ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

Bayesian Bandwidth Selection in Nonparametric Time-Varying Coefficient Models

Tingting Cheng, Jiti Gao and Xibin Zhang

February 2013

Working Paper 07/13

Bayesian Bandwidth Selection in Nonparametric Time–Varying Coefficient Models [1]

Tingting Cheng, Jiti Gao,[2] Xibin Zhang

Department of Econometrics and Business Statistics, Monash University, Australia

Abstract: Bandwidth plays an important role in determining the performance of local linear estimators. In this paper, we propose a Bayesian approach to bandwidth selection for local linear estimation of time–varying coefficient time series models, where the errors are assumed to follow the Gaussian kernel error density. A Markov chain Monte Carlo algorithm is presented to simultaneously estimate the bandwidths for local linear estimators in the regression function and the bandwidth for the Gaussian kernel error–density estimator. A Monte Carlo simulation study shows that: 1) our proposed Bayesian approach achieves better performance in estimating the bandwidths for local linear estimators than normal reference rule and cross–validation; 2) compared with the parametric assumption of either the Gaussian or the mixture of two Gaussians, Gaussian kernel error–density assumption is a data–driven choice and helps gain robustness in terms of different specification of the true error density. Moreover, we apply our proposed Bayesian sampling method to the estimation of bandwidth for the time–varying coefficient models that explain Okun's law and the relationship between consumption growth and income growth in the U.S. For each model, we also provide calibrated parametric forms of its time–varying coefficients.

Key words: Bayes factors, local linear estimator, marginal likelihood, random–walk Metropolis algorithm.

JEL Classification: C11, C14, C15

[1]The authors acknowledge constructive comments and suggestions from participants at several seminars and conferences, particularly to those who attended the seminars held at Monash University in 2012 and then Econometric Society Australasian Meetings in July 2012. The authors also acknowledge constructive discussions with Professor Maxwell King. Thanks also go to the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP1096374 and DP1095838.

[2]Address: Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Telephone: 61399031675. Fax: 61399032007. Email: jiti.gao@monash.edu.

1 Introduction

Time–varying coefficient time series models have attracted much attention of econometricians and statisticians during the last two decades, following the publication of Robinson (1989). Recent studies that are most relevant to our work include Gao and Hawthorne (2006) for a semiparametric modeling of a temperature trend function, Cai (2007) for a nonparametric trending time series model, Li, Chen and Gao (2011), and Chen, Gao and Li (2012) for semiparametric trending panel data regression and its applications in modeling economic, financial and climatological data. An example that illustrates the importance of such models in finance is the capital asset pricing model (CAPM). A traditional CAPM usually assumes a constant linear relationship between an asset's return and a market portfolio's return, and such a relationship is reflected by the beta coefficient. However, some recent studies show that the beta coefficients might vary over time (see for example, Jagannathan and Wang, 1996; Ghysels, 1998; Wang, 2003).

In this paper, our investigation is focused on a time–varying coefficient time series model given by

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \varepsilon_t, \quad t = 1, 2, \cdots, n, \tag{1}$$

where $\boldsymbol{x}_t' = (x_{t,1}, x_{t,2}, \cdots, x_{t,k})$, $\boldsymbol{\beta}_t' = (\beta_{t,1}, \beta_{t,2}, \cdots, \beta_{t,k})$ are unknown functions of t, and ε_t is assumed to be independent and identically distributed (iid) with an unknown distribution.

The elements of $\boldsymbol{\beta}_t$ are usually estimated by local linear estimators (see Cai, 2007, among others). The performance of a local linear estimator is mainly determined by its bandwidth. Robinson (1989) proposed choosing bandwidth through the cross–validation (CV) method, but even in some simple settings, cross–validation may perform poorly and exhibit a large magnitude of sample variation (see Fan, Heckman and Wand, 1995, among others). Cai and Tiwari (2000), Cai (2002) and Cai (2007) proposed a nonparametric version of Akaike information criterion (AIC) to select bandwidth for the local linear estimator. This method aims to derive an optimal bandwidth that minimizes a measure of penalized logarithmic mean squared errors, and thus, is similar to the CV. A rule–of–thumb method for choosing bandwidth is the normal reference rule (NRR) defined as

$$h = 1.06 \sigma n^{-1/5}, \tag{2}$$

where σ is the population standard deviation and is replaced by its sample measure in practice. It is a rough method for choosing bandwidths and is used in the absence of no other practical methods.

In terms of error density estimation in a regression model, its importance is highlighted by Efromovich (2005) for nonparametric regression models. In the aforementioned example of the CAPM model, people might have interest in the density of an asset's return, which is characterized by the error density and should be estimated nonparametrically. One important use of the resulting density is to approximate the value–at–risk (VaR) for holding the underlying asset. A wrongly specified error density may produce inaccurate estimate of VaR. In the current literature, residuals are commonly used as proxies of errors, and the kernel density estimator of residuals is often used as an estimator of the error density. Therefore, a two–stage procedure is required for choosing bandwidth for the kernel estimator. The first stage aims to estimate the regression function, while the second stage aims to choose bandwidth for the kernel error–density estimator. In both stages, cross–validation could be used for choosing bandwidths. There exist some investigations on error–density estimation for nonparametric regression models (see for example, Efromovich, 2005; Zhang, King and Shang, 2011). However, to our knowledge, there has been no investigation on error density estimation for time–varying coefficient time series models.

We approximate the unknown error density by a Gaussian kernel (GK) density, which is a location–mixture of n Gaussian densities given by

$$f(\varepsilon_t; h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\phi\left(\frac{\varepsilon_t - \varepsilon_i}{h}\right), \tag{3}$$

where $\phi(\cdot)$ is the density function of the standard Gaussian distribution, all the component Gaussian densities have a common variance (denoted as h^2) and individual mean values. The GK density was proposed by Zhang and King (2011) as the error density of a GARCH model. They developed a Bayesian sampling algorithm to estimate the bandwidth and other parameters for the resulting GK GARCH model. When modelling daily stock–index returns, they found that their GK GARCH model is favored against the GARCH model with Student t errors for five out of eight stock markets. Conventionally, when the error density of a regression model is unknown, one may approximate it by a mixture of several Gaussian densities with different mean and variance values. Such an approximation is at the cost of dramatically increasing the number of parameters. In contrast, the GK density is determined by only one additional parameter, which is the common standard deviation of the component densities. It is the smoothing parameter when the GK density is regarded as a kernel–form error density estimator based on random errors. Because of this kernel form, it is reasonable to expect that this GK density can well approximate the true unknown error density.

The contribution of this paper is to present a Bayesian sampling algorithm, through which we are able to simultaneously estimate the bandwidths for local linear estimators in the regression function and the bandwidth for the kernel error–density estimator. In the current literature, there exist some investigations on the advantages of Bayesian sampling approaches to bandwidth estimation against some competing methods.[3] Motivated by recent investigations on sampling algorithms for bandwidth estimation, we propose to develop a Bayesian sampling algorithm for bandwidth estimation in the time–varying linear models given by (1), where the coefficients are estimated by local linear estimators, and the error density is assumed to be the Gaussian kernel density.

We conduct Monte Carlo simulation studies to investigate the performance of our proposed sampling algorithm in estimating bandwidths in the local linear estimators of the time–varying coefficients and the bandwidth in the Gaussian kernel density.

The rest of this paper is organized as follows. Section 2 briefly describes the local linear estimator of time–varying coefficients in the time series regression model. In Section 3, we investigate the likelihood and posterior for this model with its error density assumed to be Gaussian and the GK density, respectively. Section 4 presents Monte Carlo simulation studies to evaluate the performance of our proposed method for bandwidth estimation under three assumptions of the error density. In Section 5, two empirical examples are presented to illustrate the application of our proposed Bayesian sampling algorithm. Section 6 concludes the paper.

[3] Zhang, King and Hyndman (2006) presented a Bayesian sampling approach to bandwidth estimation for kernel density estimation based on directly observed data. They showed that their sampling approach outperforms the NRR. Zhang, Brooks and King (2009) proposed a sampling algorithm for bandwidth estimation for the Nadaraya–Waston estimator in a nonparametric regression, where the error density was assumed to be Gaussian. They found that their sampling method is comparable to cross–validation, and both outperform the NRR.

2 Local linear estimator of time–varying coefficients

Following the investigation on time–varying coefficient models by Robinson (1989) and Cai (2007), we assume that $\beta_{t,i}$, for $i = 1, 2, \cdots, k$, in (1) is a function of time t in the form of

$$\beta_{t,i} = \beta_i(\tau_t), \quad \text{for } t = 1, 2, \cdots, n,$$

where $\tau_t = t/n$. Under the assumption that $\beta_i(\cdot)$ has a continuous second derivative, we approximate $\beta_i(\tau_t)$ by a linear function of $\tau \in [0,1]$ given as

$$\beta_i(\tau_t) \approx \beta_i(\tau) + \beta_i^{(1)}(\tau)(\tau_t - \tau),$$

where $\beta_i^{(1)}(\tau)$ is the first derivative of $\beta_i(\tau)$. Let $\boldsymbol{z}_t$ be a column vector whose elements are respectively, those of $\boldsymbol{x}_t$ and $\boldsymbol{x}_t(\tau_t - \tau)$ and $\boldsymbol{\theta}_\tau = \left(\beta_1(\tau), \cdots, \beta_k(\tau), \beta_1^{(1)}(\tau), \cdots, \beta_k^{(1)}(\tau)\right)'$. Then the time–varying time series model given by (1) can be rewritten as

$$y_t \approx \boldsymbol{z}_t' \boldsymbol{\theta}_\tau + \varepsilon_t. \tag{4}$$

The parameter vector $\boldsymbol{\theta}_\tau$ can be estimated by minimizing the locally weighted sum of squares:

$$\sum_{t=1}^{n} (y_t - \boldsymbol{z}_t' \boldsymbol{\theta}_\tau)^2 K_h(\tau_t - \tau), \tag{5}$$

where $K_h(u) = K(u/h)/h$, $K(\cdot)$ is a kernel function, and $h > 0$ is the bandwidth satisfying that $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

The local linear estimator of $\boldsymbol{\theta}_\tau$ is the minimizer of (5) and is given by

$$\widehat{\boldsymbol{\theta}}_\tau(h) = \begin{pmatrix} S_{n,0}(\tau) & S_{n,1}'(\tau) \\ S_{n,1}(\tau) & S_{n,2}(\tau) \end{pmatrix}^{-1} \begin{pmatrix} R_{n,0}(\tau) \\ R_{n,1}(\tau) \end{pmatrix},$$

where

$$S_{n,j}(\tau) = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' (\tau_t - \tau)^j K_h(\tau_t - \tau),$$
$$R_{n,j}(\tau) = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{x}_t (\tau_t - \tau)^j K_h(\tau_t - \tau) y_t,$$

for $j = 0, 1$ and 2. The corresponding leave–one–out local linear estimator of $\boldsymbol{\theta}_\tau$ is given by

$$\widehat{\boldsymbol{\theta}}_\tau^{(-\tau)}(h) = \begin{pmatrix} S_{n,0}^{(-\tau)}(\tau) & S_{n,1}'^{(-\tau)}(\tau) \\ S_{n,1}^{(-\tau)}(\tau) & S_{n,2}^{(-\tau)}(\tau) \end{pmatrix}^{-1} \begin{pmatrix} R_{n,0}^{(-\tau)}(\tau) \\ R_{n,1}^{(-\tau)}(\tau) \end{pmatrix},$$

where

$$S_{n,j}^{(-\tau)}(\tau) = \frac{1}{n-1}\sum_{t=1;\tau_t\neq\tau}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'(\tau_t-\tau)^j K_h(\tau_t-\tau),$$
$$R_{n,j}^{(-\tau)}(\tau) = \frac{1}{n-1}\sum_{t=1;\tau_t\neq\tau}^{n} \boldsymbol{x}_t(\tau_t-\tau)^j K_h(\tau_t-\tau)y_t,$$

for $j = 0, 1$ and 2. Therefore, the leave–one–out local linear estimator of $\boldsymbol{\beta}_\tau$ is given by

$$\widehat{\boldsymbol{\beta}}_t^{(-t)}(h) = \left(\mathbf{1}_k', \mathbf{0}_k'\right)\widehat{\boldsymbol{\theta}}_\tau^{(-\tau)}(h), \tag{6}$$

where $\mathbf{1}_k$ and $\mathbf{0}_k$ are both k–dimensional column vectors with elements being one and zero, respectively.

3 Bayesian estimation of bandwidths

In this section, we investigate the construction of likelihood and posterior for the time–varying coefficient time series model with its error density assumed to be the GK density. We first briefly illustrate the basic building block of the Bayesian approach.

Let $\boldsymbol{y}$ be a vector of observed values of the response in (1), and assume that its true error density is $f(\varepsilon;\eta)$ characterized by its parameter vector η. Plugging–in the local linear estimator of time–varying coefficients into (1), we derive an approximate density of y_t denoted as $f\left(y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h);\eta,h\right)$, where the bandwidth h is treated as an additional parameter. Therefore, an approximate likelihood function of $\boldsymbol{y}$ given (η,h) is given by

$$\ell(\boldsymbol{y}|\eta,h) = \prod_{t=1}^{n} f\left(y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h);\eta,h\right). \tag{7}$$

According to Bayes theorem, the posterior of (η,h) is (up to a normalizing constant)

$$\pi(\eta,h|\boldsymbol{y}) \propto \pi(\eta)\pi(h)\ell(\boldsymbol{y}|\eta,h), \tag{8}$$

where $\pi(\eta)$ and $\pi(h)$ are respectively, the prior densities of η and h.

3.1 Gaussian error density

When the error density of (1) is assumed to be Gaussian with mean zero and variance σ^2, the density of y_t is Gaussian with its mean approximated by $\boldsymbol{x}'\widehat{\boldsymbol{\beta}}_t^{(-t)}$, and thus, the likelihood $\boldsymbol{y}$ is approximately

$$\ell_G(\boldsymbol{y}|h,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}^2\right). \tag{9}$$

We assume that the priors of h^2 and σ^2 both the inverse Gamma (IG) density with hyperparameters (α_h,β_h) and $(\alpha_\sigma,\beta_\sigma)$, respectively. The posterior of (h^2,σ^2) is (up to a normalizing constant)

$$\begin{aligned}\pi\left(h^2,\sigma^2|\boldsymbol{y}\right) &\propto \pi(h^2)\pi(\sigma^2)\ell_G(\boldsymbol{y}|h^2,\sigma^2)\\ &\propto \pi\left(h^2\right)\left(\frac{1}{\sigma^2}\right)^{(2\alpha_\sigma+n)/2+1}\exp\left(-\frac{2\beta_\sigma+\sum_{t=1}^{n}\left\{y_t-\boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}^2}{2\sigma^2}\right).\end{aligned} \tag{10}$$

Note that $\pi(h^2,\sigma^2|\boldsymbol{y}) = \pi(\sigma^2|h^2,\boldsymbol{y})\pi(h^2|\boldsymbol{y})$, where $\pi(\sigma^2|h^2,\boldsymbol{y})$ is the conditional posterior for given h^2 and $\boldsymbol{y}$ and is an inverse Gamma density with parameters given

$$\sigma^2|h^2,\boldsymbol{y} \sim \mathrm{IG}\left(\frac{2\alpha_\sigma+n}{2}, \frac{2\beta_\sigma+\sum_{t=1}^{n}\left\{y_t-\boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}^2}{2}\right). \tag{11}$$

Integrating σ^2 out of the joint posterior given by (10), we obtain the conditional posterior of h^2 as

$$\pi(h^2|\boldsymbol{y}) \propto \pi\left(h^2\right)\left(\frac{2\beta_\sigma+\sum_{t=1}^{n}\left\{y_t-\boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}^2}{2}\right)^{-(2\alpha_\sigma+n)/2}. \tag{12}$$

The sampling algorithm consists of sampling h^2 from (12) using the random–walk Metropolis algorithm and σ^2 from (11) using the Gibbs sampler.

3.2 Gaussian kernel density of errors

Any parametric assumption about the error density of (1) is only an approximation to the unknown true error density. Traditionally, when the error density is unknown, one could approximate it by a mixture of several Gaussians. However, such a mixture density is at the cost of dramatically increasing the number of parameters. Instead, in this paper, we assume that the error density is a weighted average of a density given by

$$f(\varepsilon;b) = \sum_{i=1}^{n} w_i g(\varepsilon;\varepsilon_i,b), \tag{13}$$

where we choose the weight w_i to be $1/n$, and $g(\varepsilon;\varepsilon_i,b)$ the Gaussian density with mean ε_i and variance b^2, which is denoted as $\phi((\varepsilon-\varepsilon_i)/b)/b$, for $i=1,2,\cdots,n$. Note that $\phi(\cdot)$ is the probability density function (pdf) of the standard Gaussian distribution. Therefore, the density function given by (13) can be rewritten as

$$f(\varepsilon; b) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b}\phi\left(\frac{\varepsilon-\varepsilon_i}{b}\right), \tag{14}$$

which we call the Gaussian kernel (GK) density because it has a kernel–form expression. It is reasonable to expect that the GK density approaches the unknown true error density as the sample size increases. Moreover, it is determined by one parameter only in contrast to the large number of parameters in traditional mixture density of several Gaussians.

If $\boldsymbol{\beta}_t$ were known, the GK error density would be a well–defined density function of the errors, and we would have

$$y_t \sim f\left(\left\{y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_t\right\}; b\right) = \frac{1}{n}\sum_{s=1}^{n}\frac{1}{b}\phi\left(\frac{\left\{y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_t\right\} - \left\{y_s - \boldsymbol{x}_s'\boldsymbol{\beta}_s\right\}}{b}\right), \tag{15}$$

for $t = 1, 2, \cdots, n$.

When $\boldsymbol{\beta}_t$ is unknown, we use the leave–one–out local linear estimator of $\boldsymbol{\beta}_t$ to replace the $\boldsymbol{\beta}_t$ in (15), Therefore, the density of y_t is approximated by $f\left(\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}; b\right)$, which is explicitly expressed as

$$f\left(\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}; b\right) = \frac{1}{n}\sum_{s=1}^{n}\frac{1}{b}\phi\left(\frac{\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\} - \left\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}_s^{(-s)}(h)\right\}}{b}\right), \tag{16}$$

for $t = 1, 2, \cdots, n$.

3.2.1 An approximate likelihood

For the purpose of constructing an appropriate likelihood function, the density function given by (16) contains an unwanted term $\phi(0)/b$, which could be arbitrarily large by choosing an arbitrarily small value of b. A remedy to this problem is to exclude the tth term from this summation, and thus, the density of y_t is approximated as

$$f_{(-1)}\left(\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}; b\right) = \frac{1}{n-1}\sum_{s=1; s\neq t}^{n}\frac{1}{b}\phi\left(\frac{\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\} - \left\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}_s^{(-s)}(h)\right\}}{b}\right).$$

For given parameters h^2 and b^2, the likelihood of $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$ is approximated by

$$\begin{aligned}\ell_{GK}\left(\boldsymbol{y}|h^2, b^2\right) &\approx \prod_{i=1}^{n} f_{(-1)}\left(\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\}; b\right) \\ &= \prod_{i=1}^{n}\left\{\frac{1}{n-1}\sum_{s=1;s\neq t}^{n}\frac{1}{b}\phi\left(\frac{\left\{y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_t^{(-t)}(h)\right\} - \left\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}_s^{(-s)}(h)\right\}}{b}\right)\right\}. \end{aligned} \tag{17}$$

3.2.2 Priors

We assume that the prior of h^2 is the inverse Gamma density denoted as IG(α_h,ζ_h) with its density function given by (up to a normalizing constant)

$$\pi(h^2) = \frac{\zeta_h^{\alpha_h}}{\Gamma(\alpha_h)}\left\{\frac{1}{h^2}\right\}^{\alpha_h+1}\exp\left\{-\frac{\zeta_h}{h^2}\right\}, \tag{18}$$

where α_h and ζ_h are hyperparameters. The prior of b^2 is also the inverse Gamma density with hyperparameters α_b and ζ_b.

3.2.3 Posterior

According to Bayes theorem, the posterior of h^2 and b^2 is approximately given by

$$\pi\left(h^2, b^2|\boldsymbol{y}\right) \propto \pi(h^2)\pi(b^2)\ell_{GK}\left(\boldsymbol{y}|h^2, b^2\right),$$

from which we use the random–walk Metropolis algorithm to sample h^2 and b^2. We record h and b at each iteration after the burn–in period, and the ergodic mean of each recorded chain is an estimate of the corresponding parameter.

4 Monte Carlo simulation

The purposes of the Monte Carlo simulation study are as follows. First, with two simulated samples of two different sample sizes, we illustrate the use and effectiveness of our Bayesian sampling algorithm for estimating bandwidths in the local linear estimator and the kernel estimator of the error density. Second, we compare the proposed Bayesian method with NRR and CV through 1,000 samples generated from a time–varying coefficient time series model with the error densities being respectively, the Gaussian, a mixture of two Gaussians and a centralized exponential distribution.

The performance of the estimated bandwidth for the local linear estimator is assessed by mean square errors (MSE) of the estimated time–varying parameters. Third, as the true error density is known in the simulation study, we evaluate the accuracy of the estimated error density through the mean integrated squared errors (MISE). Fourth, we further evaluate the performance of the sampling approach under each of the three error–density assumptions, where Bayes factors are used for comparison purposes. As such, we briefly illustrate the Bayes factors as follows.

4.1 Bayes factors

In Bayesian inference, model selection can be conducted through Bayes factors of the model of interest against its competing models. The Bayes factor represents a summary of evidence provided by the data supporting the model as opposed to its competing model. It is defined as the ratio of the marginal likelihoods derived under the model of interest and its competing model, respectively.

In this paper, we employ the method proposed by Chib (1995) to approximate the marginal likelihood. Let θ denote the parameter vector and $\boldsymbol{y}$ the data. Chib (1995) showed that the marginal likelihood under model $\mathscr{A}$ is expressed as

$$P_{\mathscr{A}}(\boldsymbol{y}) = \frac{L_{\mathscr{A}}(\boldsymbol{y} \mid \theta)\pi_{\mathscr{A}}(\theta)}{\pi_{\mathscr{A}}(\theta \mid \boldsymbol{y})},$$

which is computed at the posterior estimate of θ. The numerator is the product of likelihood and the prior of θ, while the denominator is the posterior density of θ and is approximated by the kernel density estimator based on the simulation output derived through a posterior simulator. The Bayes factor of model $\mathscr{A}$ against model $\mathscr{B}$ is defined as

$$BF = \frac{P_{\mathscr{A}}(\boldsymbol{y})}{P_{\mathscr{B}}(\boldsymbol{y})},$$

with which we can make a decision on whether $\mathscr{A}$ is favored against $\mathscr{B}$ according to Jeffreys (1961) scales modified by Kass and Raftery (1995). A Bayes factor between 1 and 3 indicates that the evidence supporting $\mathscr{A}$ against $\mathscr{B}$ is not worth more than a bare mention. When the Bayes factor is between 3 and 20, $\mathscr{A}$ is favored against $\mathscr{B}$ with positive evidence; when the Bayes factor is between 20 and 150, $\mathscr{A}$ is favored against $\mathscr{B}$ with strong evidence; and when the Bayes factor is above 150, $\mathscr{A}$ is favored against $\mathscr{B}$ with very strong evidence.

4.2 Performance of the proposed Bayesian methods

We consider the following data generating model:

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \varepsilon_t, \quad \text{for } t = 1, 2, \cdots, n, \tag{19}$$

where $\boldsymbol{x}_t = (1, x_{t,2})'$, $\boldsymbol{\beta}_t = (\beta_{1,t}, \beta_{2,t})'$, $\beta_{1,t} = 0.2\exp(-0.7 + 3.5t/n)$ and $\beta_{2,t} = 2t/n + \exp(-16(t/n - 0.5)^2) - 1$. In order to generate samples, we generate $x_{t,2}$ through an $AR(1)$ model given by $x_{t,2} = 0.8x_{t-1,2} + u_t$ with u_t being generated from $N(0, 0.5^2)$; we generate ε_t from the mixture of two Gaussians defined as $0.3\text{N}(0, 0.5^2) + 0.7\text{N}(0, 2^2)$; and then we calculate y_t according to (19). The sample sizes considered here are $n = 200$ and 400.

We assume that the relationship between y_t and $\boldsymbol{x}_t$ is approximated by a time–varying coefficient time series model given by (19), where the error density is assumed to be Gaussian, a mixture of two Gaussians and the GK error density given by (14). We applied our proposed Bayesian sampling method to (19) for each simulated sample. The burn–in period contains 1,000 draws and the following 10,000 draws were recorded. The posterior mean values of parameters are presented in Table 1 for sample size $n = 200$ and Table 2 for sample size $n = 400$. The mixing performance of this posterior simulator is examined by the SIF, which can be approximately interpreted as the number of draws needed so as to obtain an independent draw from the simulated Markov chain. We also derived the marginal likelihood under each assumption of the error density.

Table 1: *Parameters estimated through Bayesian sampling under assumptions of Gaussian, mixture of 2 Gaussians and GK error densities for* $n = 200$.

Error density	Parameter	Estimate	SIF
Gaussian	h	0.5553	19.4
	σ	1.7711	1.1
	log marginal likelihood	−404.5998	
Mixture of 2 Gaussians	w	0.2255	12.7
	μ_1	−0.2611	6.4
	σ_1	0.3718	10.7
	σ_2	1.9949	11.3
	h	0.3755	19.4
	log marginal likelihood	−402.6349	
GK density	h	0.4079	7.5
	b	0.4099	5.3
	log marginal likelihood	−398.1490	

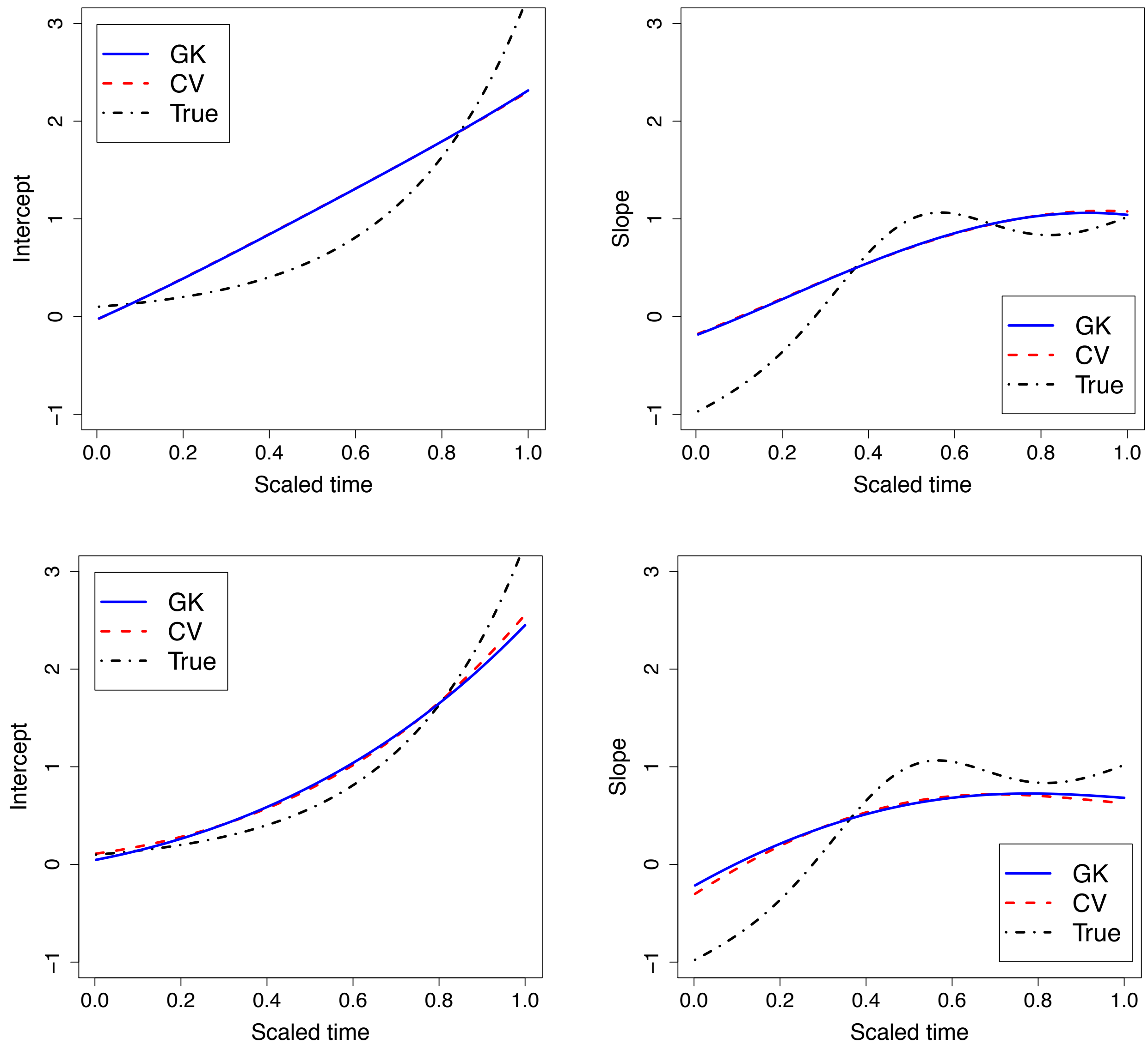
Table 2: *Parameters estimated through Bayesian sampling under assumptions of Gaussian, mixture of 2 Gaussians and GK error densities for* $n = 400$.

Error density	Parameter	Estimate	SIF
Gaussian	h	0.3111	5.8
	σ	1.7857	1.0
	log marginal likelihood	−805.7278	
Mixture of 2 Gaussians	w	0.2384	13.1
	μ_1	0.0393	5.9
	σ_1	0.4422	10.7
	σ_2	2.0310	8.5
	h	0.2989	6.5
	log marginal likelihood	−797.5236	
GK density	h	0.338206	10.5
	b	0.488289	7.4
	log marginal likelihood	−794.0806	

From Tables 1 and 2, we found that for each sample size, the marginal likelihood obtained under the GK error–density assumption is larger than that obtained under the assumption of Gaussian or the mixture of two Gaussians. When $n = 200$, the Bayes factor of the model under the GK error–density assumption is $\exp\{6.4508\}$ against the same model under the Gaussian error–density assumption and $\exp\{4.4859\}$ against the same model under the mixture of two Gaussians. Therefore, the assumption of GK error density is favored against Gaussian assumption with very strong evidence and against the mixture of two Gaussians with strong evidence. When $n = 400$, the Bayes factor of the GK error–density assumption is $\exp\{11.6472\}$ against Gaussian and $\exp\{3.4430\}$ against the mixture of two Gaussians. Thus, the GK error–density assumption is favored against Gaussian with very strong evidence and against the mixture of two Gaussians with strong evidence. Nonetheless, we cannot draw such a conclusion only based on two generated samples. In Section 4.5, we calculate the marginal likelihood under each assumption of the error density for 1,000 generated samples for both sample sizes.

We plotted the graphs of the estimated values of $\beta_1(\cdot)$ and $\beta_2(\cdot)$ in Figure 1. We found that the estimated time–varying parameters through Bayesian and CV are similar to each other. It can also be shown that Bayesian sampling performs well when the sample size is large.

Figure 1: *Estimated curves of $\beta_1(t)$ and $\beta_2(t)$ for the sample sizes of 200 (top panel) and 400 (bottom panel).*



4.3 Accuracy of the estimated bandwidth

One important issue in kernel estimation is the bandwidth choice. In this section, we compare our Bayesian sampling method with two existing bandwidth selection methods, which are the NRR and CV. The accuracy of the estimated bandwidths is measured by the mean squared errors (MSE) defined as

$$\mathrm{MSE}(h) = \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\beta}_{1,t}(\widehat{h}) - \beta_{1,t}\right)^2 + \frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\beta}_{2,t}(\widehat{h}) - \beta_{2,t}\right)^2.$$

For each sample size, we generated 1,000 samples according to the regression function given by (19) with the error density being respectively, $\mathrm{N}(0, 0.25^2)$, $0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$ and $\exp(1) - 1$. For each sample, we derived the estimate of bandwidth in the local linear estimator through NRR, CV and Bayesian sampling under each of the three error–density assumptions, which are the Gaussian, mixture of two Gaussians and GK error density. We calculated the MSE under each situation. Further, we obtained the relative efficiency improved by Bayesian sampling method compared to NRR and CV. These results are presented in Tables 3 and 4.

Table 3: *Mean and standard deviation (given in parentheses) of 1,000 MSE values derived through each bandwidth estimation method based on 1,000 generated samples.*

Source of errors	NRR	CV	Bayesian		
			Gaussian	Mixture	GK
$n = 200$					
$\mathrm{N}(0, 0.25^2)$	1.0485	1.0346	1.0287	1.0285	1.0252
	(0.0585)	(0.0580)	(0.0540)	(0.0538)	(0.0535)
$0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$	1.6059	1.1169	1.0737	1.0681	1.0703
	(0.5132)	(0.3996)	(0.3400)	(0.3274)	(0.3376)
$\exp(1) - 1$	1.2240	1.0430	1.0344	1.0185	1.0071
	(0.2423)	(0.2062)	(0.1936)	(0.1750)	(0.1776)
$n = 400$					
$\mathrm{N}(0, 0.25^2)$	1.0454	1.0386	1.0338	1.0337	1.0311
	(0.0367)	(0.0363)	(0.0356)	(0.0355)	(0.0355)
$0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$	1.2827	1.0648	1.0530	1.0458	1.0369
	(0.2770)	(0.2501)	(0.2328)	(0.2186)	(0.2209)
$\exp(1) - 1$	1.1171	1.0367	1.0293	1.0193	1.0136
	(0.1545)	(0.1405)	(0.1327)	(0.1245)	(0.1227)

Table 4: *The relative improvements achieved by Bayesian sampling in comparison with the NRR and CV.*

Source of errors	Bayesian vs NRR			Bayesian vs CV		
	Gaussian	Mixture	GK	Gaussian	Mixture	GK
$n = 200$						
$\mathrm{N}(0, 0.25^2)$	1.9%	1.9%	2.2%	0.5%	0.6%	0.9%
$0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$	31.6%	31.8%	31.8%	2.2%	2.2%	2.0%
$\exp(1) - 1$	14.7%	15.8%	16.7%	0.5%	1.5%	2.6%
$n = 400$						
$\mathrm{N}(0, 0.25^2)$	1.1%	1.1%	1.4%	0.5%	0.5%	0.7%
$0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$	17.4%	17.8%	18.5%	0.7%	1.0%	1.9%
$\exp(1) - 1$	7.6%	8.3%	8.8%	0.6%	1.4%	1.9%

When the errors were generated from $\mathrm{N}(0, 0.25^2)$, we found that for any error density assumption, the MSE derived through Bayesian sampling is smaller than that derived through either the NRR or

CV, and the relative improvements achieved by Bayesian sampling compared to NRR and CV are at least 1.9% and 0.5%, respectively, for $n = 200$; and 1.1% and 0.5%, respectively, for $n = 400$.

When the errors were generated from $0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$, the MSE derived through Bayesian sampling under any error–density assumption is smaller than that derived through either the NRR or CV. The relative improvements achieved by Bayesian sampling in comparison with the NRR and CV are at least 31.6% and 2%, respectively, for $n = 200$; and 17.4% and 0.7%, respectively, for $n = 400$.

When the errors were generated from a centralized exponential distribution with its density defined as $\exp(1) - 1$, the MSE derived through Bayesian sampling under any error–density assumption is less than that derived through either the NRR or CV. The relative improvements achieved by Bayesian sampling compared with the NRR and CV are at least 14.7% and 0.5%, respectively, for $n = 200$; and 7.6% and 0.6%, respectively, for $n = 400$.

Therefore, our Bayesian sampling method can provide more accurate results than the NRR and CV under the MSE measure.

4.4 Performance of the estimated error density

The error density of a regression model is often of great interest for inference. To measure the performance of the estimated error density under each bandwidth estimation method, we calculated the mean integrated squared errors (MISE), which is defined as

$$\mathrm{MISE} = E\int_{-\infty}^{+\infty} \left(f(x) - \widehat{f}(x)\right)^2 dx,$$

where f represents the unknown error density and $\widehat{f}$ its kernel estimator. In the Monte Carlo simulation study, for each simulated sample, the ISE is approximately by

$$\mathrm{ISE} \approx \frac{1}{m}\sum_{i=1}^{m} \left(f(x_i) - \widehat{f}(x_i)\right)^2,$$

where $x_1, x_2, \cdots, x_m$ are grid points over the majority support of $f(x)$. The MISE was approximated by the mean of the approximate ISE values of all simulated samples. The MISE results for sample sizes 200 and 400 are presented in Table 5, from which our findings are as follows.

First, when the errors were generated from $\mathrm{N}(0, 0.25^2)$, the MISE derived through Bayesian sampling under the normality assumption of errors is smaller than that derived through CV or

Table 5: *Mean and standard deviation (given in parentheses) of ISE values of the estimated error density derived through each bandwidth estimation method based on 1,000 generated samples with errors simulated from three specified densities.*

Source of errors	CV	Bayesian		
		Gaussian	Mixture	GK
$n = 200$				
$\mathrm{N}(0, 0.25^2)$	0.0133	0.0033	0.0052	0.0125
	(0.0118)	(0.0041)	(0.0103)	(0.0082)
$0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$	0.0096	0.0201	0.0070	0.0085
	(0.0044)	(0.0031)	(0.0055)	(0.0043)
$\exp(1) - 1$	0.0810	0.1744	0.0827	0.0796
	(0.0196)	(0.0074)	(0.0135)	(0.0211)
$n = 400$				
$\mathrm{N}(0, 0.25^2)$	0.0080	0.0015	0.0032	0.0091
	(0.0064)	(0.0022)	(0.0056)	(0.0057)
$0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$	0.0062	0.0197	0.0028	0.0054
	(0.0032)	(0.0022)	(0.0026)	(0.0029)
$\exp(1) - 1$	0.0662	0.1741	0.0782	0.0653
	(0.0179)	(0.0057)	(0.0102)	(0.0186)

Bayesian sampling under the assumption of a mixture of two Gaussians and GK error density assumptions. Second, when the errors were generated from $0.3\mathrm{N}(0, 0.5^2) + 0.7\mathrm{N}(0, 2^2)$, the MISE derived through Bayesian sampling under the assumption of a mixture of two Gaussians is smaller than that derived through CV or Bayesian sampling under the Gaussian and GK error–density assumptions. Third, when the errors were generated from $\exp(1) - 1$, the MISE derived through Bayesian sampling under the GK error–density assumption is smaller than that derived through CV or Bayesian sampling under the Gaussian and a mixture of two Gaussians error–density assumptions. Finally, when the true error density is neither Gaussian nor mixture of two Gaussians, the GK error–density assumption leads to a better performance than the assumption of either Gaussian or a mixture of two Gaussians. This finding indicates that the GK error–density assumption is more robust in terms of different specifications of the true error density.

4.5 Bayesian comparison among error density assumptions

The benefit of GK error–density assumption is to gain robustness in terms of error density specification, because it is a kernel density estimator of residuals, and has the capacity to approximate an unknown error density. In order to show this, we conducted a simulation study using the same 1,000 samples for each of the sample size $n = 200$ and 400 in Section 4.3. Simulation results are described as follows.

4.5.1 Gaussian distribution of errors

We generated samples by simulating the errors of (19) from the Gaussian distribution with mean zero and variance 0.25^2. The proposed Bayesian sampling was then implemented to estimate bandwidth parameters, where the error density was assumed to be the Gaussian with mean zero and variance σ^2, a mixture of two Gaussians and GK error density, respectively.

As the assumption of Gaussian errors is correct, we calculated Bayes factors of the model under the Gaussian assumption of errors against the same model under the assumptions of a mixture of two Gaussians and GK error density, respectively. For each simulated sample, we made a decision on whether the Gaussian assumption is favored against the other two based on the Jeffreys (1961) scales modified by Kass and Raftery (1995). Of all 1000 simulated samples, we reported the relative frequency that the Gaussian assumption is favored against each of the other two for sample sizes of 200 and 400, respectively. The results of such relative frequencies are presented in the second column of Table 6.

For sample size $n = 200$, the Gaussian assumption of the error density is favored against a mixture of two Gaussians in 77.2% of the simulated sample with positive evidence, 14.4% with strong evidence, and 4.6% with very strong evidence. Thus, the former is favored against the latter in 96.2% of the simulated samples with either positive, strong or very strong evidence. The Gaussian assumption is favored against the assumption of GK error density in 28.6% of the simulated sample with positive evidence, 61.9% with strong evidence, and 4.9% with very strong evidence. All together, the former is favored against the latter in 95.4% of simulated samples with either positive, strong or very strong evidence. Nonetheless, there are 1% simulated samples, through which we could find positive evidence supporting the assumption of GK error density against Gaussian.

When the sample size increases to 400, the revealed evidence supporting Gaussian against each of the two competing assumptions becomes stronger than evidence revealed when the sample size of 200. Moreover, no positive evidence could be found to support either the GK or mixture

Table 6: *Bayesian decision on the choice of error–density assumption.*

Source of errors	Gaussian		Mixture of two Gaussians		$\exp(1)-1$	
	Mixture	GK	Gaussian	GK	Gaussian	Mixture
$n = 200$						
$BF \le 1/150$	0.0%	0.0%	26.6%	78.8%	0.0%	12.6%
$1/150 < BF \le 1/20$	0.0%	0.0%	9.6%	12.0%	0.0%	3.8%
$1/20 < BF \le 1/3$	0.2%	1.0%	13.6%	6.0%	0.0%	3.6%
$1/3 < BF \le 1$	0.4%	0.6%	11.2%	0.8%	0.0%	2.2%
$1 < BF \le 3$	3.2%	3.0%	8.0%	1.6%	0.0%	3.2%
$3 < BF \le 20$	77.2%	28.6%	11.0%	0.8%	0.0%	4.8%
$20 < BF \le 150$	14.4%	61.9%	9.2%	0.0%	0.0%	5.6%
$BF > 150$	4.6%	4.9%	10.8%	0.0%	100%	64.2%
$n = 400$						
$BF \le 1/150$	0.0%	0.0%	1.4%	64.0%	0.0%	14.4%
$1/150 < BF \le 1/20$	0.0%	0.0%	0.6%	18.4%	0.0%	1.0%
$1/20 < BF \le 1/3$	0.0%	0.0%	1.6%	7.6%	0.0%	1.0%
$1/3 < BF \le 1$	0.0%	0.4%	3.4%	3.2%	0.0%	1.0%
$1 < BF \le 3$	3.2%	1.8%	3.2%	1.8%	0.0%	1.6%
$3 < BF \le 20$	55.4%	8.6%	10.6%	2.6%	0.0%	1.4%
$20 < BF \le 150$	30.4%	41.6%	11.4%	0.6%	0.0%	2.8%
$BF > 150$	11.0%	47.6%	67.8%	1.8%	100%	76.8%

assumption of the errors against the Gaussian assumption. This is not a surprising result, because the Gaussian assumption is the correct one.

4.5.2 A mixture density of two Gaussians of errors

When errors were generated from the mixture of two Gaussians defined as $0.3\mathrm{N}(0,0.5^2)+0.7\mathrm{N}(0,2^2)$, the assumption of a mixture of two Gaussians was supposed to be correct, but it has one location parameter, one weight parameter and two scale parameters. We calculated the Bayes factors of the model under this mixture assumption against the same model under the Gaussian and GK assumptions, respectively.

When $n = 200$, the assumption of the mixture of two Gaussians is favored against the assumption of GK error density in only 0.8% simulated samples with merely positive evidence. In contrast, the latter is favored against the former in 78.8% simulated samples with very strong evidence, in 12% samples with strong evidence, and in 6% samples with positive evidence. Nonetheless, when sample size is increased to 400, we could find strong evidence in 0.6% samples and very strong evidence in 1.8% samples supporting the mixture assumption against the GK assumption. Meanwhile, the revealed evidence supporting the GK assumption against the mixture assumption becomes slightly weaker than evidence revealed under sample size of 200.

In terms of the comparison between the assumptions of mixture and Gaussian error densities for $n = 200$, we found that the mixture assumption is favored against the Gaussian assumption with very strong evidence in 10.8% samples, strong evidence in 9.2% samples and positive evidence 11% samples. Therefore, the former assumption is favored against the latter in 31% samples with at least positive evidence. Meanwhile, the Gaussian assumption is favored against the mixture assumption in 26.6% samples with very strong evidence, 9.6% samples with strong evidence and 13.6% samples with positive evidence. This ends up with the fact that the Gaussian assumption is favored against the mixture in 49.8% samples with at least positive evidence.

When sample size is increased to 400, the comparison between the two assumptions reveals different results. The mixture assumption is favored against the Gaussian with very strong evidence in 67.8% samples, strong evidence in 11.4% samples and positive evidence 10.6% samples. As such, the former is favored against the latter in 89.8% samples with at least positive evidence. However, the latter assumption is favored against the former in 1.4% samples with very strong evidence, 0.6% samples with strong evidence and 1.6% samples with positive evidence. These numbers indicate that the Gaussian assumption is favored against the mixture in only 3.6% samples with at least positive evidence.

The finding indicates that for small samples, the mixture assumption cannot compete with Gaussian even though the errors follow a mixture density of two Gaussians. However, for large samples, the correct assumption does have competing ability against the Gaussian assumption.

4.5.3 A centralized exponential distribution of errors

We also considered generating samples by simulating errors of (19) from a centralized exponential distribution with its density defined as $\exp(x-1)$ (for $x > 0$). In this situation, all three assumptions are incorrect. We computed the Bayes factors of the GK assumption against the other two assumptions, respectively. For each generated sample, we made a decision on whether the model under the assumption of GK error density is favored against the same model under each competing assumption. Of all simulated samples, we reported the relative frequencies that the GK assumption is favored against its competing assumptions in the last column of Table 6.

The results show that for sample sizes of 200 and 400, the assumption of GK error density is favored with very strong evidence against the assumption of Gaussian error density in all the samples. For sample size of 200, the GK assumption is favored against the mixture assumption with very strong evidence in 64.2% samples, strong evidence in 5.6% samples and positive evidence in 4.8% samples. All together, the GK assumption is favored against the mixture 74.6% samples with at least positive evidence. Meanwhile, the latter assumption is favored against the former with very strong evidence in 12.6% samples, strong evidence in 3.8% samples and positive evidence in 3.6% samples. To sum up, the mixture assumption is favored against the GK assumption in 20% samples.

For sample size of 400, the evidence supporting one assumption against the other becomes stronger than that for sample size of 200. The GK assumption is favored against the mixture with very strong evidence in 76.8% samples, while the mixture is favored against the GK with very strong evidence in 14.4% samples. Both values are larger than those derived under sample size of 200.

To conclude this subsection on comparison among the three assumptions about the error density, the GK assumption does not lead to a better performance than the correct assumption when the errors are simulated from the Gaussian. However, the GK assumption does lead to a better performance than the mixture assumption when the errors are simulated from a mixture density of two Gaussians. Moreover, the GK assumption leads clearly better performance than the two competing assumptions when the errors are simulated from a centralized exponential distribution. Thus, the simulation study reveals that the assumption of GK error density in the underlying time–varying coefficient time series model helps achieve robustness in terms of different assumptions of the error density.

5 Two empirical applications

5.1 Okun's law

Since the seminal work of Okun (1962), there have been many studies that provide evidence of the correlation between variations in unemployment rate and real output over a business cycle.

However, many studies in the current literature on Okun's law tend to focus on the lack of robustness of Okun's coefficient without questioning the constant linear relationship (see for example, Lee, 2000; Silvapulle, Moosa and Silvapulle, 2004). We employ the time–varying coefficient model to demonstrate the time–varying feature of Okun's coefficient.

The sample contains yearly paired observations of unemployment rate and GDP of the U.S. over the period from 1949 to 2010 with sample size $n = 62$. The data were collected from the OECD database. Okun's law is usually investigated through the model given by

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + \varepsilon_t,$$

where Δy_t is the yearly change of the unemployment rate, Δx_t is the yearly change of the log GDP, and $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are independent and identically distributed random errors. The time series plots of y_t, x_t, Δy_t and Δx_t are presented in Figure 2.

The ordinary least squares (OLS) estimates of α_0 and α_1 are respectively, 1.3514(0.0000) and −0.9185(0.0000) with corresponding p values given in parentheses. We fitted the time–varying coefficient model,

$$\Delta y_t = \beta_0(t) + \beta_1(t)\Delta x_t + \varepsilon_t$$

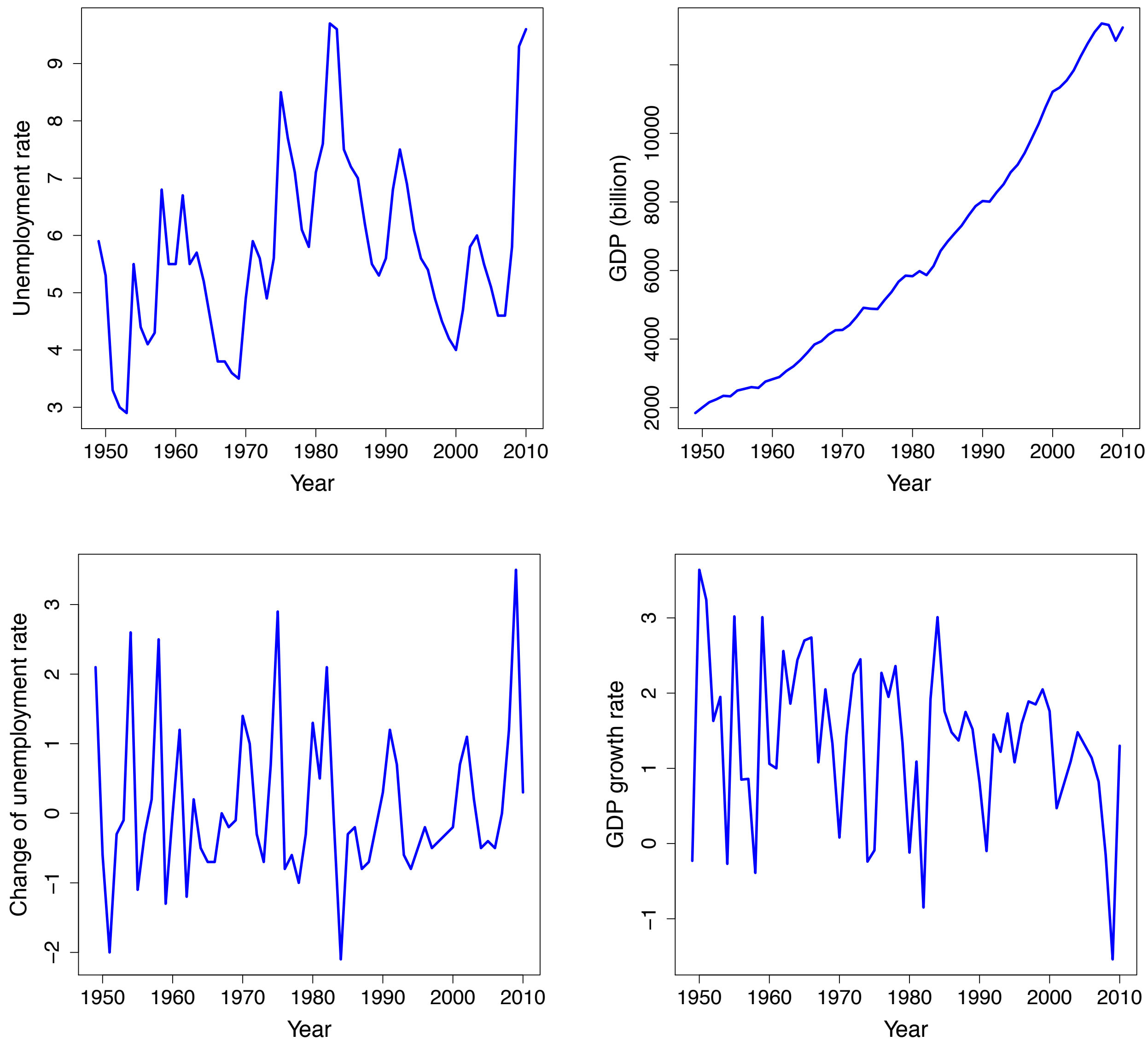
to the sample data. Local linear estimators of $\beta_0(t)$ and $\beta_1(t)$ were computed with bandwidth chosen through CV and Bayesian sampling under the GK error density. The graphs of $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$ are presented in Figure 3.

We found that the local linear estimators with bandwidth estimated through Bayesian sampling are respectively, similar to those with bandwidth selected through CV. However, the error density estimator with bandwidth derived through Bayesian sampling is different from the one with bandwidth selected through CV based on residuals. We plotted the two error–density estimators in Figure 4, where we could find clear differences at their peaks and tails.

The MSE values of the constant coefficient model and time–varying coefficient model are respectively, 0.2894 and 0.2455. The introduction of time–varying coefficients improves model fitting by 15% in comparison to the constant coefficient model.

We derived the 95% point–wise confidence intervals of the two time–varying coefficients based on the asymptotic results obtained in Cai (2007) and we plotted the confidence intervals in Figure 5.

Figure 2: *Time series plots of unemployment rate, GDP, change of unemployment rate and GDP growth rate.*



According to the patterns of the estimated time–varying coefficients shown in Figure 3, a linear function of time might be appropriate to approximate each coefficient function. Consider the time–varying coefficient model given by

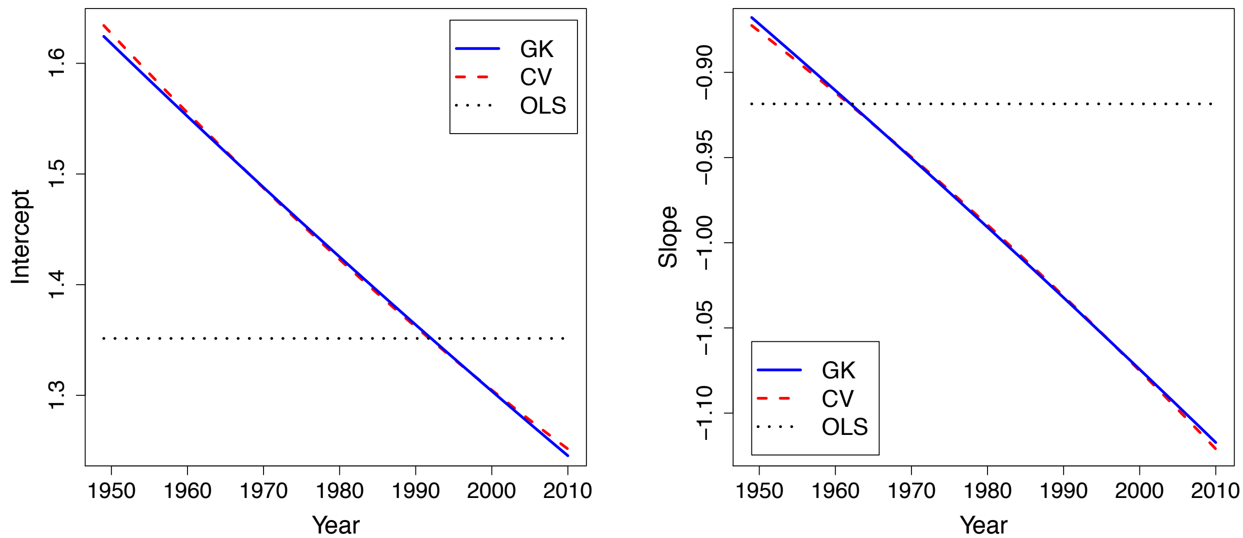
$$\Delta y_t = \beta_0^*(t) + \beta_1^*(t)\Delta x_t + \varepsilon_t, \tag{20}$$

where $\beta_j^*(t)$, for $j = 0$ and 1, are known–form functions of t/T given by

$$\beta_0^*(t) = c_0 + d_0 \times t/T, \qquad \beta_1^*(t) = c_1 + d_1 \times t/T. \tag{21}$$

The least squares estimates of the coefficients are respectively, $(\widehat{c}_0, \widehat{d}_0) = (1.6262, -0.3852)$ and

Figure 3: *Local linear estimates of $\beta_0(t)$ and $\beta_1(t)$ with bandwidth derived through CV and Bayesian sampling under GK error density.*



$(\widehat{c}_1, \widehat{d}_1) = (-0.8613, -0.2535)$. The p values of the four estimates are all zeros, and this indicates that all four coefficients are significant at the 1% level.

The graphs of the calibrated coefficient functions are presented in Figure 5, where we could find that each linear function coefficient is very close to the corresponding local linear estimator. Moreover, the MSE computed based on the calibrated time–varying coefficients is 0.2456, which is very close to that computed based on the local linear estimators of the coefficients. Figure 5 also shows that the 95% point–wise confidence intervals covered the calibrated time–varying coefficients, respectively. Thus, the calibrated coefficient functions are appropriate to capture the time–varying dynamics of Okun's coefficient. In order to validate the calibrated linear time trends for the time–varying coefficients, we tested the null hypothesis that two time–varying coefficients are linear functions of time against the alternative of local linear estimators. The test procedure is based on bootstrapping with 200 repetitions. See the Appendix for details. The wild bootstrapping test produced a p value of 0.1891. Therefore, we could not reject the null hypothesis. Moreover, our finding is consistent with Blanchard (1999), who found that the Okun's coefficient was supposed to be decreasing over time.

Figure 4: *Graphs of the estimated error densities of the time–varying coefficient Okun's law with bandwidths estimated through CV and Bayesian sampling under GK error density.*

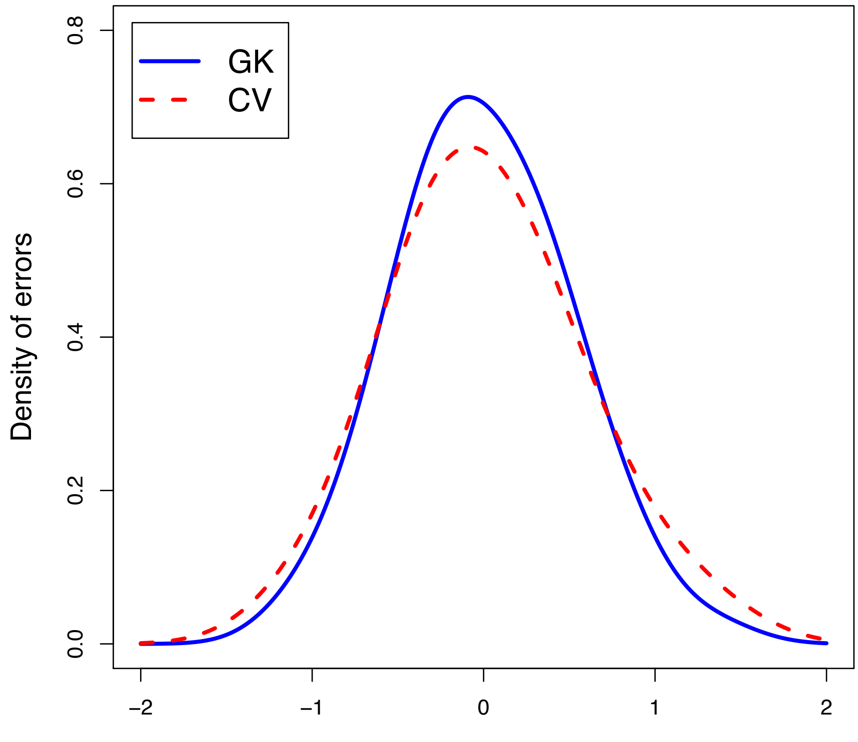
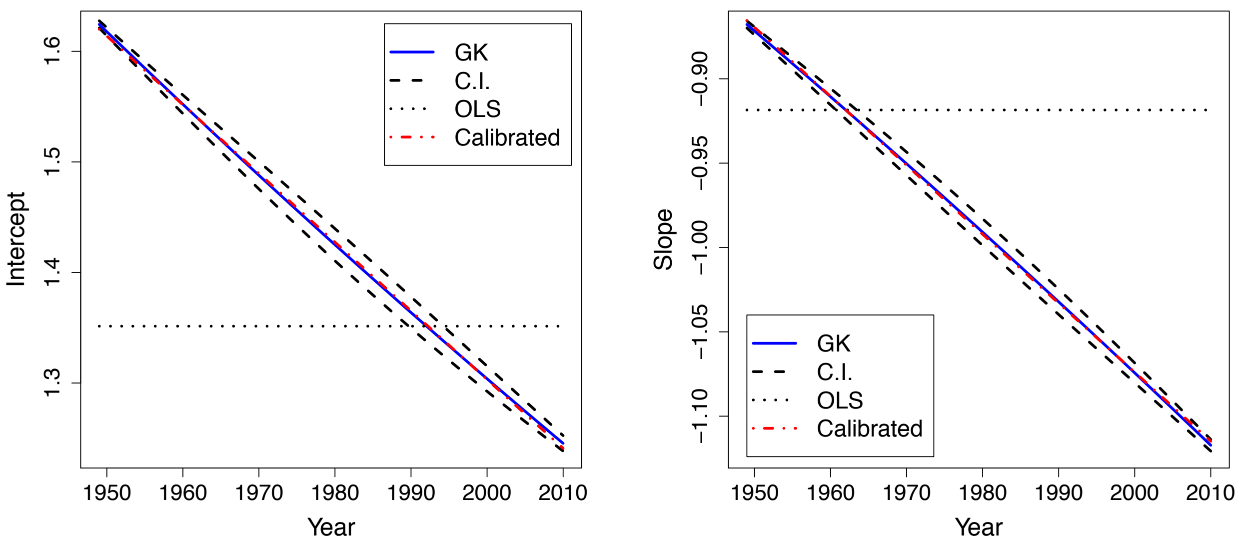


Figure 5: *Local linear estimators of $\beta_0(t)$ and $\beta_1(t)$ and their calibrated linear functions with 95% point–wise confidence interval.*



5.2 Consumption growth

The study on the relationship between consumption growth and income growth has been extensively investigated in empirical macroeconomics during the past several decades. Such a relationship is usually investigated through the linear model given by

$$y_t = \alpha_0 + \alpha_1 x_t + \varepsilon_t, \tag{22}$$

where y_t is the consumption growth, which is the first difference of logarithm of the consumption expenditure, and x_t is the income growth, which is the first difference of logarithm of the disposable income. Carrol and Summers (1991) estimated the coefficient of income growth rate α_1, which is 0.601 based on the data for 15 OECD countries during the period from 1960 to 1985. Campbell and Mankiw (1989) estimated α_1 using the U.S. quarterly data during the period from 1953 to 1986. They found that the estimate of α_1 is 0.316 through OLS, and that this estimate becomes larger when the lagged consumption growth rates are used as instrumental variables.

The relationship between consumption growth and income growth may not necessarily be a constant over time. For example, if a consumer receives additional information on his/her present or future income, he/she would adjust his/her level of consumption discontinuously to be consistent with his/her new inter–temporal budget constraint. If the consumer knows that the income growth rate will increase, he/she will adjust his/her level and rate of consumption. However, the model given by (22) indicates that the response only depends on the level of income growth rate. In this section, we treat the two coefficients of (22) are time–varying, and conduct hypothesis testing to examine whether the coefficients are constants and time–varying.

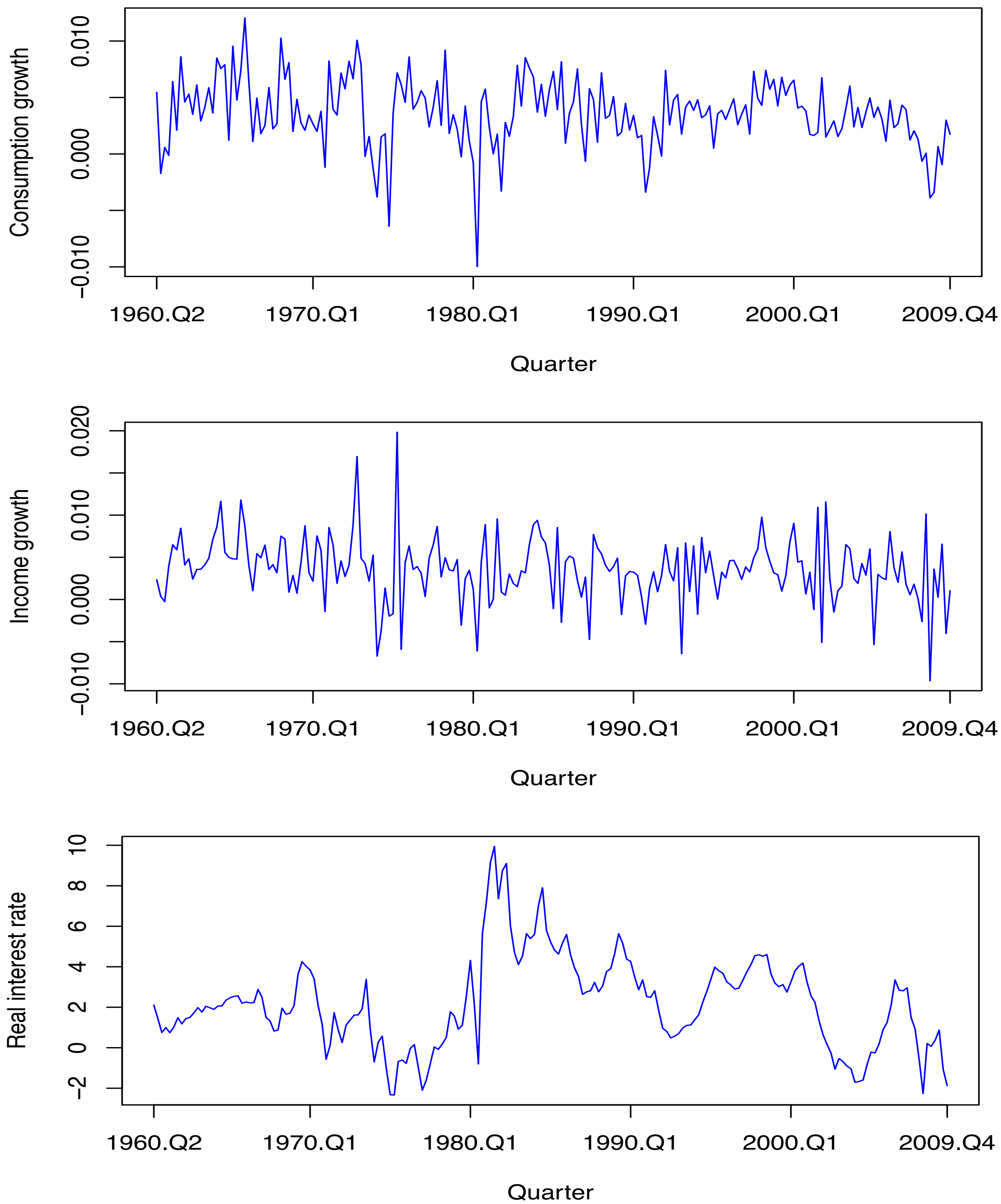
We now investigate the time–varying relationship between consumers' consumption expenditure growth and disposable income growth during the period from 1960 to 2009 in the U.S. The data are quarterly and are available at the website of the Bureau of Economic Analysis. The sample size is $n = 199$. The time series plots of y_t, x_t and the real interest rate are presented in Figure 6.

If the two coefficients are assumed to be constants, the OLS estimates of α_0 and α_1 are respectively, 0.0024 and 0.3399, with their p values being both zero. Assuming the two coefficients are time–varying, we fitted the model given by

$$y_t = \beta_0(t) + \beta_1(t) x_t + \varepsilon_t$$

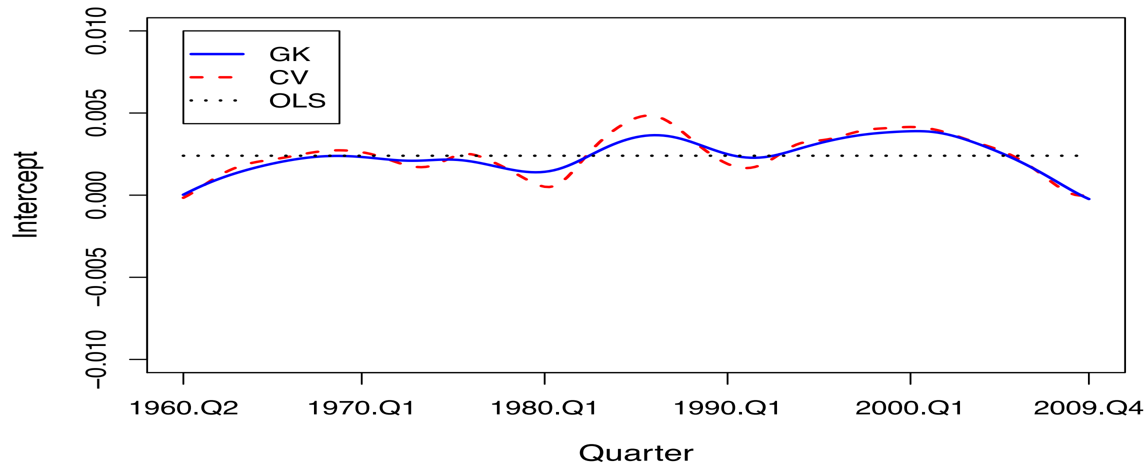
to the above sample, where the errors are assumed to be independent and identically distributed and follow an unknown distribution. Local linear estimators of $\beta_0(t)$ and $\beta_1(t)$ were computed

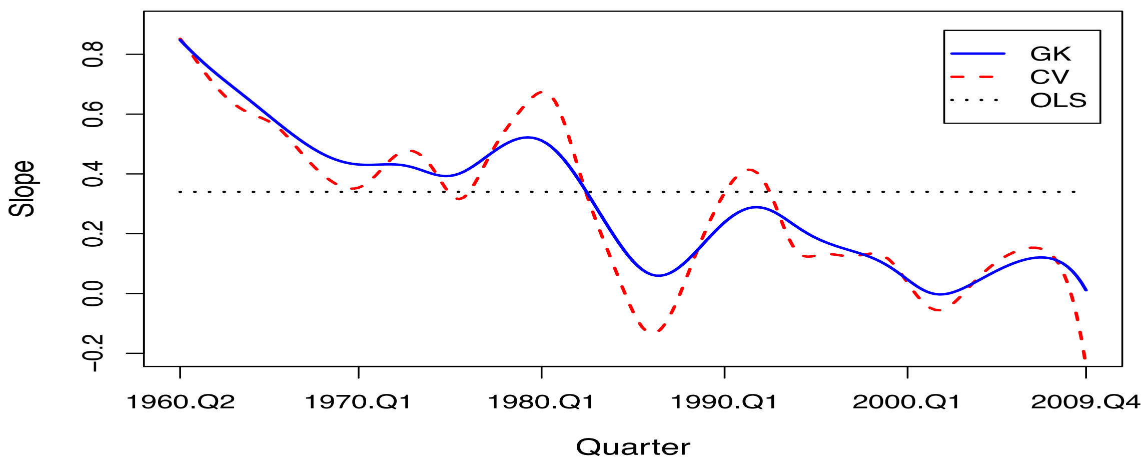
Figure 6: *Time series plots of consumption growth, income growth and real interest rate.*



with bandwidth estimated through the proposed Bayesian sampling under the GK error density. For comparison purpose, we also used CV to choose the bandwidths. The graphs of $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$ with their bandwidths estimated/chosen through Bayesian sampling and CV, are presented in Figure 7. Note that the local linear estimators of the two time–varying coefficients are determined by the bandwidth. The time–varying coefficients with their bandwidth estimated through Bayesian sampling are clearly different from those with their bandwidth chosen through the CV.

Figure 7: *Local linear estimates of $\beta_0(t)$ and $\beta_1(t)$ with bandwidth derived through CV and Bayesian sampling under GK error density.*



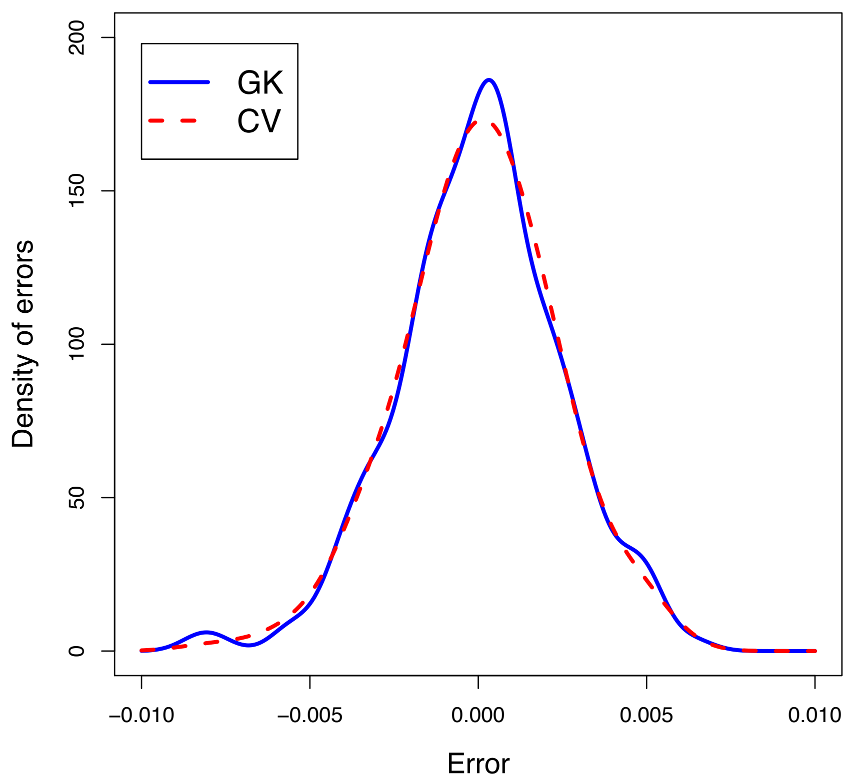


We find that with the bandwidth estimated through Bayesian sampling, the long–term trend of the slope coefficient is decreasing. This reflects a weakening relationship between consumption growth and disposable income growth during the sample period. However, dividing the sample period into several segments, we find that the slope coefficient was decreasing during the period from 1960 to 1968, which was followed by a flat period till 1974. The slope coefficient demonstrated an increasing trend between 1975 and 1979 and an obvious decreasing trend from 1980 to 1986. The slope coefficient had an increasing trend during the period from 1987 to 1991, which was followed by a decreasing period till 2002. The slope coefficient had an increasing trend from 2003 to 2007 and a decreasing trend from 2008 onwards.

The slope coefficient can be interpreted as the marginal propensity to consume (MPC). A possible explanation is that in aggregation, households increase their spending as their disposable income rises and households consume a smaller and smaller proportion of disposable income as disposable income increases. The MPC is the ratio of a change in consumption over a change in income that caused such a consumption change. From the data, we found that disposable income demonstrated an increasing trend, therefore the MPC or the slope coefficient had a long–term decreasing trend. For each specific segment during the sample period, the slope coefficient exhibited up and down due to different rates of economy growth.

The error density estimator with bandwidth estimated through Bayesian sampling differs from the that with bandwidth selected through CV based on residuals. We plotted the two error–density estimators in Figure 8, where we could find clear differences at their peaks and tails.

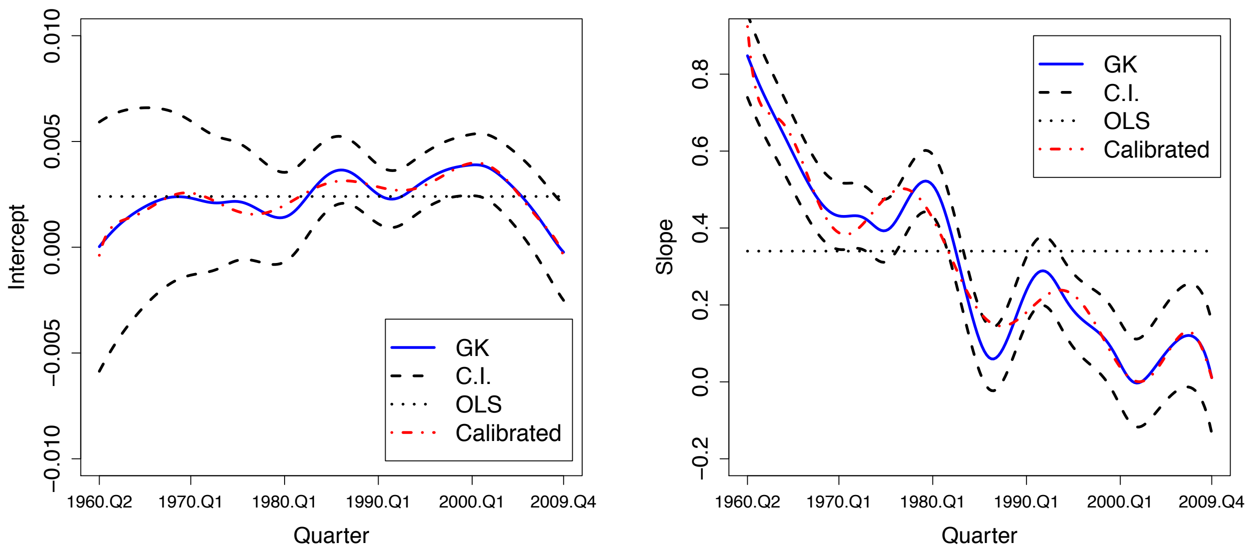
Figure 8: *Graphs of the estimated error densities with bandwidths estimated through CV and Bayesian sampling under GK error density.*



The mean squared errors (MSE) of residuals is 5.8004×10^{-6} for the time–varying coefficient model, and is 7.3398×10^{-6} for the constant–coefficient model. The introduction of time–varying coefficients improves model fitting by 21.0% in comparison to the constant coefficient model.

We derived the asymptotic 95% point–wise confidence intervals of the two time–varying coefficients, which were plotted in Figure 9. The construction of the confidence intervals is same as the one for the time–varying coefficients in Okun's law.

Figure 9: *Local linear estimators of $\beta_0(t)$ and $\beta_1(t)$ and their calibrated linear functions with 95% point–wise confidence interval.*



In order to examine whether the coefficients are time–varying, we tested the null hypothesis of constant coefficients against the alternative of time–varying coefficients. The test procedure is the same to the one used in Section 5.1 and is described in the Appendix. The wild bootstrapping test with 200 repetitions produced a p value that is approximately 0. Therefore, we rejected the null hypothesis of constant coefficients in the linear regression model at the 1% significance level.

According to the patterns of the estimated time–varying coefficients shown in Figure 7, a polynomial function of time might be appropriate to approximate each coefficient function. Consider the time–varying coefficient model given by

$$y_t = \beta_0^*(t) + \beta_1^*(t) x_t + \varepsilon_t, \tag{23}$$

where $\beta_j^*(t)$, for $j = 0$ and 1, are assumed to be the tenth–order polynomials of $\tau_t = t/T$, respectively. Fitting this model to the sample, we derived the coefficient estimates of the two polynomials. The two estimated polynomials are

$$\begin{aligned}\beta_0^*(t) =& -0.0009 + 0.1116\tau_t - 2.442\tau_t^2 + 28.77\tau_t^3 - 184.6\tau_t^4 + 687.2\tau_t^5 \\ & - 1554\tau_t^6 + 2165\tau_t^7 - 1816\tau_t^8 + 840.5\tau_t^9 - 164.9\tau_t^{10}, \\ \beta_1^*(t) =& 1.006 - 18.43\tau_t + 440.8\tau_t^2 - 5303\tau_t^3 + 33510\tau_t^4 - 121700\tau_t^5 \\ & + 267600\tau_t^6 - 362400\tau_t^7 + 295600\tau_t^8 - 133100\tau_t^9 + 25430\tau_t^{10}.\end{aligned}$$

The p values of all estimated coefficients are zeros, indicating that all coefficients are significant at the 1% significance level.

The graphs of the above–calibrated coefficient functions are presented in Figure 9, where we could find that each polynomial coefficient is very close to its corresponding local linear estimator. Moreover, the MSE computed based on the calibrated time–varying coefficients is 6.2117×10^{-6}, which is very close to that computed based on the local linear estimators of the coefficients. Figure 9 also shows that the 95% point–wise confidence intervals covered the calibrated time–varying coefficients, respectively. Therefore, the calibrated polynomial coefficients are appropriate to capture the time–varying nature of the coefficients.

In order to validate the calibrated polynomials for the time–varying coefficients, we tested the null hypothesis that the two time–varying coefficients are the tenth–order polynomials of time against the alternative of local linear estimators. The test procedure is the same as the one used in Section 5.1. The wild bootstrapping test with 200 repetitions produced a p value of 0.1592. Therefore, we could not reject the null hypothesis at 5% significance level.

6 Conclusion

In this paper, we have presented a Bayesian sampling approach to bandwidth estimation for the local linear estimator of time–varying coefficients in linear regression models, where the errors are assumed to follow the Gaussian kernel error density. From Monte Carlo simulation studies, we have found that our proposed Bayesian method achieves better performance than the NRR and CV in estimating the bandwidths for local linear estimators in the regression function. Compared with the parametric assumption of either the Gaussian or the mixture of two Gaussians, the GK error density in the underlying time–varying coefficient time series model helps achieve robustness in terms of different specifications of the true error density.

Applying the proposed sampling algorithm to the estimation of bandwidths for the time–varying coefficient regression models that reflect Okun's law and the relationship between consumption growth and income growth, we have found that this sampling approach works very well for the time–varying coefficient time series model. For both applications, we have found that the estimated error densities derived through CV and Bayesian sampling with GK error density are obviously different to each other at their peaks and tails. In addition, based on the corresponding nonparametric estimates, we have proposed linear functions as approximations to time–varying coefficients of Okun's law and tenth–order polynomials as approximations to time–varying coefficients of consumption growth model.

Appendix: Nonparametric Specification Testing

To examine whether the coefficients are time–varying, we test the null hypothesis

$$\mathrm{H}_0 : \beta_0(t) = \alpha_0,\ \beta_1(t) = \alpha_1,$$

against the alternative hypothesis given by

$$\mathrm{H}_1 : \beta_0(t) \text{ and } \beta_1(t) \text{ have a nonparametric specification.}$$

The test statistic is defined as

$$TS = \frac{RSS_0 - RSS_1}{RSS_1}, \tag{24}$$

where RSS_0 is the residual sum of square (RSS) under the null hypothesis, and RSS_1 is the RSS under the alternative hypothesis. The null hypothesis is rejected for a large value of TS. The p–value is computed by employing the following bootstrap procedure.

1) Estimate $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t))'$ by local linear estimation and get the resulting estimator $\widehat{\boldsymbol{\beta}}(t)$, $t = 1, 2, \cdots, n$.

2) For each $t = 1, 2, \cdots, n$, generate

$$y_t^* = \boldsymbol{x}'_t \widehat{\boldsymbol{\beta}}(t) + \varepsilon_t^*,$$

where $\varepsilon_t^* = \widehat{\sigma} u_t$, $\widehat{\sigma}$ is the standard deviation of $\{y_t - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}(t) : t = 1, 2, \cdots, n\}$ and $\{u_t\}$ is a sequence of independent and identically distributed random variables drawn from a prespecified distribution with mean zero and unit variance, such as N(0, 1).

3) Use the data set $\{(y_t^*, \boldsymbol{x}_t) : t = 1, 2, \cdots, n\}$ to estimate $\alpha = (\alpha_0, \alpha_1)'$ and $\boldsymbol{\beta}(\cdot)$. Calculate the corresponding RSS_0^* and RSS_1^*, and then compute TS^*.

4) Repeat Steps 2) and 3) for B times to obtain the empirical distribution for TS^*. Then, the p–value of the test is computed by $\frac{1}{B} \sum_{i=1}^{B} I(TS_i^* \geq TS)$, where $I(\cdot)$ is an indicator function.

References

Blanchard, O. (1999), *Macroeconomics*, Prentice Hall International, London.

Cai, Z. (2002), 'A two–stage approach to additive time series models', *Statistica Neerlandica* **56**, 415–433.

Cai, Z. (2007), 'Trending time–varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**, 163–188.

Cai, Z. and Tiwari, R. C. (2000), 'Application of a local linear autoregressive model to BOD time series', *Environmetrics* **11**, 341–350.

Campbell, J. Y. and Mankiw, N. G. (1989), Consumption, income, and interest rates: Reinterpreting the time series evidence, *in* O. J. Blanchard and S. Fischer, eds, 'NBER Macroeconomics Annual 1989', MIT Press, pp. 185–246.

Carrol, C. D. and Summers, L. H. (1991), Consumption growth parallels income growth: Some new evidence, *in* B. D. Bernheim and J. B. Shoven, eds, 'National Saving and Economic Performan', University of Chicago Press, pp. 305–348.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross–sectional dependence', *Journal of Econometrics* **171**, 71–85.

Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**, 1313–1321.

Efromovich, S. (2005), 'Estimation of the density of regression errors', *The Annals of Statistics* **33**, 2194–2227.

Fan, J., Heckman, N. E. and Wand, M. P. (1995), 'Local polynomial kernel regression for generalized linear models and quasi–likelihood functions', *Journal of the American Statistical Association* **90**, 141–150.

Gao, J. and Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *The Econometrics Journal* **9**, 332–355.

Ghysels, E. (1998), 'On stable factor structures in the pricing of risk: Do time–varying betas help or hurt?', *The Journal of Finance* **53**, 549–573.

Jagannathan, R. and Wang, Z. (1996), 'The conditional CAPM and the cross–section of expected returns', *The Journal of Finance* **51**, 3–53.

Jeffreys, H. (1961), *Theory of Probability*, Oxford University Press, Oxford.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**, 773–795.

Lee, J. (2000), 'The robustness of Okun's law: Evidence from OECD countries', *Journal of Macroeconomics* **22**, 331–356.

Li, D., Chen, J. and Gao, J. (2011), 'Non–parametric time–varying coefficient panel data models with fixed effects', *The Econometrics Journal* **14**, 387–408.

Okun, A. (1962), 'Potential GNP: Its measurement and significance', *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* pp. 89–104.

Robinson, P. M. (1989), Nonparametric estimation of time–varying parameters, *in* P. Hackl, ed., 'Statistical Analysis and Forecasting of Economis Structural Change', Springer, Berlin, pp. 253–264.

Silvapulle, P., Moosa, I. A. and Silvapulle, M. J. (2004), 'Asymmetry in Okun's law', *Canadian Journal of Economics* **37**, 353–374.

Wang, K. Q. (2003), 'Asset pricing with conditioning information: A new test', *The Journal of Finance* **58**, 161–196.

Zhang, X., Brooks, R. D. and King, M. L. (2009), 'A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state–price density estimation', *Journal of Econometrics* **153**, 21–32.

Zhang, X. and King, M. L. (2011), Bayesian semiparametric GARCH models, Working paper, Monash University.
URL: *http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011*

Zhang, X., King, M. L. and Hyndman, R. J. (2006), 'A Bayesian approach to bandwidth selection for multivariate kernel density estimation', *Computational Statistics and Data Analysis* **50**, 3009–3031.

Zhang, X., King, M. L. and Shang, H. L. (2011), Bayesian estimation of bandwidth for a nonparametric regression model with an unknown error density, Working paper, Monash University.
URL: *http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011*