)$ may be assumed to be known for sampling and computation.

This section considers the case where there is a vector of hyperparameters involved in $\pi(\theta)$, denoted as $\pi(\theta;\gamma)$. We still assume that $\pi(\theta;\gamma)$ is available for sampling and computation when the value of γ is given. We introduce the following definition:

$$g(x;\gamma) = \frac{\int \theta f(x|\theta)\pi(\theta;\gamma)d\theta}{\int f(x|\theta)\pi(\theta;\gamma)d\theta} \equiv \frac{q(x;\gamma)}{p(x;\gamma)}, \tag{5.1}$$

where $p(x;\gamma) = \int f(x|\theta)\pi(\theta;\gamma)d\theta$ and $q(x;\gamma) = \int \theta f(x|\theta)\pi(\theta;\gamma)d\theta$.

As pointed out before, the functional form of $g(x;\gamma)$ in most cases is not available for computation. Since γ is the same involved in $q(x;\gamma)$ as in $p(x;\gamma)$, we propose to estimate γ by a nonparametric maximum likelihood estimation method. Similar ideas have been used in Kristensen and Shin (2012) for estimating unknown parameters involved in a class of fully parametric models.

For each given γ, suppose that we can sample $\{(\theta_j = \theta_j(\gamma), x_j = x_j(\gamma)) : 1 \le j \le m\}$ from $f(x|\theta)\pi(\theta;\gamma)$. For each given γ, we then estimate $p(x;\gamma)$ by

$$p_m(x;\gamma) = \frac{1}{mh^r}\sum_{j=1}^{m} L\left(\frac{x_j(\gamma) - x}{h}\right), \tag{5.2}$$

where r is the dimensionality of $x = (x_1, \cdots, x_r)^\tau$, $L(\cdot)$ is a probability density function defined on R^r, and h is a bandwidth parameter.

Assume that the data set $\{X_i : i = 1, 2, \cdots, n\}$ available for us is a sequence of i.i.d. random variables. Define a normalised log–likelihood function of the form

$$L_n(\gamma) = \frac{1}{\nu_n}\sum_{i=1}^{n} \log(p(X_i;\gamma)), \tag{5.3}$$

where $\nu_n \to \infty$ is a sequence of positive real numbers.

We estimate γ by $\gamma_n = \arg\max_{\gamma\in\Gamma} L_n(\gamma)$, in which Γ is a subset of R^c with c being the dimensionality of γ.

The corresponding version for $p_m(x;\gamma)$ is then defined as

$$L_{mn}(\gamma) = \frac{1}{\nu_n}\sum_{i=1}^{n} \log(p_m(X_i;\gamma)), \tag{5.4}$$

and γ is then estimated by $\gamma_{mn} = \arg\max_{\gamma\in\Gamma} L_{mn}(\gamma)$. Note that in practice, $L_{mn}(\gamma)$ may need to replaced by a truncated form. Such an issue is discussed in Remark A.1 just after the proof of Theorem 5.1 in the Appendix below.

In order to study asymptotic properties for γ_n and γ_{mn}, we need to introduce the following notation:

$$\begin{aligned}
&S_n(\gamma) = \frac{\partial L_n(\gamma)}{\partial \gamma}, \quad H_n(\gamma) = \frac{\partial^2 L_n(\gamma)}{\partial \gamma \partial \gamma'},\\
&G_{ni}(\gamma) = \frac{\partial^3 L_n(\gamma)}{\partial \gamma \partial \gamma' \partial \gamma_i}, \quad I_n(\gamma) = \frac{1}{\nu_n}\sum_{i=1}^{n} E\left[\frac{\partial \log(p(X_1;\gamma)}{\partial \gamma}\frac{\partial \log(p(X_1;\gamma)}{\partial \gamma'}\right],\\
&l_n(\gamma) = \text{diag}(I_n(\gamma)) = \text{the diagonal elements of matrix } I_n(\gamma),\\
&U_n(\gamma) = l_n^{-1}(\gamma)S_n(\gamma), \quad V_n(\gamma) = l_n^{-\frac{1}{2}}(\gamma)H_n(\gamma)l_n^{-\frac{1}{2}}(\gamma) \quad \text{and}\\
&W_{ni}(\gamma) = l_n^{-\frac{1}{2}}(\gamma)G_{ni}(\gamma)l_n^{-\frac{1}{2}}(\gamma). \qquad (5.5)
\end{aligned}$$

We are now able to introduce the following assumptions.

Assumption 5.1: (i) The parameter space is given by a sequence of local neighbourhoods: $\Gamma_n = \{\gamma : ||\sqrt{l_n(\gamma)}\,(\gamma - \gamma_0)\,|| \leq \epsilon\} \subset R^r$, where γ_0 is the true value of γ, $\epsilon > 0$ and $l_n^{-1}(\gamma) = O_P(1)$.

(ii) Let the data set $\{X_i\}$ be a sequence of i.i.d. random variables having the same density function as $p(x;\gamma)$. Let $p(x;\gamma)$ and $p_2(x;\gamma) = \int f_x^{(2)}(x|\theta)\pi(\theta;\gamma)d\theta$ satisfy

$$\sup_{x\in R^r}\sup_{\gamma\in\Gamma_n}\left(p^{-1}(x;\gamma)\left[1 + p_2(x;\gamma)\right]\right) < \infty.$$

(iii) $L_n(\gamma)$ is three times differentiable with its derivatives satisfying:

(a) $\left(\sqrt{\nu_n}U_n(\gamma_0), V_n(\gamma_0)\right) \to_D (U_\infty, V_\infty)$, where U_∞ and V_∞ are random variables with $P(V_\infty < 0) = 1$; and

(b) $\max_{1\leq i\leq r}\sup_{\gamma\in\Gamma_n}||W_{ni}(\gamma)|| = O_P(1)$.

Assumption 5.2: (i) For each given γ, $\{x_j = x_j(\gamma)\}$ is a sequence of i.i.d. random variables. The function $x_j(\gamma)$ is a differentiable with respect to γ and $\dot{x}_j(\gamma)$ denotes the first–order derivative with $\max_{j\geq 1} E\left[||\dot{x}_j(\gamma)||^2\right] < \infty$ for each given γ.

(ii) Define

$$\begin{aligned}
B_1(\gamma) &= \sup_{x\in R^d} E\left[\dot{x}_1(\gamma)|x_1(\gamma) = x\right]p(x;\gamma) \quad \text{and}\\
B_2(\gamma) &= \sup_{x\in R^d} ||x||^{\delta_0}\; E\left[\dot{x}_1(\gamma)|x_1(\gamma) = x\right]p(x;\gamma)
\end{aligned}$$

for some $\delta_0 \geq r$. There are some $0 < C_1, C_2 < \infty$ such that $\sup_{\gamma\in\Gamma} B_i(\gamma) \leq C_i < \infty$ for $i = 1, 2$.

Assumption 5.3: (i) Let $L(\cdot)$ be a symmetric and bounded probability kernel function.

(ii) There are some constants $0 < C, C_1, C_2 < \infty$ and $\mu > 1$ such that $L(u)$ is differentiable with $\left|\left|\frac{\partial L(u)}{\partial u}\right|\right| \leq C_1$ and $\left|\left|\frac{\partial L(u)}{\partial u}\right|\right| \leq C_2\; |||u||^{\mu}$ when $||u|| \geq C$. In addition, $L(u) \leq C_3\; ||u||^{\mu}$ when $||u|| \geq C$ for some $0 < C_3 < \infty$.

(iii) The bandwidth h satisfies $h \to 0$, $mh^r \to \infty$ and $\frac{\log(m)}{m\;h^r} \to 0$ as $m \to \infty$.

(iv) The sequence ν_n satisfies $nh^2 = o\left(\sqrt{\nu_n}\right)$ and $\frac{n^2 \log(m)}{mh^r} = o(\nu_n)$ as $(m, n) \to (\infty, \infty)$.

Assumptions 5.1–5.3 are standard for this kind of problem. For example, Assumption 5.1 is similar to Assumptions C1 and C3 of Kristensen and Shin (2012). Assumption 5.1 can be simplified to a set of standard conditions that are similar to the conditions of Theorems 4.1.2–4.1.4 of Amemiya (1985) for the standard maximum likelihood estimation. Assumptions 5.2 and 5.3(i)(ii)(iii) are simplified versions of A.2–A.5 of Kristensen (2009). Assumption 5.3(iv) imposes some conditions on the rates of convergence of (m, n). For example, when $\nu_n = n$, it simply requires $nh^4 \to 0$ and $\frac{n \log(m)}{mh^r} \to 0$, which can be satisfied with many different choices of (m, n). While the conditions may not be the weakest possible, they are sufficient for the establishment of the main theorem; its proof is given in Section 9.7.

Theorem 5.1: *Let Assumptions 5.1–5.3 hold. Then, as $n \to \infty$*

$$\sqrt{\nu_n}\, l_n^{\frac{1}{2}}(\gamma_0)\,(\gamma_{mn} - \gamma_0) \to_D \xi, \tag{5.6}$$

where $\xi = -\frac{U_\infty}{V_\infty}$ with (U_∞, V_∞) being defined in Assumption 5.1(iii).

Theorem 5.1 shows that γ_{mn} is still able to achieve the standard rate of convergence of an order of $n^{-\frac{1}{2}}$ when $\nu_n = n$, even though a kernel estimation is involved in the construction of the nonparametric maximum likelihood function. This is basically because of the following derivations:

$$\begin{aligned}\sqrt{\nu_n}\, l_n^{\frac{1}{2}}(\gamma_0)\,(\gamma_{mn} - \gamma_0) &= \sqrt{\nu_n}\, l_n^{\frac{1}{2}}(\gamma_0)\,(\gamma_{mn} - \gamma_n) + \sqrt{\nu_n}\, l_n^{\frac{1}{2}}(\gamma_0)\,(\gamma_n - \gamma_0)\\ &= \sqrt{\nu_n}\, l_n^{\frac{1}{2}}(\gamma_0)\,(\gamma_n - \gamma_0) + o_P(1) \to_D \xi\end{aligned} \tag{5.7}$$

as shown in Section 9.7 below. With a suitable rate of convergence for γ_{mn} to γ, the corresponding versions of $g(x; \gamma)$ or $g(T_n; \gamma)$ may be consistently estimated in the same way as in Sections 2–4.

We now conclude the establishment of the theory of this paper. In Sections 6 and 7 below, we will evaluate the proposed theory and estimation methods using both simulated and real data examples.

6 Numerical Evidence

This section evaluates the finite sample performance of the theory and the proposed estimation methods. Examples 6.1–6.3 consider the case where $\{(x_j, \theta_j)\}$ can be sampled directly from $f(x|\theta)\pi(\theta)$ and $\{X_i; 1 \le i \le n\}$ is drawn from $p(x)$. Example 6.4 consider the case where an MCMC algorithm is used for sampling.

Example 6.1: Consider the case where $f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$ for $-\infty < x < \infty$, and $\pi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}}$. This implies $g(x) = \frac{x}{2}$.

In this example and Example 6.2 below, we use $(m, n) = (2000, 200)$. Figures 1 and 2 below give the corresponding estimates of $g(x)$ for Example 6.1.

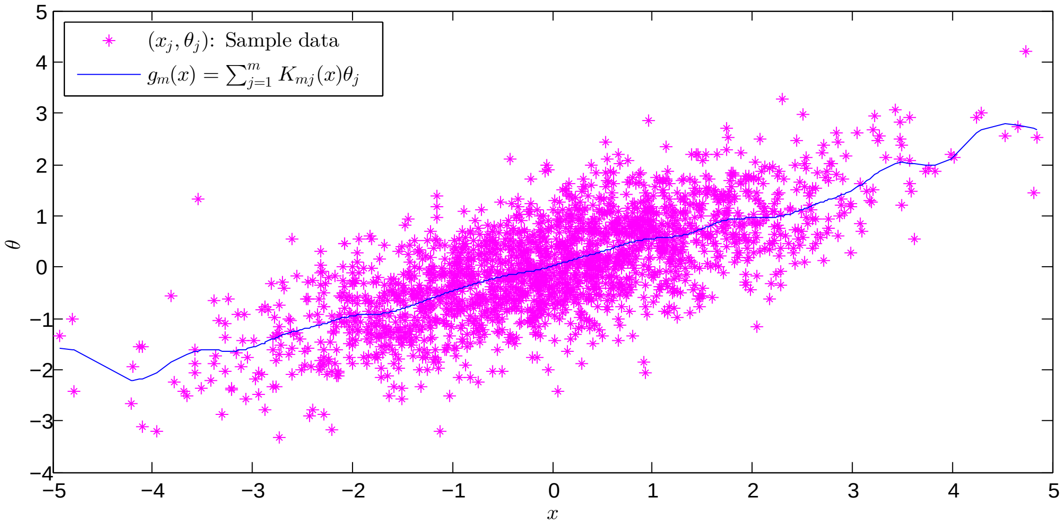

Figure 1: $h_{cv} = 0.2204$, $g_m(x) = \sum_{j=1}^{m} K_{mj}(x)\theta_j$

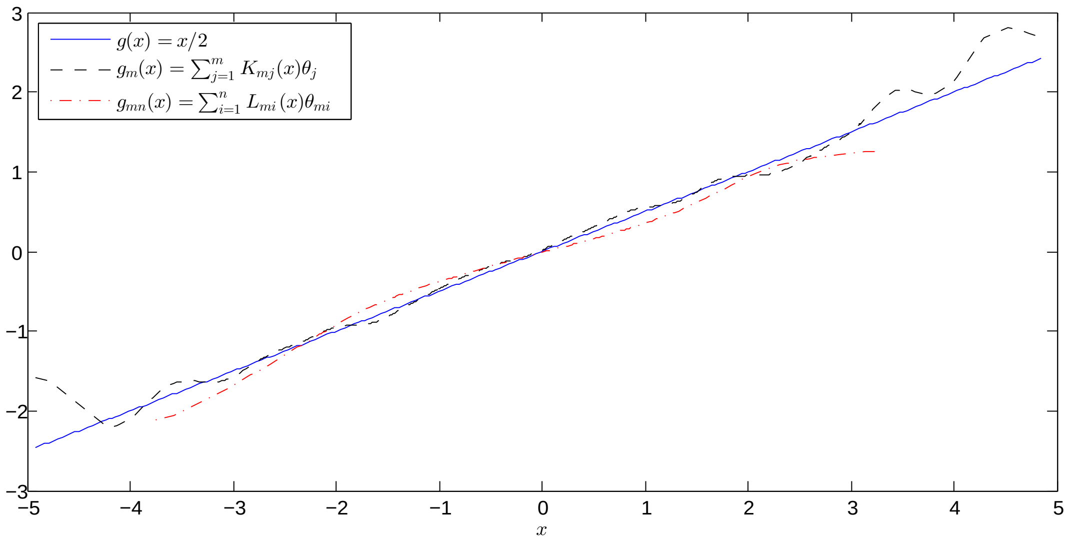

Figure 2: For $g_{mn}(x) = \sum_{i=1}^{n} L_{mi}(x)\theta_{mi}$, $h_{cv} = 0.6720$; For $g_m(x) = \sum_{j=1}^{m} K_{mj}(x)\theta_j$, $h_{cv} = 0.453$, X_i is drawn from $p(x) = N(0, 2)$.

Example 6.2: Consider the case where $f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$ for $-\infty < x < \infty$, and $\pi(\theta) = \frac{1}{6} I[-3 \le \theta \le 3]$. This implies

$$p(x) = \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta = \frac{1}{6}\left(\Phi(x+3) - \Phi(x-3)\right)$$

$$q(x) = \int_{-\infty}^{\infty} \theta f(x|\theta)\pi(\theta)d\theta = \frac{1}{6\sqrt{2\pi}}\left(e^{-\frac{(x+3)^2}{2}} - e^{-\frac{(x-3)^2}{2}}\right) + x\, p(x), \tag{6.1}$$

which implies $g(x) = \frac{q(x)}{p(x)} = x + \frac{1}{\sqrt{2\pi}} \cdot \frac{\left(e^{-\frac{(x+3)^2}{2}} - e^{-\frac{(x-3)^2}{2}}\right)}{\Phi(x+3) - \Phi(x-3)}$, in which $\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv$.

Figures 3 and 4 below give the corresponding estimates of $g(x)$ for Example 6.2.

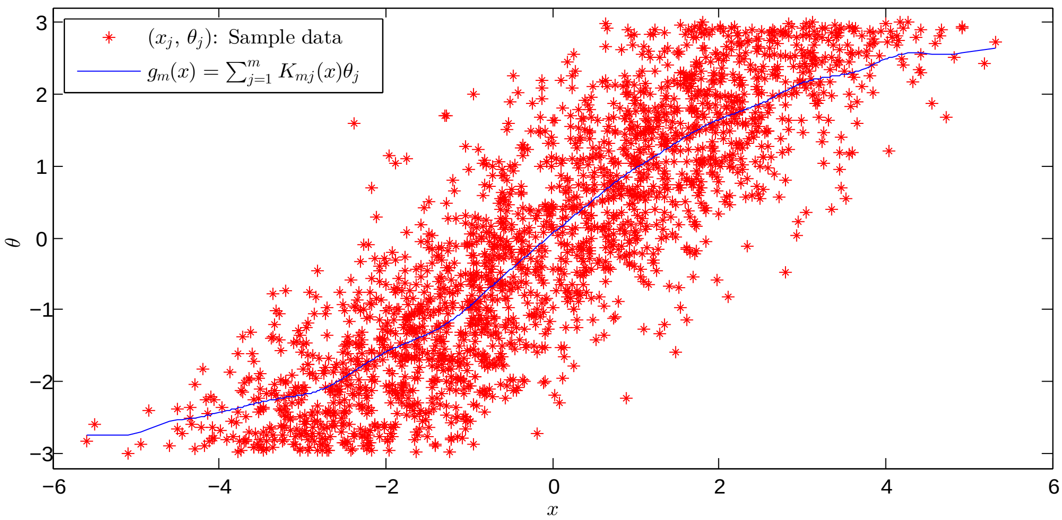

Figure 3: $h_{cv} = 0.2946$, $g_m(x) = \sum_{j=1}^{m} K_{mj}(x)\theta_j$

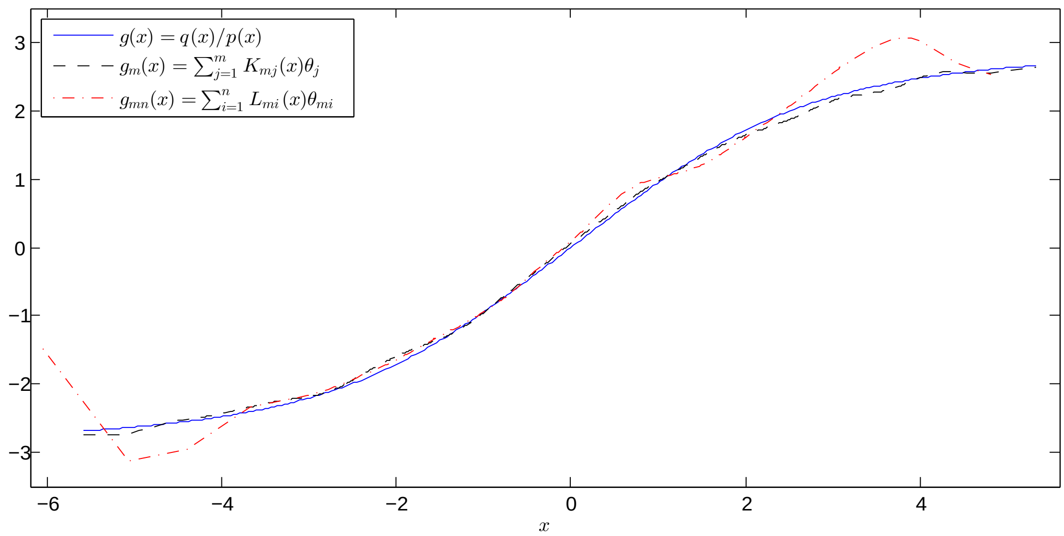

Figure 4: For $g_{mn}(x) = \sum_{i=1}^{n} L_{mi}(x)\theta_{mi}$, $h_{cv} = 0.4812$; For $g_m(x) = \sum_{j=1}^{m} K_{mj}(x)\theta_j$, $h_{cv} = 0.2946$, X_i is drawn from $p(x)$.

Figures 1–4 show that the proposed estimates $g_m(x)$ and $g_{mn}(x)$ are both very close to the true function $g(x)$.

Example 6.3: Let $X_i = \theta_0 + e_i$, where $e_i \sim N(0, 1)$. Parameter θ can be estimated by $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, and $\theta_0 = 1$. In this case, we have $g(T_n) = \frac{n}{1+n}T_n$.

Let $\theta_j \sim N(0, 1)$ and generate T_{nj} by a mean model of the form:

$$T_{nj} = \theta_j + \sigma_n\, \xi_j,\ j = 1, 2, \cdots, m, \tag{6.2}$$

where $\theta_j \sim N(0, 1)$, $\xi_j \sim N(0, 1)$ and $\sigma_n^2 = \frac{1}{n}$.

To evaluate the finite sample performance of the proposed estimates, we use the following estimators:

$$\widetilde{g}_m(T_n) = \frac{\sum_{j=1}^m \theta_j f(T_n|\theta_j)}{\sum_{j=1}^m f(T_n|\theta_j)}, \quad \widehat{g}_m(T_n) = \frac{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)\theta_j}{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)}, \tag{6.3}$$

$$\widehat{g}_m^*(T_n) = \frac{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)\theta_{Nj}^*}{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)}, \quad \overline{g}_N(T_n) = \frac{1}{N}\sum_{i=1}^N \theta_i(T_n). \tag{6.4}$$

where the last one denotes the sample mean of the direct MCMC draws, and $K(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ and h is a bandwidth chosen by the leave–one–out cross–validation method.

The main task is summarised as follows. Consider **Case A**: $(m,n) = (1000, 100)$; **Case B**: $(m,n) = (2000, 200)$. Let the number of replications, $M = 1000$.

- Calculate $\text{bias}_1 = \left|\overline{T}_n - \theta\right|$ and $\text{std}_1 = \sqrt{\frac{1}{M}\sum_{i=1}^M \left(T_n(i) - \overline{T}_n\right)^2}$, where $\overline{T}_n = \frac{1}{M}\sum_{i=1}^M T_n(i)$, and $T_n(i)$ is the value at i–th replication of $T_n = \frac{1}{n}\sum_{i=1}^n X_i$;
- Calculate $\text{bias}_2 = \left|\overline{\widetilde{g}}_m(\overline{T}_n) - g(\overline{T}_n)\right|$ and $\text{std}_2 = \sqrt{\frac{1}{M}\sum_{i=1}^M \left(\widetilde{g}_{mi}(\overline{T}_n) - \overline{\widetilde{g}}_m(\overline{T}_n)\right)^2}$; and
- Calculate $\text{bias}_3 = \left|\overline{\widehat{g}}_m(\overline{T}_n) - g(\overline{T}_n)\right|$ and $\text{std}_3 = \sqrt{\frac{1}{M}\sum_{i=1}^M \left(\widehat{g}_{mi}(\overline{T}_n) - \overline{\widehat{g}}_m(\overline{T}_n)\right)^2}$.

Table 6.1: The biases and standard deviations

case A: $(m,n) = (1000, 100)$				case B: $(m,n) = (2000, 200)$			
bias_1	0.0013	std_1	0.1761	bias_1	0.0007	std_1	0.1179
bias_2	0.0297	std_2	0.0303	bias_2	0.0152	std_2	0.0159
bias_3	0.0189	std_3	0.0097	bias_3	0.0094	std_3	0.0056

Table 6.1 shows that $g_{mk}(\cdot)$ is the best performer in terms of the finite sample performance of the standard deviations, while T_n has the smallest bias in each case. This is because T_n is an unbiased estimator of θ.

In Example 6.4 below, we will consider the case where resamples are used in the nonparametric kernel estimation.

Example 6.4: Simulate $X_i = \theta_0 + e_i$ with $e_i \sim N(0,1)$ for $i = 1, 2, \cdots, n$, where $\theta_0 = 1$. Let $T_n = \frac{1}{n}\sum_{i=1}^n X_i$ and generate T_{nj} as in Example 6.1. Generate θ_{ij} from $f(\theta|T_{nj})$ for $i = 1, 2, \cdots, N$ and each fixed j for $1 \le j \le m$.

We use the following proposal density functions for the generations of the resamples.

- **Mixture density 1: Proposal density**

  $f(\theta|T_n) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right) + \frac{1}{2\sqrt{2\pi T_n^2}} \exp\left(-\frac{\theta^2}{2T_n^2}\right)$, implying $g(T_n) = 0$.

- **Mixture density 2: Proposal density**

  $f(\theta|T_n) = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(\theta-1)^2}{2}\right) + \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(\theta-T_n)^2}{2}\right)$, implying $g(T_n) = 0.5 + 0.5T_n$.

- **Mixture density 3: Proposal density**

  $f(\theta|T_n) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right) + \frac{1}{2}(1+T_n^2) \exp\left(-(1+T_n^2)\theta\right)$, implying $g(T_n) = \frac{1}{\sqrt{2\pi}} + \frac{1}{2(1+T_n^2)}$.

As in Example 6.3, we use $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and choose h by the leave–one–out cross–validation method. Let the number of replications, $M = 1000$. We consider two cases: **Case A**: $(m, n, N) = (1000, 100, 1000)$, and **Case B**: $(m, n, N) = (2000, 200, 2000)$.

Define $\overline{g}_{Ni}(T_n(i)) = \frac{1}{N}\sum_{j=1}^{N} \theta_j(T_n(i))$ and $\overline{\overline{g}}_N(\overline{T}_n) = \frac{1}{M}\sum_{i=1}^{M} \overline{g}_{Ni}(T_n(i))$, where $\overline{T}_n = \frac{1}{M}\sum_{i=1}^{M} T_n(i)$, and $T_n(i)$ is the value at i–th replication of $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

- Calculate $\text{bias}_4 = \left|\overline{\overline{g}}_N(\overline{T}_n) - g(\overline{T}_n)\right|$ and $\text{std}_4 = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\overline{g}_{Ni}(T_n(i)) - \overline{\overline{g}}_N(\overline{T}_n)\right)^2}$;
- Calculate $\text{bias}_5 = \left|\overline{\widehat{g}}_m(\overline{T}_n) - g(\overline{T}_n)\right|$ and $\text{std}_5 = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\widehat{g}_{mi}(\overline{T}_n) - \overline{\widehat{g}}_m(\overline{T}_n)\right)^2}$; and
- Calculate $\text{bias}_6 = \left|\overline{\widehat{g}}^*_m(\overline{T}_n) - g(\overline{T}_n)\right|$ and $\text{std}_6 = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\widehat{g}^*_{mi}(\overline{T}_n) - \overline{\widehat{g}}^*_m(\overline{T}_n)\right)^2}$.

Table 6.2: The biases and standard deviations

Case A				Case B			
Mixture density 1							
bias_4	0.0048	std_4	0.0317	bias_4	0.0022	std_4	0.0261
bias_5	0.0185	std_5	0.0089	bias_5	0.0105	std_5	0.0053
bias_6	0.0011	std_6	0.0014	bias_6	0.0006	std_6	0.0007
Mixture density 2							
bias_4	0.0069	std_4	0.0281	bias_4	0.0031	std_4	0.0206
bias_5	0.0192	std_5	0.0092	bias_5	0.0115	std_5	0.0050
bias_6	0.0009	std_6	0.0008	bias_6	0.0004	std_6	0.0005
Mixture density 3							
bias_4	0.0053	std_4	0.0328	bias_4	0.0038	std_4	0.0239
bias_5	0.0178	std_5	0.0096	bias_5	0.0094	std_5	0.0048
bias_6	0.0012	std_6	0.0011	bias_6	0.0007	std_6	0.0006

The estimated biases and standard deviations are given in Table 6.2 above. Table 6.2 shows that the nonparametric estimate based on the resamples has the smallest standard deviation in each case, and $g_{mk}(\cdot)$ is better than the sample average of the resamples in terms of the standard deviations. In addition, Table 6.2 also shows that the biases and standard deviations are smaller for the case of $(N, m, n) = (2000, 2000, 200)$ than those for the case of $(N, m, n) = (1000, 1000, 100)$.

In the following example, we consider estimating unknown parameters involved in a GARCH model before we use the same model in an empirical evaluation.

7 GARCH model estimation and implementation

Consider a GARCH (1,1) model of the form:

$$
\begin{aligned}
y_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \quad t = 1, 2, \cdots, n, \\
\sigma_t^2 &= b_0 + b_1 y_{t-1}^2 + b_2 \sigma_{t-1}^2,
\end{aligned}
$$

where $\theta = (b_0, b_1, b_2)'$ denotes a vector of unknown parameters. Let T_n denote the maximum likelihood estimates of θ. Our aim is to estimate the conditional mean $E(\theta|T_n)$. Now we consider the following estimation methods.

We compute the maximum likelihood estimates of b_0, b_1 and b_2 based on a given data series and denote the estimates as $\overline{T}_n = (\widehat{b}_0, \widehat{b}_1, \widehat{b}_2)'$.

We then propose the following nonparametric kernel estimation method:

- First, we simulate $\theta_j = (b_j^0, b_j^1, b_j^2)$, for $j = 1, 2, \cdots, m$, from the prior density $\pi(\theta)$. We assume $b_0 \sim \text{Uniform}(0, 1)$, $b_2 \sim \text{Uniform}(0, 1)$ and $b_1 \sim \text{Uniform}(0, 1 - b_2)$.

- Second, we simulate $T_{n,j}$ from a limiting distribution of the maximum likelihood estimator of θ, which is a normal distribution with mean θ_j and variance $\frac{1}{n}\Sigma$, where Σ can be computed with a closed–form expression provided by Ma (2008). When n is large enough, we have
$$T_{n,j} = (T_{n,j}^0, T_{n,j}^1, T_{n,j}^2) \sim N\left(\theta_j, \frac{1}{n}\Sigma\right).$$

- Define $g(T_n) = \left(g_0\left(T_n^0\right), g_1\left(T_n^1\right), g_2\left(T_n^2\right)\right)^{\tau}$, in which $g_0(T_n^0) = E(b_0|T_n)$, $g_1(T_n^1) = E(b_1|T_n)$ and $g_2(T_n^2) = E(b_2|T_n)$. Based on $(\theta_j, T_{n,j})$, we then estimate $g(T_n)$ by the nonparametric Nadaraya–Waston kernel estimate, denoted by $g_{NW}(\cdot)$.

The conventional importance sampling estimate is denoted by $g_{IM}(\cdot)$. To distinguish the nonparametric kernel estimation based on MCMC samples from the conventional Bayesian sample mean, we use MC1 to denote the latter, while MC2 to denote the former.

MC1 method is summarised as follows.

(i) Assume $\pi(\theta|T_n) = (1+T_{n0}^2)I[1 < b_0 < \frac{1}{1+T_{n0}^2}](1+T_{n1}^2)I[1 < b_2 < \frac{1}{1+T_{n2}^2}]I[0 < b_1 < 1-b_2]$.

(ii) We can sample θ_j from $\pi(\theta|\overline{T}_n)$, for $j = 1, 2, \cdots, m$, and then we approximate the conditional mean of the form $g_{MC1}(\cdot) = \frac{1}{m}\sum_{j=1}^m \theta_j$.

MC2 method is described as follows.

(i) Given $T_{n,j}$, we simulate a sample $\theta_{j1}, \theta_{j2}, \cdots, \theta_{jN}$ from $\pi(\theta_j|T_{n,j})$ and then we compute the average denoted as $\overline{\theta}_j = \frac{1}{N}\sum_{i=1}^N \theta_{ji}$.

(ii) Based on $\left(\overline{\theta}_j, T_{n,j}\right)$, we then estimate $g(T_n)$ by the conventional Nadaraya–Waston method denoted as $g_{MC2}(\cdot)$.

We then consider two types of error distributions in Sections 7.1 and 7.2 below.

7.1 Gaussian error density

We simulated 1000 samples from the following GARCH (1,1) model:

$$y_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \quad t = 1, 2, \cdots, n$$
$$\sigma_t^2 = b_0 + b_1 y_{t-1}^2 + b_2 \sigma_{t-1}^2,$$

where $\theta = (b_0, b_1, b_2)' = (0.5, 0.15, 0.7)'$.

- **Case A:** $n = 500$ and $m = 5000$.
- **Case B:** $n = 1000$ and $m = 10000$.

For each replication, we estimate θ by $\widehat{\theta}$ and the bias and standard deviation for **Case A** and **Case B** are given in Tables 7.1 and 7.2, respectively.

7.2 Chi–squared error density

We simulated 1000 samples from the following GARCH (1,1) model:

$$y_t = \sigma_t \varepsilon_t, \quad \varepsilon_t = \frac{u_t - 1}{\sqrt{2}}, \quad u_t \sim \chi^2(1), \quad t = 1, 2, \cdots, n$$
$$\sigma_t^2 = b_0 + b_1 y_{t-1}^2 + b_2 \sigma_{t-1}^2,$$

where $n = 1000$ and $\theta = (b_0, b_1, b_2)' = (0.5, 0.15, 0.7)'$.

For each replication, we estimate θ by $\widehat{\theta}$ and the bias and standard deviation for **Case A** and **Case B** are given in Tables 7.3 and 7.4, respectively. For $i = 0, 1, 2$, define

$$\text{bias}_{b_i} = \frac{1}{1000}\sum_{r=1}^{1000}\left(\widehat{b}_{i,r} - b_i\right), \quad \text{bias}_\theta = \text{bias}_{b_0} + \text{bias}_{b_1} + \text{bias}_{b_2},$$
$$\sigma_{ij} = \frac{1}{1000}\sum_{r=1}^{1000}\left(\widehat{b}_{i,r} - \overline{\widehat{b}_i}\right)\left(\widehat{b}_{j,r} - \overline{\widehat{b}_j}\right), \quad \text{std}_\theta = \sqrt{\sum_{i,j}\sigma_{ij}^2}.$$

Tables 7.1–7.4 below give the biases and standard deviations of the maximum likelihood (ML) estimate, the importance sampling average (IMS) estimate, the simple MC mean (MC1) estimate, the nonparametric NW kernel (NW) estimate based on direct samples, and the nonparametric NW kernel estimate based on MCMC samples (MC2).

Table 7.1: The biases and standard deviations for Case A with Normal error

Case A	bias				std			
	bias_{b_0}	bias_{b_1}	bias_{b_2}	bias_{θ}	std_{b_0}	std_{b_1}	std_{b_2}	std_{θ}
MLE	0.1117	0.0043	−0.0396	0.0764	0.3540	0.0537	0.1396	0.2456
IMS	0.1174	0.0027	−0.0390	0.0811	0.1389	0.0487	0.0746	0.1010
MC1	−0.1304	0.1760	−0.3520	−0.3064	0.0886	0.0210	0.0418	0.0693
NW	0.0006	−0.0042	−0.1464	−0.1499	0.0042	0.0499	0.0766	0.0248
MC2	−0.1158	0.1632	−0.3265	−0.2791	0.0039	0.0132	0.02630	0.0167

Table 7.2: The biases and standard deviations for Case B with Normal error

Case B	bias				std			
	bias_{b_0}	bias_{b_1}	bias_{b_2}	bias_{θ}	std_{b_0}	std_{b_1}	std_{b_2}	std_{θ}
MLE	0.0452	0.0009	−0.0151	0.0310	0.2068	0.0363	0.0850	0.1449
IMS	0.0917	0.0019	−0.0294	0.0642	0.1309	0.0344	0.0624	0.0914
MC1	−0.1143	0.1796	−0.3593	−0.2940	0.0600	0.0135	0.0267	0.0476
NW	−0.0080	−0.0072	−0.1542	−0.1694	0.0112	0.0349	0.0469	0.0119
MC2	−0.1091	0.1611	−0.3221	−0.2701	0.0034	0.0083	0.0165	0.0073

Table 7.3: The biases and standard deviations for Case A with Chi–squared error

Case A	bias				std			
	bias_{b_0}	bias_{b_1}	bias_{b_2}	bias_{θ}	std_{b_0}	std_{b_1}	std_{b_2}	std_{θ}
MLE	0.1667	0.0360	−0.0788	0.1238	0.5112	0.1598	0.2236	0.3801
IMS	0.1435	0.0147	−0.0875	0.0707	0.1967	0.1060	0.1860	0.1492
MC1	−0.1404	0.1708	−0.3417	−0.3112	0.1135	0.0314	0.0627	0.0866
NW	−0.0048	0.0232	−0.1712	−0.1528	0.0277	0.1253	0.1768	0.0169
MC2	−0.1168	0.1611	−0.3223	−0.2780	0.0120	0.0272	0.0545	0.0097

Table 7.4: **The biases and standard deviations for Case B with Chi-squared error**

Case B	bias				std			
	bias_{b_0}	bias_{b_1}	bias_{b_2}	bias_{θ}	std_{b_0}	std_{b_1}	std_{b_2}	std_{θ}
MLE	0.1128	0.0204	−0.0523	0.0808	0.3790	0.1071	0.1712	0.2731
IMS	0.1269	0.0131	−0.0589	0.0811	0.1923	0.0866	0.1301	0.1458
MC1	−0.1306	0.1742	−0.3484	−0.3048	0.0939	0.0251	0.0501	0.0716
NW	−0.0045	0.0126	−0.1725	−0.1644	0.0217	0.0956	0.1290	0.0235
MC2	−0.1103	0.1595	−0.3191	−0.2690	0.0065	0.0205	0.0411	0.0108

Note that the bandwidth used in either the nonparametric NW estimate or the nonparametric NW estimate based on the resample is chosen by the normal reference rule. Note also that Tables 7.1–7.4 show that in terms of the standard deviation performance, MC2, the nonparametric estimate based on the resample, in each case, outperforms its natural competitors.

In the following subsection, using a set of real data, we examine the finite–sample performance of NW and MC2 with their natural competitors of the unknown parameters involved in the GARCH model.

7.3 Real data

We downloaded the S&P 500 daily closing prices, p_t from http://finance.yahoo.com. The date t return is calculated as $y_t = \log(p_t/p_{t-1})$. We consider the case where S&P 500 daily returns are from the 4th of January 2007 to the 31st of May 2013 with the rolling period starting from the 3rd of January 2011. Figure 5 below gives the plot of y_t for this first period. To evaluate whether the inclusion of the period of the main global financial crisis has any impact on the finite–sample performance, we also consider the returns for the period of the 5th January 2009 to the 2nd September 2014 rolling from 2nd January 2013. Figure 6 below gives the plot of y_t for this second period.

Unlike in simulation, we cannot replicate the data. Instead, we propose a so–called forward–rolling method to use the consecutive sections of the data for the evaluation of standard deviations of the estimators proposed above. Let T denote the number of all the observations and n denote the rolling sample size. We consider the first case where $T = 1613$, the size of rolling samples, $n = 1007$, and the number of rolling samples, $R = T - n = 606$. We also consider the second case where $T = 1425$, the size of rolling samples, $n = 1005$, and the number of rolling samples, $R = T - n = 420$.

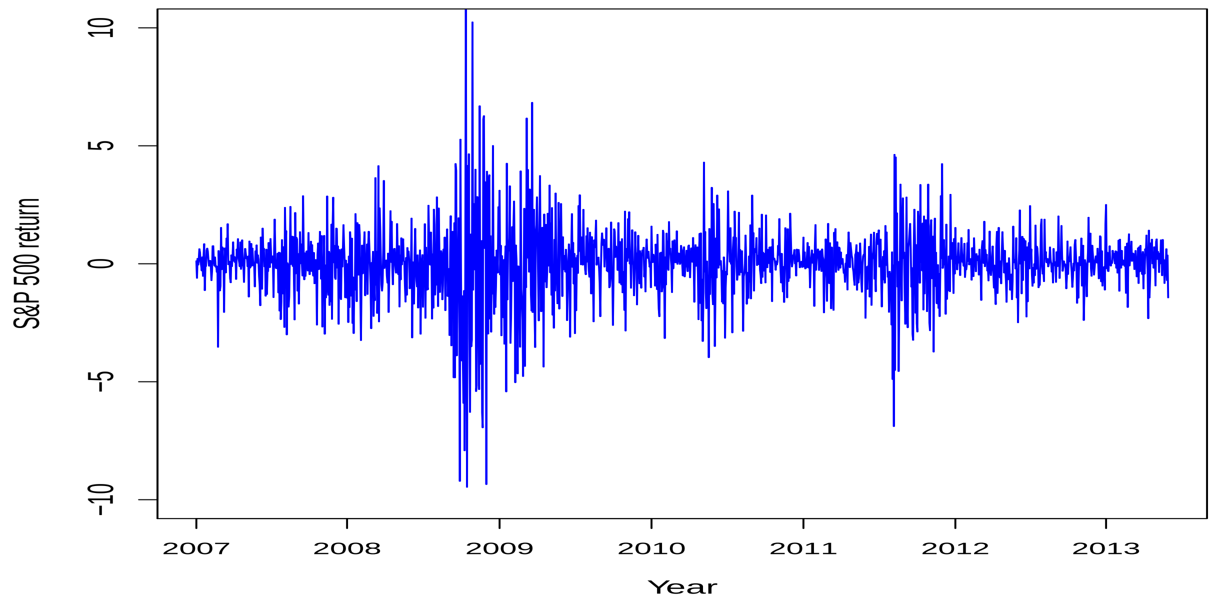

Figure 5: Plot of y_t for the period of the 4th of January 2007 to the 31st of May 2013.

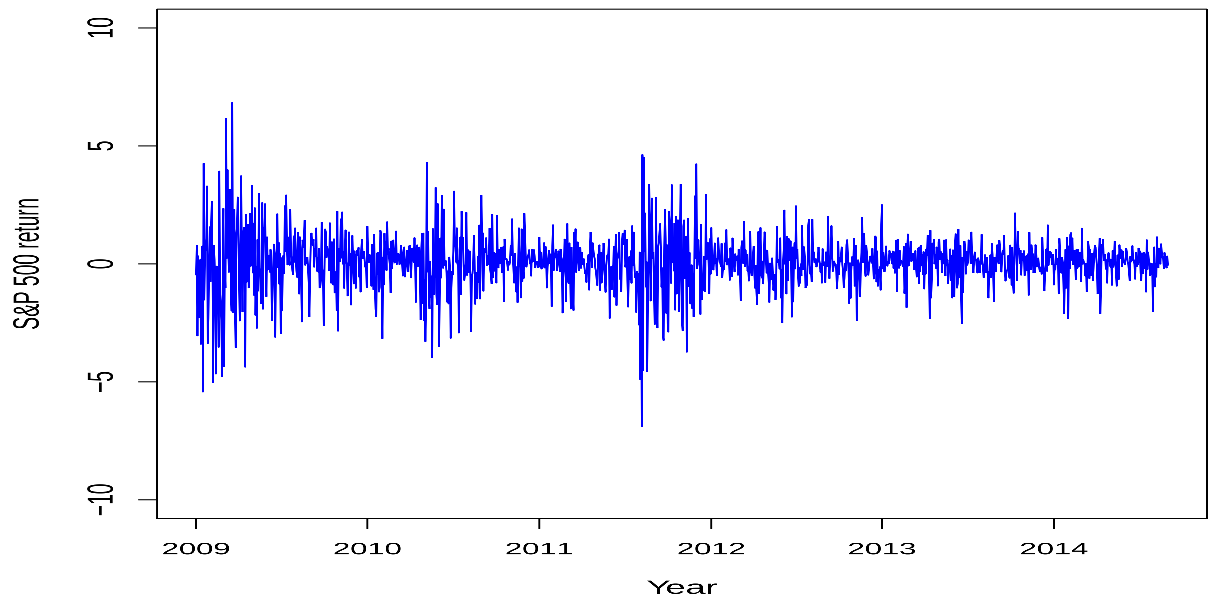

Figure 6: Plot of y_t for the period of the 5th January 2009 to the 2nd September 2014.

For the r–th rolling sample with $r = 1, 2, \cdots, T - n$, we compute parameter estimates denoted as $\widehat{b}_{i,r},\ i = 0, 1, 2$ using the five methods outlined in the previous section. We then compute forward–rolling standard deviation (FRSD) by

$$\mathrm{FRSD}_{b_i} = \sqrt{\frac{1}{T-n}\sum_{r=1}^{T-n}\left(\widehat{b}_{i,r} - \overline{\widehat{b}_i}\right)^2} \quad \text{and} \quad \mathrm{FRSD}_{\theta} = \sqrt{\sum_{i,j}\sigma_{ij}^2} \tag{7.1}$$

for $i = 0, 1, 2$, where $\overline{\widehat{b}_i} = \frac{1}{T-n}\sum_{r=1}^{T-n}\widehat{b}_{i,r}$ and σ_{ij} is as defined before for $i, j = 0, 1, 2$.

Table 7.5 below gives the Forward–rolling standard deviations (FRSDs) for the first period, while Table 7.6 below reports the FRSDs for the second period. For the nonparametric kernel based estimates: NW and MC2, the usual normal reference rule was used for the bandwidth choice.

Table 7.5: FRSDs of S&P 500 returns for the first period

Estimates	$FRSD_{b_0}$	$FRSD_{b_1}$	$FRSD_{b_2}$	$FRSD_{\theta}$
MLE	0.0028	0.0069	0.0084	0.0044
IMS	0.1203	0.0046	0.0623	0.0679
MC1	0.0029	0.0026	0.0028	0.0038
NW	0.0005	0.0062	0.0103	0.0036
MC2	0.00008	0.0020	0.0040	0.0013

Table 7.6: FRSDs of S&P 500 returns for the second period

Estimates	$FRSD_{b_0}$	$FRSD_{b_1}$	$FRSD_{b_2}$	$FRSD_{\theta}$
MLE	0.0029	0.0063	0.0117	0.0052
IMS	0.0168	0.0053	0.0242	0.0140
MC1	0.0029	0.0029	0.0038	0.0039
NW	0.0005	0.0059	0.0091	0.0027
MC2	0.0001	0.0018	0.0035	0.0009

While Column 4 of Figure 7.5 shows that MC1 has slightly smaller forward–rolling standard deviation than that of MC2, the last column of Table 7.5 shows that MC2 uniformly outperforms its natural competitors in terms of the proposed forward–rolling standard deviation. Similar conclusions can be made for Table 7.6. Without including the period of the main global financial crisis, there are some improvements on the FRSDs, particularly with the IMS method.

In summary, both the simulation and the real data evaluation show that MC2 has the smallest standard deviation, which supports that the large–sample theory is verifiable in the finite–sample situations.

8 Conclusions and Discussions

This paper has proposed some closed–form estimation and approximation methods for some infeasible computational issues. We have also developed some new asymptotic theory to support the proposed estimation and approximation methods. The proposed estimation and approximation theory has been evaluated by the simulated examples. Meanwhile, an empirical example has also been included to show that the proposed nonparametric estimation method based on resamples for the unknown parameters involved in the GARCH model has the smallest forward–rolling standard deviation.

There are several topics that are left for future research. The first topic is that issues regarding model overidentification will be carefully examined. The second topic is that new estimation and approximation methods are needed to deal with the case where there are structural breaks with the conditional distribution of X given θ. In this case, one may need to extend the work by Andrews and Fair (1988), and then Imbens and Kalyanaraman (2012) for our case. The third topic is how to deal with various estimation and approximation issues for the case where there is a type of endogeneity involved in the conditional distribution of X given θ. Estimation issues for unknown parameters and functions involved in general structural models will also be considered.

9 Appendix

9.1 Proofs of Theorems 2.1 and 2.2

Let $A_{mj}(x) = \frac{K\left(\frac{x_j-x}{h}\right)}{\sqrt{\sum_{k=1}^m K^2\left(\frac{x_k-x}{h}\right)}}$. We have

$$\sqrt{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)}\left(g_m(x)-g(x)\right) = \sqrt{\frac{\sum_{j=1}^m K^2\left(\frac{x_j-x}{h}\right)}{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)}}\cdot\sum_{j=1}^m A_{mj}(x)e_j$$
$$+\frac{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)\left(g(x_j)-g(x)\right)}{\sqrt{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)}}. \tag{10.1}$$

The converge of the first term of (10.1) in distribution to a Normal random variable follows directly from Assumptions 2.1 and 2.3(i). Assumption 2.1(ii)(iii) implies that $g(x) = E[\theta|x]$ is twice differentiable and the second–order derivative is continuous. Thus, the second term of equation (10.1) is of an order of $O\left(\sqrt{mh^{r+4}}\right)$.

To prove Theorem 2.2, we have a look at the following decomposition:

$$
\begin{aligned}
g_{mmn}(x) - g(x) &= g_{mn}(x) - g_m(x) + g_m(x) - g(x) \\
&= \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)\varepsilon_{mi}}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} + \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g_m(X_i) - g_m(x))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} + g_m(x) - g(x) \\
&\equiv B_{1mn}(x) + B_{2mn}(x) + g_m(x) - g(x), \qquad (10.2)
\end{aligned}
$$

where $B_{1mn}(x) = \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)\varepsilon_{mi}}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)}$ and $B_{2mn}(x) = \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g_m(X_i) - g_m(x))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)}$.

Let us first deal with $B_{1mn}(x)$. Under Assumptions 2.1–2.3, the standard conditions required for establishing the central limit theorem are satisfied. Thus, we have as $n \to \infty$

$$
\begin{aligned}
&\sqrt{\frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)}{\widehat{\sigma}_{mn}^2}} \sqrt{\frac{\sum_{i=1}^n L^2\left(\frac{X_i - x}{b}\right)}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)}} \cdot \sum_{i=1}^n \left(\frac{L\left(\frac{X_i - x}{b}\right)}{\sqrt{\sum_{i=1}^n L^2\left(\frac{X_i - x}{b}\right)}} \right) \varepsilon_{mi} \\
&\to_D \; N\left(0, \sigma^2(L)\right). \qquad (10.3)
\end{aligned}
$$

Let us then deal with $B_{2mn}(x)$. Observe that

$$
\begin{aligned}
B_{2mn}(x) &= \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g_m(X_i) - g_m(x))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} = \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g_m(X_i) - g(X_i))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} \\
&+ \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g(X_i) - g(x))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} + \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g(x) - g_m(x))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} \\
&\equiv C_{1mn}(x) + C_{2mn}(x) + C_{3mn}(x), \qquad (10.4)
\end{aligned}
$$

where we have used the decomposition: $g_m(X_i) - g_m(x) = g_m(X_i) - g(X_i) + g(X_i) - g(x) + g(x) - g_m(x)$, and $C_{3mn}(x) = g(x) - g_m(x)$.

Meanwhile, using the following decomposition:

$$
g_m(x) - g(x) = \sum_{j=1}^m K_{mj}(x)\theta_j - g(x) = \sum_{j=1}^m K_{mj}(x)e_j + \sum_{j=1}^m K_{mj}(x)\left(g(x_j) - g(x)\right) \qquad (10.5)
$$

we have

$$
\begin{aligned}
C_{1mn}(x) &= \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)(g_m(X_i) - g(X_i))}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} = \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)\left(\sum_{j=1}^m K_{mj}(X_i)e_j\right)}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} \\
&+ \frac{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)\left(\sum_{j=1}^m K_{mj}(X_i)\left(g(x_j) - g(X_i)\right)\right)}{\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right)} \equiv D_{1mn}(x) + D_{2mn}(x), \qquad (10.6)
\end{aligned}
$$

where $K_{mj}(x) = \frac{K\left(\frac{x_j - x}{h}\right)}{\sum_{k=1}^m K\left(\frac{x_k - x}{h}\right)}$.

Define $D_{3mn}(x) = \frac{1}{mhnb}\sum_{j=1}^m \left(\sum_{i=1}^n L\left(\frac{X_i - x}{b}\right) K\left(\frac{x_j - X_i}{h}\right)\right) e_j$ and let $E[e_1^2] = 1$ for notational simplicity. We then have

$$E\left[D_{3mn}^2(x)\right] = \frac{1}{m^2h^2n^2b^2}\left(\sum_{j=1}^m\sum_{i=1}^n K^2\left(\frac{x_j - X_i}{h}\right) L^2\left(\frac{X_i - x}{b}\right)\right)$$

$$+ \frac{2}{m^2h^{2r}n^2b^{2r}}\left(\sum_{j=1}^m\sum_{i=2}^n\sum_{k=1}^{i-1} K\left(\frac{x_j - X_i}{h}\right) K\left(\frac{x_j - X_k}{h}\right) L\left(\frac{X_i - x}{b}\right) L\left(\frac{X_k - x}{b}\right)\right)$$

$$\equiv \frac{1}{m^2h^{2r}n^2b^{2r}}\left(D_{4mn}(x) + 2D_{5mn}(x)\right). \tag{10.7}$$

Simple calculation implies as $(m, n) \to (\infty, \infty)$

$$E\left[D_{4mn}(x)\right] = \sum_{j=1}^m\sum_{i=1}^n E\left[K^2\left(\frac{x_j - X_i}{h}\right) L^2\left(\frac{X_i - x}{b}\right)\right]$$

$$= mn \int\int K^2\left(\frac{u - v}{h}\right) L^2\left(\frac{v - x}{b}\right) p(v)f(u)dudv$$

$$= mnh^rb^r\ (1 + o(1)) \cdot \int\int K^2(x)L^2(y)dydx \cdot p(x) \cdot f(x), \tag{10.8}$$

where $p(\cdot)$ and $f(\cdot)$ denote the density functions of X_i and x_j, respectively.

Similarly, we have as $\frac{h}{b} \to 0$

$$E\left[D_{5mn}(x)\right] = \sum_{j=1}^m\sum_{i=2}^n\sum_{k=1}^{i-1} E\left[K\left(\frac{x_j - X_i}{h}\right) K\left(\frac{x_j - X_k}{h}\right) L\left(\frac{X_i - x}{b}\right) L\left(\frac{X_k - x}{b}\right)\right]$$

$$= mn^2(1 + o(1)) \cdot \int\int\int K\left(\frac{u - v}{h}\right) K\left(\frac{u - w}{h}\right) L\left(\frac{v - x}{b}\right) L\left(\frac{w - x}{b}\right) q(v)q(w)f(u)$$

$$\times\ dudvdw = mn^2h^{2r}b^r\ (1 + o(1)) \cdot \int L^2(u)du \cdot q^2(x) \cdot p(x). \tag{10.9}$$

Equations (10.7)–(10.9) then imply

$$D_{3mn}(x) = O_P\left(\frac{1}{\sqrt{mb^r}}\right) \quad \text{and} \quad D_{1mn}(x) = O_P\left(\frac{1}{\sqrt{mb^r}}\right). \tag{10.10}$$

We now deal with $D_{2mn}(x)$ involved in (10.6). By Assumption 2.1(ii), using the standard derivation for the bias term (see Theorem 2.2 of Li and Racine 2007), we have

$$\sum_{j=1}^m K_{mj}(x)\left(g(x_j) - g(x)\right) = (1 + o_P(1))\ Q(x) \cdot h^2,$$

where

$$Q(x) = \frac{\int u^2K(u)du\ (1 + o(1))}{2} \cdot \sum_{j=1}^r\left(g^{(jj)}(x) + \frac{2g^{(j)}(x)f^{(j)}(x)}{f(x)}\right),$$

in which $f(\cdot)$ is the density function of $\{x_j\}$.

We then define $D_{6mn}(x) = \frac{h^2}{nb^r}\sum_{i=1}^n L\left(\frac{X_i-x}{b}\right)Q(X_i)$. As $(m,n)\to(\infty,\infty)$, we then have

$$\begin{aligned}
E\left[D_{6mn}^2(x)\right] &= \frac{h^4}{n^2b^{2r}}\left(\sum_{i=1}^n E\left[L^2\left(\frac{X_i-x}{b}\right)Q^2(X_i)\right]\right.\\
&+ \left. 2\sum_{i=2}^n\sum_{j=1}^{i-1} E\left[L\left(\frac{X_i-x}{b}\right)L\left(\frac{X_j-x}{b}\right)Q(X_i)Q(X_j)\right]\right)\\
&= \frac{h^4}{n^2b^{2r}}(1+o(1))Q^2(x)\left(nb^r\int L^2(u)du\cdot p(x)+n(n-1)b^{2r}p^2(x)\right)\\
&= h^4Q^2(x)p(x)\ (1+o(1))\left(\frac{1}{nb^r}\int L^2(u)du+p(x)\right), \qquad (10.11)
\end{aligned}$$

which implies

$$D_{6mn}(x) = O_P(h^2) \quad \text{and} \quad D_{2mn}(x) = O_P(h^2). \qquad (10.12)$$

Similarly, we have as $n\to\infty$

$$C_{2mn}(x) = \frac{\sum_{i=1}^n L\left(\frac{X_i-x}{b}\right)(g(X_i)-g(x))}{\sum_{i=1}^n L\left(\frac{X_i-x}{b}\right)} = b^2\ (1+O_P(1))\cdot R_n(x), \qquad (10.13)$$

where $R_n(x)$ is a continuous function.

Therefore, under Assumptions 2.1–2.3, equations (10.2)–(10.4), (10.10) and (10.12)–(10.13) complete the proof of Theorem 2.2.

9.2 Proofs of Theorems 2.3 and 2.4

The main ideas for the proofs of Theorems 2.3 and 2.4 are very similar to those for Theorems 2.1 and 2.2. In addition, the idea dealing with the multivariate case for Theorem 2.3 is also very similar to that for Theorem 2.4. Therefore, we just give the main idea for the proof of Theorem 2.3.

Observe that

$$\begin{aligned}
\sqrt{\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)}\ (g_m(x)-g(x)) &= \sqrt{\frac{\sum_{j=1}^m K^2\left(\frac{x_j-x}{h}\right)}{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)}}\cdot\sum_{j=1}^m W_{mj}(x)e_j\\
&+ \frac{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)(g(x_j)-g(x))}{\sqrt{\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)}}\\
&\equiv \sqrt{\frac{\sum_{j=1}^m K^2\left(\frac{x_j-x}{h}\right)}{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)}}\cdot J_{m1}(x)+\sqrt{\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)}\cdot J_{m2}(x), \qquad (10.14)
\end{aligned}$$

where $W_{mj}(x) = \frac{K\left(\frac{x_j-x}{h}\right)}{\sqrt{\sum_{j=1}^m K^2\left(\frac{x_j-x}{h}\right)}}$, $J_{m1}(x) = \sum_{j=1}^m W_{mj}(x)e_j$ and $J_{m2}(x) = \frac{\sum_{j=1}^m K\left(\frac{x_j-x}{h}\right)(g(x_j)-g(x))}{\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)}$.

In order to show that $J_{m1}(x)\to_D N(0,\Sigma)$, it suffices to show that $A^\tau J_{m1}(x)\to_D N(0, A^\tau\Sigma A)$ for any constant vector $A=(a_1,\cdots,a_d)^\tau$ satisfying $A^\tau A=1$. The proof of the latter follows trivially.

Meanwhile, the bias term $J_{m2}(x)$ may also be dealt with through computing the bias term $A^\tau J_{m2}(x)$, and this follows trivially. This therefore completes the proof of Theorem 2.3.

9.3 Proof of Theorem 3.1

Observe that

$$\sqrt{m\sum_{i=1}^{n} L\left(\frac{X_i-x}{h}\right)}\left(g_{mn}(x)-g(x)\right) = \frac{\sqrt{\sum_{i=1}^n L^2\left(\frac{X_i-x}{h}\right)}}{\sqrt{\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)}}$$

$$\times \frac{\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)\overline{e}_{mi}}{\sqrt{\sum_{i=1}^n L^2\left(\frac{X_i-x}{h}\right)}} + \frac{\sqrt{m}\ \sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)\left(g(X_i)-g(x)\right)}{\sqrt{\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)}}, \tag{10.15}$$

where $\overline{e}_{mi} = \frac{1}{\sqrt{m}}\sum_{j=1}^m e_{ij}$.

By Peligrad (1987), we have for some $0 < C < \infty$

$$E\left[\overline{e}_{mi}^4\right] \leq \frac{1}{m^2}\left(m^2\ E[e_{ij}^4] + \left(E\left[\sum_{j=1}^m e_{ij}\right]^2\right)^2\right) \leq C < \infty, \tag{10.16}$$

which, along with the conditions of Theorem 3.1 of this paper and Theorem 2.22 of Fan and Yao (2003), implies

$$\frac{1}{\sqrt{nh^r}\sigma_m(x)}\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)\overline{e}_{mi} \to_D N\left(0, \Sigma^2(x)\right), \tag{10.17}$$

where $\sigma_m^2(x) = \frac{1}{m}\sum_{j=1}^m E\left[e_{ij}^2|X_i=x\right] + \frac{2}{m}\sum_{j_1=2}^m\sum_{j_2=1}^{j_1-1} E\left[e_{ij_1}e_{ij_2}|X_i=x\right] \to \gamma(x) + 2\sum_{j=1}^\infty \gamma_j(x)$, and $\Sigma^2(x) = p(x)\cdot\int K^2(u)du$, in which $p(x)$ is the density of X_1.

The second term of equation (10.15) can be dealt with as before. Therefore, we have completed the proof of Theorem 3.1.

9.4 Proof of Theorem 3.2

The proof of the first part follows in the same way from equation (10.15). In a similar fashion to equation (10.16), by Assumption 6.4(ii), we have almost surely,

$$E\left[\overline{e}_{mi}^2|\mathcal{F}_{m,i-1}\right] = \frac{1}{m}\sum_{j=1}^m E\left[\varepsilon_{ij}^2|\mathcal{F}_{m,i-1}\right] + \frac{2}{m}\sum_{j_1=2}^m\sum_{j_2=1}^{j_1-1} E\left[\varepsilon_{i,j_1}\varepsilon_{i,j_2}|\mathcal{F}_{m,i-1}\right]$$

$$= \frac{1}{m}\sum_{j=1}^m E\left[\varepsilon_{ij}^2|\mathcal{F}_{m,i-1}\right] \to \sigma_\varepsilon^2, \tag{10.18}$$

$$E\left[\overline{e}_{mi}^4|\mathcal{F}_{m,i-1}\right] \leq \frac{1}{m^2}\left(m^2\ E[e_{ij}^4|\mathcal{F}_{m,i-1}] + \left(E\left[\left(\sum_{j=1}^m e_{ij}\right)^2|\mathcal{F}_{m,i-1}\right]\right)^2\right) \leq C < \infty,$$

which, along with an application of Theorem 3.2 of Gao and Phillips (2013), imply

$$\sqrt{\frac{\sum_{i=1}^n L^2\left(\frac{X_i-x}{h}\right)}{\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)}}\cdot\frac{\sum_{i=1}^n L\left(\frac{X_i-x}{h}\right)\overline{e}_{mi}}{\sqrt{\sum_{i=1}^n L^2\left(\frac{X_i-x}{h}\right)}}\to_D N\left(0,\Sigma_1^2\right), \tag{10.19}$$

where Σ_1^2 is the same as defined in Theorem 3.2. The second term can be dealt with in a similar way to what has been done before. This completes the proof of Theorem 3.2.

9.5 Proof of Theorem 4.1

Let $e_{nj}=\theta_j-E[\theta_j|T_{nj}]$. Observe that

$$\sqrt{\sum_{j=1}^n K\left(\frac{T_{nj}-T_n}{h}\right)}\,(g_{km}(T_n)-g(T_n))=\sqrt{\frac{\sum_{j=1}^m K^2\left(\frac{T_{nj}-T_n}{h}\right)}{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)}}\cdot\sum_{j=1}^m P_{mj}(T_n)e_{nj}$$

$$+\frac{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)(g(T_{nj})-g(T_n))}{\sqrt{\sum_{j=1}^n K\left(\frac{T_{nj}-T_n}{h}\right)}}$$

$$\equiv\sqrt{\frac{\sum_{j=1}^m K^2\left(\frac{T_{nj}-T_n}{h}\right)}{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)}}\cdot Q_{m1}(T_n)+\sqrt{\sum_{j=1}^n K\left(\frac{T_{nj}-T_n}{h}\right)}\cdot Q_{m2}(T_n), \tag{10.20}$$

where $P_{mj}(T_n)=\frac{K\left(\frac{T_{nj}-T_n}{h}\right)}{\sqrt{\sum_{j=1}^m K^2\left(\frac{T_{nj}-T_n}{h}\right)}}$, $Q_{m1}(T_n)=\sum_{j=1}^m P_{mj}(T_n)e_{nj}$ and

$$Q_{m2}(T_n)=\frac{\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)(g(T_{nj})-g(T_n))}{\sum_{j=1}^n K\left(\frac{T_{nj}-T_n}{h}\right)}. \tag{10.21}$$

Let $f_n(\cdot)$ be the density of T_{nj}. Simple calculation implies for $i=1,2$, $\sum_{j=1}^m E\left[K^i\left(\frac{T_{nj}-T_n}{h}\right)\right]=(1+o(1))\;mh\int K^i(v)dv\cdot f_n(T_n)$ when T_n is given. Such results, along with the law of large numbers, imply that $\frac{1}{mhf_n(T_n)}\sum_{j=1}^m K^i\left(\frac{T_{nj}-T_n}{h}\right)=(1+o_P(1))\int K^i(v)dv$ for $i=1,2$.

Thus, we have as $n\to\infty$

$$E\left[\frac{1}{mh}\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)(g(T_{nj})-g(T_n))\right]=(1+o(1))h^2C(T_n)\cdot\int u^2K(u)du, \tag{10.22}$$

where $C(T_n)=\frac{1}{2}p(T_n)g^{(2)}(T_n)+g^{(1)}(T_n)p^{(1)}(T_n)$.

Let $D(T_n)=p(T_n)\left(g^{(1)}(T_n)\right)^2$. Similarly, we have as $n\to\infty$

$$\mathrm{Var}\left[\frac{1}{\sqrt{mh}}\sum_{j=1}^m K\left(\frac{T_{nj}-T_n}{h}\right)(g(T_{nj})-g(T_n))\right]=\frac{1}{mh}\sum_{j=1}^m \mathrm{Var}\left(K\left(\frac{T_{nj}-T_n}{h}\right)(g(T_{nj})-g(T_n))\right)$$

$$\le\frac{1}{mh}\sum_{j=1}^m E\left[K^2\left(\frac{T_{nj}-T_n}{h}\right)(g(T_{nj})-g(T_n))^2\right]=(1+o(1))\;h^2\;D(T_n)\cdot\int u^2K^2(u)du,$$

which, along with equation (10.22), implies as $n \to \infty$

$$\frac{\sqrt{mh}}{\lambda_n} Q_{m2}(T_n) = \left(O_P\left(\frac{h}{\lambda_n}\right) + O_P\left(\frac{\sqrt{mh^5}}{\lambda_n}\right)\right) \cdot \gamma(T_n) = O_P(1) \tag{10.23}$$

by Assumption 4.2(ii), where $\gamma(T_n) = \frac{1}{2} g^{(2)}(T_n) + \frac{p^{(1)}(T_n)}{p(T_n)} g^{(1)}(T_n) + g^{(1)}(T_n)$.

Using the result that $E[e_1^2|T_{n1}] = \mathrm{Var}[\theta|T_n] = \lambda_n^2$ as well as Assumptions 4.1 and 4.2, the standard conditions required for establishing the central limit theorem are satisfied. The proof of Theorem 4.1 is therefore completed.

9.6 Proof of Theorem 4.2

Recall that $p(T_n) = \int f(T_n|\theta)\pi(\theta)d\theta$, $q(T_n) = \int \theta f(T_n|\theta)\pi(\theta)d\theta$ and $\widehat{p}(T_n) = \frac{1}{m}\sum_{i=1}^m f(T_n|\theta_i)$. By the definition of $g_{am}(T_n)$, we have

$$\sqrt{m}\, \widehat{p}(T_n) \left(g_{am}(T_n) - g(T_n)\right) = \frac{1}{\sqrt{m}} \sum_{i=1}^m \left(\theta_i f(T_n|\theta_i) - \frac{q(T_n)}{p(T_n)} f(T_n|\theta_i)\right)$$
$$= \sum_{i=1}^m \frac{1}{\sqrt{m}} \cdot \left(\theta_i f(T_n|\theta_i) - \frac{q(T_n)}{p(T_n)} f(T_n|\theta_i)\right) \equiv \sum_{i=1}^m Y_{mi}, \tag{10.24}$$

where $Y_{mi} = \frac{1}{\sqrt{m}} \cdot \left(\theta_i f(T_n|\theta_i) - \frac{q(T_n)}{p(T_n)} f(T_n|\theta_i)\right)$.

By Assumption 4.3, the standard conditions required for establishing the central limit theorem of a sum of i.i.d. random variables are satisfied. Therefore, we have for fixed n and as $m \to \infty$

$$\sqrt{\frac{m\, p^2(T_n)}{\sigma^2(T_n)}} \cdot \frac{\widehat{p}(T_n)}{p(T_n)} \left(g_{am}(T_n) - g(T_n)\right)$$
$$= \frac{1}{\sqrt{m\ \sigma^2(T_n)}} \sum_{i=1}^m \left(\theta_i f(T_n|\theta_i) - \frac{q(T_n)}{p(T_n)} f(T_n|\theta_i)\right) \to_D N(0,1), \tag{10.25}$$

which completes the proof of Theorem 4.2.

9.7 Proof of Theorem 5.1

9.7.1 Lemmas

In order to prove Theorem 5.1, we need to introduce the following lemmas.

Lemma 9.1: Suppose that Assumption 5.1 holds. Let $\sup_{\gamma\in\Gamma_n} |L_{mn}(\gamma) - L_n(\gamma)| = o_P\left(\nu_n^{-\frac{1}{2}}\right)$ for $m = m(n) \to \infty$. Then, we have $\sqrt{\nu_n}\ l_n^{\frac{1}{2}}(\gamma_0)\left(\gamma_{mn} - \gamma_n\right) = o_P(1)$ and $\sqrt{\nu_n}\ l_n^{\frac{1}{2}}(\gamma_0)\left(\gamma_{mn} - \gamma_0\right) \to_D \xi$, where ξ is as defined in Theorem 5.1.

The proof of Lemma 9.1 follows from that of Theorem A.5 of Kristensen and Shin (2012).

Lemma 9.2: Let Assumptions 5.2 and 5.3 hold. Then as $(m, n) \to (\infty, \infty)$

$$\sup_{x\in R^r} \sup_{\gamma\in\Gamma_n} |p_m(x;\gamma) - E\left[p_m(x;\gamma)\right]| = O_P\left(\sqrt{s_m}\right). \tag{10.26}$$

where $\Gamma \subset R^r$ is the parameter space of γ, and $s_m = \frac{\log(m)}{m\ h^r}$.

The proof of Lemma 9.2 follows immediately from Theorem 1(ii) of Kristensen (2009) because the setting here is all i.i.d.

9.7.2 Proof of Theorem 5.1

The main idea here is to apply Lemma 9.2 to verify $\sup_{\gamma\in\Gamma_n} |L_{mn}(\gamma) - L_n(\gamma)| = o_P\left(\nu_n^{-\frac{1}{2}}\right)$ for $m = m(n) \to \infty$.

Observe that

$$\begin{aligned}
L_{mn}(\gamma) - L_n(\gamma) &= \frac{1}{\nu_n}\sum_{i=1}^{n}((\log(p_m(X_i;\gamma)) - (\log(p(X_i;\gamma))) \\
&= \frac{1}{\nu_n}\sum_{i=1}^{n}\log\left(1 + \frac{p_m(X_i;\gamma) - p(X_i;\gamma)}{p(X_i;\gamma)}\right) \\
&= \frac{1}{\nu_n}\sum_{i=1}^{n}\frac{1}{p(X_i;\gamma)}\cdot(p_m(X_i;\gamma) - p(X_i;\gamma)) + \frac{1}{\nu_n}\sum_{i=1}^{n}\Delta_{mn}(X_i;\gamma) \\
&\equiv R_{mn1}(\gamma) + R_{mn2}(\gamma), \qquad (10.27)
\end{aligned}$$

where $\Delta_{mn}(x;\gamma)$ is a function of terms with orders higher than $\frac{1}{p(X_i;\gamma)}\cdot(p_m(X_i;\gamma) - p(X_i;\gamma))$, and such higher order terms are negligible in the evaluation of the order of $L_{mn}(\gamma) - L_n(\gamma)$.

Using Lemma 9.2, we have

$$\begin{aligned}
|R_{mn1}(\gamma)| &= \frac{1}{\nu_n}\left|\sum_{i=1}^{n}\frac{1}{p(X_i;\gamma)}\cdot(p_m(X_i;\gamma) - p(X_i;\gamma))\right| \qquad (10.28)\\
&\leq \frac{1}{\nu_n}\sum_{i=1}^{n}\frac{1}{p(X_i;\gamma)}\cdot|p_m(X_i;\gamma) - E_1\left[p(X_i;\gamma)\right]| + \frac{1}{\nu_n}\sum_{i=1}^{n}\frac{1}{p(X_i;\gamma)}\cdot|E_1\left[p_m(X_i;\gamma)\right] - p(X_i;\gamma)| \\
&\leq \frac{1}{\nu_n}\sum_{i=1}^{n}\sup_{x\in R^r}\sup_{\gamma\in\Gamma_n}\left(\frac{1}{p(x;\gamma)}\cdot|p_m(x;\gamma) - E\left[p(x;\gamma)\right]|\right) \\
&+ \frac{1}{\nu_n}\sum_{i=1}^{n}\sup_{x\in R^r}\sup_{\gamma\in\Gamma_n}\left(\frac{1}{p(x;\gamma)}\cdot|E\left[p_m(x;\gamma)\right] - p(x;\gamma)|\right) \leq \frac{C\ n}{\nu_n}\left(\sqrt{\frac{\log(m)}{mh^r}} + h^2\right) = o_P\left(\sqrt{\nu_n}\right)
\end{aligned}$$

using the standard result that $|E[p_m(x,\gamma)] - p(x;\gamma)| \leq C(x,\gamma)h^2$ as well as Assumptions 5.1(ii) and 5.3(iv) in particular, where $E_1[U]$ denotes the conditional expectation of U given X_i, and $C(x,\gamma)$ is a function involving $p_2(x;\gamma)$ that is the second–order partial derivative of $p(x;\gamma)$ with respect to x.

This shows that $\sup_{\gamma\in\Gamma_n} |L_{mn}(\gamma) - L_n(\gamma)| = o_P\left(\nu_n^{-\frac{1}{2}}\right)$ required in Lemma 9.1 is satisfied. Therefore, the proof of Theorem 5.1 then follows from Lemma 9.1.

Remark 9.1: As mentioned in Section 5, one may replace $L_{mn}(\gamma)$ by a truncated version of the form

$$L_{mnc}(\gamma) = \frac{1}{\nu_n}\sum_{i=1}^{n} w_{mn}(X_i)\log(p_m(X_i;\gamma)), \qquad (10.29)$$

where $w_{mn}(X_i) = 1$ if $p_m(X_i;\gamma) > c_{mn}$ and $w_{mn}(X_i) = 0$ if $p_m(X_i;\gamma) < \frac{c_{mn}}{2}$, where $c_{mn} > 0$ and $c_{mn} \to 0$ as $(m,n) \to (\infty,\infty)$.

Let $f_m(x;\gamma) = p^{-1}(x;\gamma)$. We then have

$$\begin{aligned} L_{mnc}(\gamma) - L_{mn}(\delta) &= -\frac{1}{\nu_n}\sum_{i=1}^{n} I\left[\frac{c_{mn}}{2} \le p_m(X_i;\gamma) \le c_{mn}\right] \log(p_m(X_i;\gamma)) \\ &= \frac{1}{\nu_n}\sum_{i=1}^{n} I\left[\log(c_{mn}) \le \log(f_m(X_i;\gamma)) \le \log(2c_{mn})\right] \log(f_m(X_i;\gamma)), \end{aligned} \tag{10.30}$$

which implies

$$\begin{aligned} &\sqrt{\nu_n}\; E\left[\sup_{\gamma\in\Gamma_n} |L_{mnc}(\gamma) - L_{mn}(\gamma)|\right] \\ &\le \frac{1}{\sqrt{\nu_n}}\sum_{i=1}^{n} E\left[\sup_{\gamma\in\Gamma_n} \left(|I\left[\log(c_{mn}) \le \log(f_m(X_i;\gamma)) \le \log(2c_{mn})\right]| \cdot |\log(f_m(X_i;\gamma))|\right)\right] \\ &\le |\log(c_{mn})|^{-\delta} \cdot \frac{n}{\sqrt{\nu_n}} E\left[\sup_{\gamma\in\Gamma_n} \left(|\log(f_m(X_1;\gamma))|^{1+\delta}\right)\right] = o(1) \end{aligned} \tag{10.31}$$

when $\frac{n}{\sqrt{\nu_n}\;|\log(c_{mn})|^{\delta}} = o(1)$ and $\int \sup_{\gamma\in\Gamma_n}\left(|\log(p(x;\gamma))|^{1+\delta}\, p(x;\gamma)\right) dx < \infty$ for some $\delta > 0$. Note that it is possible to choose c_{mn} such that $\frac{n}{\sqrt{\nu_n}\;|\log(c_{mn})|^{\delta}} = o(1)$ for a suitable $\delta > 0$. When $|\log(c_{mn})| = (mn)^c$ for some $c > 0$ and $\nu_n = n$, for example, the condition is satisfied if $\frac{\sqrt{n}}{(mn)^{c\delta}} \to 0$.

REFERENCES

Amemiya, T. (1985), *Advanced Econometrics* (Cambridge: Harvard University Press).

Andrews, D. W. K. and Fair, R. C. (1988), "Inference in Nonlinear Econometric Models with Structural Change", *Review of Economic Studies*, **LV**, 615–640.

Beaumont, M., Zhang, W. and Balding, D. (2002), "Approximate Bayesian Computation in Population Genetics", *Genetics*, **162**, 2025–2035.

Blum, M. G. B. (2010), "Approximate Bayesian Computation: a Nonparametric Perspective", *Journal of the American Statistical Association*, **105**, 1178–1187.

Blum, M. G. B. (2013), "A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation", *Statistical Science*, **28**, 189–208.

Brooks, S., Gelman, A., Jones, G. and Meng, X. (2011), *Handbook of Markov Chain Monte Carlo* (London:. Chapman & Hall/CRC).

Carlin, B. and Louis, T. (1996), *Bayes and Empirical Bayes Methods for Data Analysis* (London: Chapman and Hall/CRC).

Chen, X. (2007), "Large Sample Sieve Estimation of Semi–Nonparametric Models", *Handbook of Econometrics*, **6**, 5549–5632.

Chernozhkov, V. and Hong, H. (2003), "An MCMC Approach to Classical Estimation", *Journal of Econometrics*, **115**, 293–346.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm", *Journal of the American Statistical Association*, **49**, 327–335.

CREEL, M. AND KRISTENSEN, D. (2011), "Indirect Likelihood Inference", Working paper is available at the website: http://econpapers.repec.org/paper/aubautbar/931.13.htm.

EFRON, B. (1996), "Empirical Bayes Methods for Combining Likelihoods (with discussion)", *Journal of the American Statistical Association*, **91**, 538–565.

FAN, J. AND YAO, Q. (2003), *Nonlinear Time Series: Non– and Parametric Methods* (New York: Springer).

FEARNHEAD, P. AND PRANGLE, D. (2012), "Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic Approximate Bayesian Computation (with discussion)", *Journal of the Royal Statistical Society Series B*, **74**, 419–474.

GALLANT, R. AND TAUCHEN, G. (1996), "Which Moments to Match ?", *Econometric Theory*, **12**, 363–390.

GAO, J. (2007), *Nonlinear Time Series: Semi– and Non–Parametric Methods* (London: Chapman & Hall/CRC).

GAO, J. AND HONG H. (2014), "A Computational Implementation of GMM", Working paper available at http://ssrn.com/abstract=2503199.

GAO, J. AND PHILLIPS, P. C. B. (2013), "Semiparametric Estimation in Triangular System Equations with Nonstationrity", *Journal of Econometrics*, **176**, 5979.

GELFAND, A. AND SMITH, A. F. M. (1990), "Sampling–based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, **85**, 398–409.

GEWEKE, J. (1989), "Bayesian Inference in Econometric Models using Monte Carlo Integration", *Econometrica*, **57**, 1317–1339.

GEWEKE, J. (2005), *Contemporary Bayesian Econometrics and Statistics* (New jersey: John Wiley & Sons).

HANSEN, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, **50**, 1029–1054.

IMBENS, G. AND KALYANARAMAN, K. (2012), "Optimal Bandwidth Choice for the Regression Discontinuity Estimator", *Review of Economic Studies*, **79**, 933–959.

KRISTENSEN, D. (2009), "Uniform Convergence Rates of Kernel Estimators with Heterogeneous Dependent Data", *Econometric Theory*, **25**, 1433–1445.

KRISTENSEN, D. AND SHIN, Y. (2012), "Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood", *Journal of Econometrics*, **167**, 76–94.

LI, Q. AND RACINE, J. (2007), *Nonparametric Econometrics: Theory and Practice* (New Jersey: Princeton University Press).

LIU, J. (2001), *Monte Carlo Strategies in Scientific Computing* (New York: Springer).

MA, J. (2008), "A Closed–form Asymptotic Variance–Covariance Matrix for the Quasi–Maximum Likelihood Estimator of the GARCH(1,1) Model"
(working paper available at http://papers.ssrn.com/sol3/papers.cfm?abstract-id=889461).

PELIGRAD, M. (1987), "On Central Limit Theorem for ρ–Mixing Sequences", *Annals of Probability*, **15**, 1387–1394.

PHILLIPS, P. C. B. AND HANSEN, B. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, **57**, 99-125.

TIERNEY, L. (1994), "Markov Chains for Exploring Posterior Distributions", *Annals of Statistics*, **22**, 1701–1762.