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 for a nonparametric regression model with mixed types of regressors

Xibin Zhang, Maxwell L. King, Han Lin Shang

June 2013

Working Paper 13/13

Bayesian bandwidth selection for a nonparametric regression model with mixed types of regressors

Xibin Zhang,[1] Maxwell L. King,

Department of Econometrics and Business Statistics, Monash University, Australia

Han Lin Shang

ESRC Centre for Population Change, University of Southampton, United Kingdom

May 27, 2013

Abstract: We propose a sampling approach to bandwidth estimation for a nonparametric regression model with continuous and discrete types of regressors and unknown error density. The unknown error density is approximated by a location–mixture of Gaussian densities with means being the individual errors, and variance a constant parameter. This error density has a form of a kernel density estimator of errors with its bandwidth being the common standard deviation. We derive an approximate likelihood and posterior for bandwidth parameters, and a sampling algorithm is also developed. Monte Carlo simulation studies show that the proposed Bayesian sampling approach leads to better accuracy of the resulting estimators, especially the error density estimator, than the cross–validation. We apply the proposed sampling method to bandwidth estimation for a nonparametric regression model of the Australian All Ordinaries (Aord) daily returns on the overnight S&P 500 return and an indicator of the FTSE return. With the estimated bandwidths, we obtain the one–day–ahead density forecast of the Aord return and a distribution–free measure of value–at–risk. We also use the proposed sampling method to estimate bandwidths for the kernel estimator of the joint density of GDP growth rate, its year level and OECD status.

Keywords: cross–validation, exceedance, Nadaraya-Watson estimator, random–walk Metropolis algorithm, unknown error density, value–at–risk

JEL Classification: C11, C14, C35

[1]Address: 900 Dandenong Road, Caulfield East, Victoria 3145, Australia. Telephone: +61 3 99032130. Fax: +61 3 99032007. Email: xibin.zhang@monash.edu.

1 Introduction

Nonparametric regression is an important tool for exploring the unknown relationship between a response variable and a set of explanatory variables also known as regressors. A simple and commonly used estimator of the regression function is the Nadaraya–Watson (NW) estimator proposed by Nadaraya (1964) and Watson (1964). In many empirical applications of nonparametric regression models, regressors are often of mixed types such as continuous and categorical, etc. In such a situation, Li and Racine (2003) proposed estimating the regression function by the NW–type estimator with different types of regressors being assigned different kernel functions. Since their work, there have been many subsequent theoretical and methodological investigations on nonparametric regression with mixed types of regressors (see for example, Racine and Li, 2004; Hall, Li, and Racine, 2007; Li and Racine, 2007; Maasoumi, Racine, and Stengos, 2007; Hayfield and Racine, 2008; Li, Ouyang, and Racine, 2009; Su, Chen, and Ullah, 2009; Ma and Racine, 2013). It has been generally accepted that the performance of the NW estimator is mainly determined by its bandwidths. In the current literature, the cross–validation (CV) technique is often used for choosing bandwidths. This paper aims to investigate a Bayesian sampling approach to bandwidth estimation for the NW estimator involving mixed types of regressors.

The popularity of CV is accredited to its simplicity and reasonably good performance. However, this bandwidth selection method has some limitations. First, the CV technique might fail to produce meaningful bandwidths in some empirical studies. As discussed by Li and Zhou (2005), there are necessary and sufficient conditions to ensure the optimality of CV for bandwidth selection in nonparametric regression models. It may not be possible to examine whether these conditions hold in applications. Second, even though the CV method does not require an assumption about the error density, it provides no direct solutions to error density estimation. However, error density estimation is important for assessing the goodness of fit of a specified error distribution (see for example, Loynes, 1980; Akritas and Van Keilegom, 2001; Cheng and

Sun, 2008); for testing symmetry, skewness and kurtosis of the residual distribution (Ahmad and Li, 1997; Dette et al., 2002; Neumeyer and Dette, 2007); for statistical inference, prediction and model validation (Efromovich, 2005; Muhsal and Neumeyer, 2010); and for estimating the density of the response variable (Samb, 2011; Escanciano and Jacho-Chávez, 2012). Therefore, being able to estimate the error density is as important as being able to estimate the regression function.

We present a Bayesian sampling approach to bandwidth estimation for the NW estimator involving mixed types of regressors, where the independent identically distributed (iid) errors follow a kernel–form error density studied previously by Zhang and King (2011) for GARCH models and Zhang, King, and Shang (2011) for nonparametric regression models with continuous regressors. We develop a sampling algorithm, which we position as an extension to the algorithm proposed by Zhang et al. (2011) in the sense that the regressors are of mixed types. The contribution of our investigation is that our sampling algorithm leads to a posterior estimate of the response density, where the response variable is modeled as an unknown function of continuous and discrete explanatory variables. The importance of such models can be explained through an example.

Suppose we are interested in the distribution of daily returns of the All Ordinaries index of the Australian stock market. Many market analysts believe that since the beginning of the global financial crisis, the Australian stock market generally follows the overnight performance of the U.S. and several European markets. As the U.S. stock market is believed to have a leading effect on other markets worldwide, we may choose the daily return of the S&P 500 index as a regressor, and a binary variable indicating the sign of the daily return of FTSE index as another regressor. With our proposed sampling algorithm, we are able to not only estimate bandwidths for the NW estimator and the kernel–form error density, but also derive a posterior estimate of the response density. None of the existing bandwidth selection methods can achieve this purpose in a simultaneous manner. When CV is used to choose bandwidths for the NW estimator, one might have to apply the likelihood cross–validation to the residuals so as to derive a kernel density estimator of residuals. This would result in a two–stage procedure for choosing bandwidths in the regression and error–density estimators, and we show that it is inferior to our sampling procedure.

We conduct Monte Carlo simulation studies to compare the in–sample and out–of–sample performances between Bayesian sampling and CV in choosing bandwidths for the regression estimator and error density estimator. The proposed Bayesian sampling approach to bandwidth estimation leads to more accurate estimators than the CV approach in most situations, while the latter performs as good as the former in only a few occasions.

Our proposed sampling algorithm is empirically validated through an application to bandwidth estimation in the nonparametric regression of the Aord daily return on the overnight S&P 500 daily return and a binary variable showing the sign of overnight FTSE daily return. An important and very useful output from this sampling algorithm is the one–day–ahead forecasted density of the Aord daily return, which we use to calculate value–at–risk. We also modify the sampling algorithm for the purpose of choosing bandwidths in kernel condition density estimation of GDP growth rate of a country given its OECD status and the year value of growth–rate observations.

The rest of the paper is organized as follows. Section 2 presents a brief description of the NW estimator when regressors contain continuous and discrete variables. In Section 3, we derive the likelihood and the posterior for bandwidth parameters. A sampling algorithm is also presented. Section 4 presents Monte Carlo simulation studies that aim to examine the performance of the proposed sampling method for bandwidth estimation. In Section 5, we use the sampling method to estimate bandwidths for a nonparametric regression model of stock returns. We modify the poposed sampling algorithm to estimate bandwidths in kernel conditional density estimation of a country's GDP growth rate in Section 6. Section 7 concludes the paper.

2 Nadaraya–Watson estimator with discrete regressors

We present a brief description of the NW estimator of the unknown regression function that contains continuous and discrete explanatory variables. More details can be found in the textbook by Li and Racine (2007). We consider the nonparametric regression model given by

$$y_i = m(\boldsymbol{x}_i) + \varepsilon_i, \qquad i = 1, 2, \cdots, n, \tag{1}$$

where y_i is an observation of a scalar response, $\boldsymbol{x}_i = \left(\boldsymbol{x}_i^{(c)}, \boldsymbol{x}_i^{(d)}\right)$ with $\boldsymbol{x}_i^{(c)}$ being a vector of p continuous variables and $\boldsymbol{x}_i^{(d)}$ a vector of q discrete variables, and $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are iid errors with an unknown probability density function denoted as $f(\varepsilon)$. The discrete variables can be either ordered or unordered, which in turn affects the choice of kernel functions.

The flexibility of model (1) stems from the fact that the unknown regression function $m(\cdot)$ does not need to have a specific functional form. With some smoothness properties, $m(\cdot)$ is estimated by the NW estimator of the form given by

$$\widehat{m}(\boldsymbol{x}; \boldsymbol{h}, \boldsymbol{\lambda}) = \frac{\sum_{i=1}^{n} \mathscr{K}_{\boldsymbol{h},\boldsymbol{\lambda}}(\boldsymbol{x}, \boldsymbol{x}_i)\, y_i}{\sum_{i=1}^{n} \mathscr{K}_{\boldsymbol{h},\boldsymbol{\lambda}}(\boldsymbol{x}, \boldsymbol{x}_i)}, \tag{2}$$

where $\mathscr{K}_{\boldsymbol{h},\boldsymbol{\lambda}}(\boldsymbol{x}, \boldsymbol{x}_i) = K_{\boldsymbol{h}}^{(c)}\left(\boldsymbol{x}^{(c)} - \boldsymbol{x}_i^{(c)}\right) \times K_{\boldsymbol{\lambda}}^{(d)}\left(\boldsymbol{x}^{(d)}, \boldsymbol{x}_i^{(d)}\right)$ is a generalized product kernel that admits continuous and discrete regressors. The kernel function for continuous regressors is a product of p identical Gaussian kernel functions expressed as

$$K_{\boldsymbol{h}}^{(c)}\left(\boldsymbol{x}^{(c)} - \boldsymbol{x}_i^{(c)}\right) = \prod_{j=1}^{p} \frac{1}{h_j} \phi\left(\frac{x_j^{(c)} - x_{i,j}^{(c)}}{h_j}\right),$$

where $x_j^{(c)}$ and $x_{i,j}^{(c)}$ are respectively, the jth elements of $\boldsymbol{x}^{(c)}$ and $\boldsymbol{x}_i^{(c)}$, $\boldsymbol{h} = (h_1, h_2, \cdots, h_p)'$ is a vector of bandwidths associated with the p continuous regressors, and $\phi(\cdot)$ is the standard Gaussian density being used as the kernel function for a continuous variable throughout this paper.

If the jth element of $\boldsymbol{x}^{(d)}$ is binary, the kernel function is Aitchison and Aitken's (1976) kernel given by

$$K_{\lambda_j}^{(d)}\left(x_j^{(d)}, x_{i,j}^{(d)}\right) = \begin{cases} 1 & \text{if } x_j^{(d)} = x_{i,j}^{(d)} \\ \lambda_j & \text{otherwise} \end{cases}, \tag{3}$$

where $x_j^{(d)}$ and $x_{i,j}^{(d)}$ are respectively, the jth elements of $\boldsymbol{x}^{(d)}$ and $\boldsymbol{x}_i^{(d)}$, and $\lambda_j \in (0, 0.5)$ is the bandwidth, for $j = 1, 2, \cdots, q$. If the jth element of $\boldsymbol{x}^{(d)}$ is categorical, its kernel is Li and Racine's (2007) kernel expressed as

$$K_{\lambda_j}^{(d)}\left(x_j^{(d)}, x_{i,j}^{(d)}\right) = \begin{cases} 1 & \text{if } x_j^{(d)} = x_{i,j}^{(d)} \\ \lambda_j^{\left|x_j^{(d)} - x_{i,j}^{(d)}\right|} & \text{otherwise} \end{cases}, \tag{4}$$

for $j = 1, 2, \cdots, q$, and $i = 1, 2, \cdots, n$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_q)'$ is a vector of bandwidths assigned to the corresponding q discrete regressors. These bandwidths are restricted to be within $(0, 1)$.

When discrete explanatory variables are included in a nonparametric regression model, there exists a conventional frequency estimator of the regression function in the literature. This frequency approach involves splitting data into cells based on different values of the discrete variables and using the data in each cell to derive an estimator of the regression function. Li and Racine (2007) found that the NW estimator given by (2) is strongly supported against the frequency estimator for both theoretical and practical reasons.

It is noteworthy that if $\lambda_j = 0$ for all $j = 1, 2, \cdots, q$, then $K_{\lambda_j}^{(d)}\left(x_j^{(d)}, x_{i,j}^{(d)}\right)$ becomes an indicator function taking values 1 and 0, corresponding to the classical frequency estimator. When $\lambda_j = 1$, this indicates that $x_j^{(d)}$ is "smoothed out" and becomes an irrelevant variable (Hall, Li, and Racine, 2007). Similarly, if h_j is very large in the kernel function given by (2), it indicates that $x_j^{(c)}$ is smoothed out and has no explanatory effect on the response variable. As a by–product, the ability of distinguishing irrelevant variables from relevant variables makes the resulting NW estimator very attractive, in comparison to the frequency estimator.

The performance of the NW estimator is mainly determined by its bandwidths, and in the current literature, bandwidths are often chosen through CV with the CV function defined as

$$\mathrm{CV}(\boldsymbol{h},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \widehat{m}_{(-i)}\left(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda}\right)\right]^2 w(\boldsymbol{x}_i),$$

where $\widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})$ is the leave–one–out NW estimator given by

$$\widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda}) = \frac{\sum_{j=1;j\neq i}^{n}\mathcal{K}_{\boldsymbol{h},\boldsymbol{\lambda}}\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)y_j}{\sum_{j=1;j\neq i}^{n}\mathcal{K}_{\boldsymbol{h},\boldsymbol{\lambda}}\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)},$$

and $w(\cdot)$ is a weight function taking values in $[0,1]$. The purpose of introducing the weight function into $\mathrm{CV}(\boldsymbol{h},\boldsymbol{\lambda})$ is to avoid difficulties caused either by division by zero or by the slow convergence rate when $\boldsymbol{x}_i$ is near the boundary of the support of $\boldsymbol{x}$ (see Li and Zhou, 2005). We follow the work by Su, Chen, and Ullah (2009) and choose

$$w(\boldsymbol{x}_i) = \prod_{j=1}^{p+q}\boldsymbol{I}\left(\left|x_{i,j} - \overline{x}_i\right| \leq 1.5 s_i\right), \tag{5}$$

where $\boldsymbol{I}(\cdot)$ is an indicator function, and $\overline{x}_i$ and s_i are the sample mean and standard deviation of the ith explanatory variable, for $i = 1,2,\cdots,p+q$.

In some empirical studies, the CV might fail to produce meaningful results of the bandwidths, and one possible reason is described as follows. Li and Zhou (2005) argued that there are some conditions to ensure the optimality of the CV function. However, it is impossible to check whether all these conditions hold due to the unknown regression function. This problem has motivated us to investigate an alternative approach to bandwidth estimation, namely a Bayesian sampling approach. Note that our sampling approach does not require the weighting function given by (5) as required by the CV.

3 Bayesian estimation of bandwidths

We now consider the nonparametric regression model given by (1) and assume that the iid errors follow an unknown distribution with its density approximated by

$$\widetilde{f}(\varepsilon;b) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b}\phi\left(\frac{\varepsilon-\varepsilon_i}{b}\right), \tag{6}$$

where $\phi(\cdot)$ is the standard Gaussian probability density function. Being previously introduced by Zhang and King (2011) into GARCH models, this density function of errors is a mixture of n identical standard Gaussian densities with different means located at individual errors and a common standard deviation b. Moreover, from the view of kernel smoothing, this error density can be regarded as a kernel density based on independent errors, where b plays the role of bandwidth or smoothing parameter. Note that the error density given by (6) is different from the one proposed by Yuan and de Gooijer (2007), who proposed using residuals as proxies of errors and employed the kernel density estimator of residuals to approximate the true error density.

In this paper, we investigate bandwidth estimation in the NW estimator of the regression function, which is an unknown function of continuous and discrete regressors, as well as the bandwidth in the kernel–form error density given by (6). Prior to our proposed investigation, Zhang et al. (2011) introduced the error density given by (6) into the nonparametric regression model with continuous regressors, where the unknown regression function was estimated by the NW estimator. They presented a sampling algorithm that is able to simultaneously sample the bandwidths in the NW estimator and the kernel–form error density. Consequently, our investigation represents an extension of the sampling algorithm proposed by Zhang et al. (2011) by considering a wider group of regressors than theirs.

In order to conduct Bayesian sampling for the purpose of estimating bandwidths, we treat the bandwidths in the NW estimator of the regression function and the kernel–form error density as parameters. Even though chosen bandwidths depend on sample size as revealed by existing asymptotic results, such a treatment will not cause problems for a fixed–size sample.

Let $(y_i, \boldsymbol{x}_i)$, for $i = 1, 2, \cdots, n$, denote the observations of $(y, \boldsymbol{x})$. Under the error density given by (6), we have

$$y_i \sim \widetilde{f}\left(\{y_i - m(\boldsymbol{x}_i)\}; b\right),$$

for $i = 1, 2, \cdots, n$. As $m(\cdot)$ is unknown, we replace it with its leave–one–out NW estimator. Thus, the density of y_i is approximated by

$$\widetilde{f}\left(\{y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})\};b\right) \approx \frac{1}{n}\sum_{j=1}^{n}\frac{1}{b}\phi\left(\frac{\{y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})\} - \{y_j - \widehat{m}_{(-j)}(\boldsymbol{x}_j;\boldsymbol{h},\boldsymbol{\lambda})\}}{b}\right), \quad (7)$$

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

3.1 An approximate likelihood

The vector of all bandwidths denoted as $(\boldsymbol{h}', \boldsymbol{\lambda}', b)'$ are treated as parameters, given which we wish to derive an approximate likelihood of $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$. For this purpose, we cannot use the approximate density of y_i given by (7) directly because it contains an undesirable term $\phi(0)/b$. The existence of this term will make the likelihood function arbitrarily large by letting b be arbitrarily small. Following the suggestion of Zhang et al. (2011), we exclude the jth term when $\{y_j - \widehat{m}_{(-j)}(\boldsymbol{x}_j;\boldsymbol{h},\boldsymbol{\lambda})\} = \{y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})\}$, from the summation given in (7). Let

$$J_i = \left\{j : y_j - \widehat{m}_{(-j)}(\boldsymbol{x}_j;\boldsymbol{h},\boldsymbol{\lambda}) \neq y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda}), \text{ for } j = 1, 2, \cdots, n\right\},$$

and n_i the number of terms excluded from the summation in (7). Therefore, the density of y_i is approximated by

$$\widetilde{f}\left(\{y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})\};b\right) = \frac{1}{n-n_i}\sum_{j\in J_i}\frac{1}{b}\phi\left(\frac{\{y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})\} - \{y_j - \widehat{m}_{(-j)}(\boldsymbol{x}_j;\boldsymbol{h},\boldsymbol{\lambda})\}}{b}\right),$$

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

Let $\boldsymbol{h}^2$ denote a row vector whose elements are the squared elements of $\boldsymbol{h}$. Thus, given $(\boldsymbol{h}^2, \boldsymbol{\lambda}', b^2)'$, the likelihood of $\boldsymbol{y}$ is approximated by

$$\ell\left(\boldsymbol{y}|\boldsymbol{h}^2, \boldsymbol{\lambda}, b^2\right) \approx \prod_{i=1}^{n}\left\{\frac{1}{n-n_i}\sum_{j\in J_i}\frac{1}{b}\phi\left(\frac{\{y_i - \widehat{m}_{(-i)}(\boldsymbol{x}_i;\boldsymbol{h},\boldsymbol{\lambda})\} - \{y_j - \widehat{m}_{(-j)}(\boldsymbol{x}_j;\boldsymbol{h},\boldsymbol{\lambda})\}}{b}\right)\right\}. \quad (8)$$

3.2 Priors

We follow Zhang et al.'s (2011) choices of priors of bandwidths for continuous regressors and the kernel–form error density. Let $\pi(h_j^2)$ denote the prior of h_j^2, for $j = 1, 2, \cdots, p$, and $\pi(b^2)$ the prior of b^2. As the Gaussian kernel is used for each variate in this situation, each bandwidth is also considered as the standard deviation of the corresponding Gaussian density. Therefore, the prior of each squared bandwidth is chosen to be an inverse Gamma density. Thus, the prior of h_j^2 is

$$\pi\left(h_j^2\right) = \frac{(\beta_h)^{\alpha_h}}{\Gamma(\alpha_h)}\left(\frac{1}{h_j^2}\right)^{\alpha_h+1} \exp\left\{-\frac{\beta_h}{h_j^2}\right\}, \text{ for } j = 1, 2, \cdots, p,$$

and the prior of b^2 is

$$\pi\left(b^2\right) = \frac{(\beta_b)^{\alpha_b}}{\Gamma(\alpha_b)}\left(\frac{1}{b^2}\right)^{\alpha_b+1} \exp\left\{-\frac{\beta_b}{b^2}\right\},$$

where the hyperparameters are chosen as $\alpha_h = \alpha_b = 1$ and $\beta_h = \beta_b = 0.05$ (see for example, Geweke, 2009).

Let $\pi(\lambda_j)$ denote the prior of λ_j, the bandwidth assigned to the jth discrete regressor, for $j = 1, 2, \cdots, q$. The prior of λ_j is assumed to be a uniform density defined on (z_a, z_b). For binary regressors, $z_a = 0$ and $z_b = 0.5$; and for ordered categorical regressors, $z_a = 0$ and $z_b = 1$. See for example, Hayfield and Racine (2008), for discussion of restrictions on these smoothing parameters.

The joint prior of $\left(\boldsymbol{h}^2, \boldsymbol{\lambda}', b^2\right)'$ is the product of all the marginal priors and is denoted as $\pi\left(\boldsymbol{h}^2, \boldsymbol{\lambda}, b^2\right)$.

3.3 An approximate posterior

An approximate posterior of $\left(\boldsymbol{h}^2, \boldsymbol{\lambda}', b^2\right)'$ is obtained as the product of the approximate likelihood given by (8) and the joint prior, and is expressed as (up to a normalizing constant)

$$\pi\left(\boldsymbol{h}^2, \boldsymbol{\lambda}, b^2 \middle| \boldsymbol{y}\right) \propto \ell\left(\boldsymbol{y} | \boldsymbol{h}^2, \boldsymbol{\lambda}, b^2\right) \times \pi\left(\boldsymbol{h}^2, \boldsymbol{\lambda}, b^2\right). \tag{9}$$

We use the adaptive random–walk Metropolis algorithm proposed by Garthwaite, Fan, and Sisson (2010) to sample $(\boldsymbol{h}^2, \boldsymbol{\lambda}', b^2)'$. The sampling procedure is described as follows.

Step 1: Specify a Gaussian proposal distribution, and start the sampling iteration process by choosing an arbitrary value of $(\boldsymbol{h}^2, \boldsymbol{\lambda}', b^2)'$ and denoting it as $(\boldsymbol{h}^2_{(0)}, \boldsymbol{\lambda}'_{(0)}, b^2_{(0)})'$. For example, the elements of $\boldsymbol{h}^2_{(0)}$ and $b^2_{(0)}$ can be any values on $(0,1)$ and the elements of $\boldsymbol{\lambda}_{(0)}$ can be any values on $(0,0.5)$ for binary regressors and $(0,1)$ for categorical regressors.

Step 2: At the kth iteration, the current state $\boldsymbol{h}^2_{(k)}$ is updated as $\boldsymbol{h}^2_{(k)} = \boldsymbol{h}^2_{(k-1)} + \gamma_{(k-1)}\boldsymbol{u}/||\boldsymbol{u}||$, where $\boldsymbol{u}$ is drawn from the proposal density which is the p dimensional standard Gaussian density, and $\gamma_{(k-1)}$ is an adaptive tuning parameter with an arbitrary initial value $\gamma_{(0)}$. The updated $\boldsymbol{h}^2_{(k)}$ is accepted with a probability given by

$$\min\left\{ \frac{\pi\left(\boldsymbol{h}^2_{(k)}, \boldsymbol{\lambda}_{(k-1)}, b^2_{(k-1)} | \boldsymbol{y}\right)}{\pi\left(\boldsymbol{h}^2_{(k-1)}, \boldsymbol{\lambda}_{(k-1)}, b^2_{(k-1)} | \boldsymbol{y}\right)}, 1 \right\}.$$

Step 3: The tuning parameter for the next iteration is set to

$$\gamma_{(k)} = \begin{cases} \gamma_{(k-1)} + c(1-\xi)/k & \text{if } \boldsymbol{h}^2_{(k)} \text{ is accepted} \\ \gamma_{(k-1)} - c\xi/k & \text{if } \boldsymbol{h}^2_{(k)} \text{ is rejected} \end{cases},$$

where $c = \gamma_{(k-1)}/(\xi - \xi^2)$ is a constant, and ξ is the optimal target acceptance probability, which is 0.234 for multivariate updating and 0.44 for univariate updating (see for example, Roberts and Rosenthal, 2009; Garthwaite et al., 2010).

Step 4: Update $\boldsymbol{\lambda}_{(k-1)}$ and $b^2_{(k-1)}$ in the same way as described by Steps 2 and 3.

Step 5: Repeat Steps 2–4, discard the burn–in period of iterations, and the draws after the burn–in period are recorded and denoted as $\left\{\left(\boldsymbol{h}_{(k)}, \boldsymbol{\lambda}'_{(k)}, b_{(k)}\right)' : k = 1, 2, \cdots, M\right\}$.

Upon completing the above iterations, we use the ergodic mean (or posterior mean) of each simulated chain as an estimate of each bandwidth. The diagnostic checking of the mixing performance of each simulated chain is conducted through the simulation inefficiency factor (SIF) discussed by Kim, Shephard, and Chib (1998), Meyer and Yu (2000) and Tse, Zhang, and Yu (2004). The SIF is approximately interpreted as the number of draws needed in order to derive independent draws.

In the following analyses, the burn–in period is taken as $1,000$ iterations and the number of recorded iterations after the burn–in period is $10,000$. The number of batches is 200, and there are 50 draws within each batch.

4 Monte Carlo simulation study

The purpose of this Monte Carlo simulation study is two–fold. First, we demonstrate the merit of the proposed Bayesian sampling approach to bandwidth estimation in comparison to the CV method for bandwidth selection discussed by Racine and Li (2004). Second, we show that the bandwidth for the kernel–form error density estimated through Bayesian sampling leads to a more accurate error–density estimator than its competitor, a two–stage procedure in which bandwidth for the kernel estimator of residuals is selected by likelihood cross–validation.

4.1 Accuracy of regression estimator

The first data generating process (DGP) includes binary and continuous explanatory variables and is given by

$$y_i = \sum_{j=1}^{4} x_{i,j}^{(d)} + \sum_{j=1}^{4} \sum_{\substack{k \neq j \\ k=1}}^{4} \frac{1}{2} x_{i,j}^{(d)} x_{i,k}^{(d)} + \sum_{j=1}^{4} z_{i,j}^{(d)} m_1 \left(x_i^{(c)} \right) + m_2 \left(x_i^{(c)} \right) + u_i,$$

for $i = 1, 2, \cdots, n$, where $x_{i,j}^{(d)}$ takes values of 0 and 1 with an equal probability of 0.5, $j = 1, 2, \cdots, 4$, $z_i^{(c)}$ is drawn from the uniform distribution on $[0, 2]$, $m_1\left(x_i^{(c)}\right) = \sin\left(x_i^{(c)}\pi\right)$; $m_2\left(x_i^{(c)}\right) = x_i^{(c)} - 0.5\left(x_i^{(c)}\right)^2 + 0.3\left(x_i^{(c)}\right)^3$, and u_i is drawn from $N(0, 1)$.

Let $\boldsymbol{x}^{(d)}$ denote a vector of the four discrete regressors, and the relationship between the response and the five regressors is modeled as

$$y_i = m\left(\boldsymbol{x}^{(d)}, x^{(c)}\right) + u_i, \tag{10}$$

where u_i, for $i = 1, 2, \cdots, n$, are assumed to be independent. We consider two approaches to the estimation of $m(\cdot)$. The first approach is the NW estimator, whose bandwidths are estimated by the proposed sampling method under the assumption that independent errors have the kernel–form error density given by (6). For comparison purposes, we also use CV to choose bandwidths for this approach. The second approach is the conventional frequency estimator, which is equivalent to the situation of setting zero bandwidths for all discrete regressors in the NW estimator (Li and Racine, 2003, 2007, Chapter 3). To measure the performance of each approach, we generate $2n$ observations denoted as $\left\{\left(y_i, \boldsymbol{x}_i^{(d)}, x_i^{(c)}\right) : 1 \leq i \leq 2n\right\}$, for $n = 100, 200$ and 500, respectively. We use the first n observations for estimation and in–sample evaluation, and the last n observations for out–of–sample evaluation (see also Su et al., 2009). We consider the average squared error (ASE) as an evaluation measure for both in–sample and out–of–sample evaluation:

$$\text{ASE}_{\text{in}} = \frac{1}{n}\sum_{i=1}^{n}\left[m\left(\boldsymbol{x}_i^{(d)}, x_i^{(c)}\right) - \widehat{m}\left(\boldsymbol{x}_i^{(d)}, x_i^{(c)}; \boldsymbol{h}, \boldsymbol{\lambda}, b\right)\right]^2,$$
$$\text{ASE}_{\text{out}} = \frac{1}{n}\sum_{i=1}^{n}\left[m\left(\boldsymbol{x}_{n+i}^{(d)}, x_{n+i}^{(c)}\right) - \widehat{m}\left(\boldsymbol{x}_{n+i}^{(d)}, x_{n+i}^{(c)}; \boldsymbol{h}, \boldsymbol{\lambda}, b\right)\right]^2 w\left(\boldsymbol{x}_i\right),$$

where $\widehat{m}\left(\boldsymbol{x}_i^{(d)}, x_i^{(c)}; \boldsymbol{h}, \boldsymbol{\lambda}, b\right)$ is the NW estimator of $m(\cdot,\cdot)$ based on the sample of the first n observations with bandwidths selected through Bayesian sampling, and $w(\boldsymbol{x}_i)$ is the weight function given by (5) for the out–of–sample evaluation. The purpose of the weight function is to trim those out–of–sample observations that lie outside the data range of the in–sample observations. When bandwidths of the NW estimator are selected through CV, the bandwidth parameter b should be removed from the expression of $\widehat{m}$; and when the frequency estimator is considered, both $\boldsymbol{\lambda}$ and b should be removed from the expression of $\widehat{m}$. We report the mean, median, standard deviation (SD) and interquartile range (IQR) of ASE averaged over the 1,000 Monte Carlo replications. These results are presented in Table 1.

Table 1 shows that the NW estimator with bandwidths estimated through Bayesian sampling outperforms the same estimator with bandwidths selected through CV, and that the NW estimator, no matter which bandwidth selection method is used, outperforms the conventional frequency estimator. As the sample size n increases, the differences in estimation and forecast

n	Model	In–sample ASE				Out–of–sample ASE			
		Mean	Median	SD	IQR	Mean	Median	SD	IQR
100	Bayesian	0.5562	0.5390	0.1301	0.1520	1.0740	0.9692	0.4451	0.4827
	CV	0.5877	0.5192	0.3017	0.1320	1.0990	0.9927	0.4525	0.5230
	Frequency	0.6098	0.5707	0.1931	0.1590	1.3680	1.2840	0.4858	0.5990
200	Bayesian	0.3571	0.3512	0.0571	0.0750	0.5233	0.4998	0.1567	0.1683
	CV	0.3577	0.3520	0.0567	0.0727	0.5355	0.5070	0.1617	0.1740
	Frequency	0.4115	0.3952	0.0865	0.0877	0.6132	0.5927	0.1533	0.1947
500	Bayesian	0.1994	0.1978	0.0255	0.0332	0.2158	0.2120	0.0415	0.0489
	CV	0.1997	0.1989	0.0253	0.0333	0.2175	0.2146	0.0411	0.0506
	Frequency	0.2298	0.2169	0.0562	0.0436	0.2332	0.2300	0.0420	0.0529

Table 1: *Average and variation measures of the ASE values derived through 1,000 samples simulated according to the first DGP. The blue colored numbers represent the minimum mean ASEs.*

accuracy among the three methods become smaller.

A conclusion drawn from this Monte Carlo simulation result is that generally speaking, Bayesian sampling slightly outperforms CV for the purpose of bandwidth estimation in the NW estimator.

The second DGP includes ordered categorical and continuous variables and is expressed by

$$y_i = x_{i,1}^{(c)} + x_{i,2}^{(c)} + x_{i,1}^{(d)} + x_{i,2}^{(d)} + u_i,$$

where $x_{i,j}^{(c)}$ is drawn from $N(0,1)$, for $j = 1$ and 2, $x_{i,j}^{(d)}$ takes values from $\{0, 1, \cdots, 5\}$ with an equal probability of 1/6, for $j = 1$ and 2, and the error term u_i is drawn from $N(0,1)$, for $i = 1, 2, \cdots, n$.

Let $\boldsymbol{x}^{(c)}$ denote the vector of two continuous regressors and $\boldsymbol{x}^{(d)}$ the vector of two discrete regressors. The relationship between the response and regressors is modeled by (10). To estimate the regression function, we consider the conventional frequency estimator, and four NW–type estimators differing by their kernel functions for discrete regressors — the unordered and ordered kernels. We expect that the ordered kernel should dominate the unordered kernel estimator in a finite sample, because the observations of the discrete regressors have a natural order. The performance of each regression estimator is measured by the in–sample ASE and out–of–sample

ASE. The results are given in Table 2, with four different sample sizes.

n	Model	In–sample ASE				Out–of–sample ASE			
		Mean	Median	SD	IQR	Mean	Median	SD	IQR
100	Bayesian (ordered)	0.5035	0.4925	0.1044	0.1306	0.6236	0.5903	0.2255	0.2993
	Bayesian (unordered)	0.5127	0.4995	0.1096	0.1371	0.6382	0.6003	0.2588	0.2948
	CV (ordered)	0.5535	0.4828	0.1055	0.1348	0.6553	0.6163	0.2469	0.3138
	CV (unordered)	0.6333	0.6232	0.1097	0.1411	0.7958	0.7660	0.2902	0.3525
	Frequency	0.8950	0.8516	0.2504	0.3071	1.9830	1.9150	0.6116	0.7890
200	Bayesian (ordered)	0.3758	0.3689	0.0615	0.0793	0.4085	0.3978	0.1128	0.1525
	Bayesian (unordered)	0.3845	0.3794	0.0619	0.0790	0.4124	0.3998	0.1134	0.1554
	CV (ordered)	0.3812	0.3755	0.0611	0.0778	0.4252	0.4098	0.1205	0.1641
	CV (unordered)	0.4968	0.4918	0.0647	0.0844	0.5456	0.5351	0.1497	0.1928
	Frequency	0.6438	0.6261	0.1348	0.1747	1.0660	1.0430	0.2816	0.3550
500	Bayesian (ordered)	0.2610	0.2580	0.0315	0.0420	0.2466	0.2434	0.0526	0.0726
	Bayesian (unordered)	0.2646	0.2606	0.0381	0.0441	0.2489	0.2441	0.0600	0.0734
	CV (ordered)	0.2633	0.2595	0.0322	0.0438	0.2562	0.2529	0.0563	0.0770
	CV (unordered)	0.3458	0.3425	0.0343	0.0455	0.3325	0.3280	0.0663	0.0885
	Frequency	0.4204	0.4146	0.0610	0.0746	0.4895	0.4883	0.0907	0.1220
1000	Bayesian (ordered)	0.1948	0.1947	0.0185	0.0249	0.1714	0.1715	0.0287	0.0367
	Bayesian (unordered)	0.1965	0.1958	0.0230	0.0253	0.1718	0.1717	0.0286	0.0380
	CV (ordered)	0.1968	0.1955	0.0298	0.0251	0.1731	0.1728	0.0282	0.0370
	CV (unordered)	0.2554	0.2557	0.0210	0.0285	0.2300	0.2289	0.0367	0.0466
	Frequency	0.3040	0.3026	0.0353	0.0486	0.2992	0.2997	0.0450	0.0573

Table 2: *Average and variation measures of the ASE values derived through 1,000 samples simulated according to the second DGP. The blue colored numbers represent the minimum mean ASEs.*

According to the in–sample and out–of–sample ASE measures of the regression estimator, Bayesian sampling for bandwidth estimation performs better than CV, and the choice of ordered discrete kernel outperforms its unordered counterpart. We also find that the conventional frequency estimator performs worse than the NW–type estimator regardless the kernel types for discrete regressors, bandwidth selection methods and sample sizes considered. Moreover, as the sample size increases, the advantage of Bayesian sampling for bandwidth estimation against CV for the same purpose becomes less evident.

A conclusion drawn from this Monte Carlo simulation results is that ordered kernel should be used for ordered categorical variables, and Bayesian sampling performs better than CV for the purpose of bandwidth estimation.

The third DGP includes ordered categorical and continuous regressors and is given by

$$y_i = 1 + \sqrt{x_i^{(d)}} + x_i^{(c)} + u_i, \text{ for } i = 1, 2, \cdots, n,$$

where $x_i^{(d)}$ is drawn from $\{0, 1, \cdots, 4\}$ with an equal probability of $1/5$, $x_i^{(c)}$ is drawn from $N(0,1)$, and the error term u_i is drawn from $N(0,1)$. As the response variable is affected by the discrete regressor through its square root, the actual distance between categories 0 and 1 is 1, between categories 1 and 2 is $\sqrt{2} - 1 \approx 0.41$, between categories 2 and 3 is $\sqrt{3} - \sqrt{2} \approx 0.32$, and between categories 3 and 4 is $\sqrt{4} - \sqrt{3} \approx 0.27$ (see also Racine and Li, 2004).

The purpose of this simulation example is to assess whether the investigated approaches perform adequately in a situation where the distance between any pair of successive observations is not a fixed constant. To estimate the unknown relationship between the response and the two regressors, we employed four NW–type estimators and the conventional frequency estimator, which were used for estimating the regression function in the second DGP. We tabulate the results derived from the third DGP in Table 3, where we present the in–sample and out–of–sample ASEs averaged over the 1,000 Monte Carlo replications for each estimator.

From Table 3, we find that the conventional frequency estimator performs worse than the kernel estimator with bandwidths selected by the Bayesian method, but does better than the kernel estimator with bandwidths selected by CV.

Table 3 shows that the NW estimator with bandwidths estimated through Bayesian sampling outperforms the same estimator with bandwidths selected through CV, regardless which discrete kernel is used, ordered or unordered. Moreover, no matter which bandwidth estimation method is used, Bayesian or CV, the unordered discrete kernel leads to a better performance than its ordered counterpart. However, when Bayesian sampling is used for bandwidth estimation, the result is a little bit surprising, because the unordered discrete kernel outperforms its ordered

n	Model	In–sample ASE				Out–of–sample ASE			
		Mean	Median	SD	IQR	Mean	Median	SD	IQR
50	Bayesian (ordered)	0.2134	0.1979	0.0919	0.1255	0.2410	0.2110	0.1343	0.1570
	Bayesian (unordered)	0.2090	0.1939	0.0884	0.1203	0.2291	0.2006	0.1266	0.1494
	CV (ordered)	0.2311	0.2135	0.0994	0.1170	0.2670	0.2260	0.1674	0.1780
	CV (unordered)	0.2785	0.2628	0.1116	0.1416	0.3023	0.2540	0.1996	0.2267
	Frequency	0.2151	0.1958	0.0963	0.1142	0.3886	0.3525	0.1982	0.2371
100	Bayesian (ordered)	0.1349	0.1282	0.0530	0.0723	0.1375	0.1228	0.0702	0.0801
	Bayesian (unordered)	0.1340	0.1272	0.0521	0.0700	0.1342	0.1216	0.0679	0.0776
	CV (ordered)	0.1605	0.1523	0.0530	0.0684	0.1480	0.1300	0.0795	0.0915
	CV (unordered)	0.1755	0.1676	0.0628	0.0793	0.1610	0.1412	0.0898	0.1041
	Frequency	0.1431	0.1369	0.0532	0.0685	0.2002	0.1852	0.0890	0.1069
200	Bayesian (ordered)	0.0841	0.0808	0.0283	0.0401	0.0790	0.0746	0.0332	0.0404
	Bayesian (unordered)	0.0841	0.0810	0.0282	0.0393	0.0782	0.0737	0.0328	0.0398
	CV (ordered)	0.1178	0.1151	0.0298	0.0383	0.0846	0.0786	0.0382	0.0434
	CV (unordered)	0.1075	0.1044	0.0348	0.0450	0.0884	0.0815	0.0390	0.0478
	Frequency	0.0979	0.0948	0.0298	0.0386	0.1155	0.1033	0.0697	0.0538

Table 3: *Average and variation measures of the ASE values derived through 1,000 samples simulated according to the third DGP. The blue colored numbers represent the minimum mean ASEs.*

counterpart by small margins, which become even smaller as the sample size increases. In contrast, when CV is used for bandwidth selection, the unordered discrete kernel clearly outperform the ordered one.

Once again, the conventional frequency estimator performs worse than the NW–type estimators. It is not surprising that as sample size increases, all the ASE values, as well as the difference between one ASE value and its competing counterpart, becomes smaller.

To conclude this simulation study, we found that the proposed Bayesian sampling outperforms its counterpart, CV, in choosing bandwidths for the NW estimator, and that when the actual distance between any pair of successive observations is not a fixed constant for a categorical regressor, the unordered kernel should be used.

The fourth and fifth DGPs include binary and continuous regressors and are given by

$$y_i = x_{i,1}^{(d)} + x_{i,2}^{(d)} + x_{i,1}^{(c)} + x_{i,2}^{(c)} + u_i,$$

and

$$y_i = x_{i,1}^{(d)} + x_{i,2}^{(d)} + x_{i,1}^{(d)} x_{i,2}^{(d)} + x_{i,1}^{(c)} + x_{i,2}^{(c)} + x_{i,1}^{(c)} x_{i,2}^{(c)} + u_i,$$

for $i = 1, 2, \cdots, n$, where $x_{i,1}^{(d)}$ and $x_{i,2}^{(d)}$ take values from $\{0, 1\}$ with equal probabilities of 0.5, $x_{i,1}^{(c)}$ and $x_{i,2}^{(c)}$ are drawn from $N(0, 1)$, and u_i is drawn from $N(0, 1)$. The fifth DGP differs from the fourth by the inclusion of two interaction terms.

To estimate the regression functions for the two DGPs, we employed the NW estimator with bandwidths estimated/selected through Bayesian sampling and CV, as well as the conventional frequency estimator. The in–sample and out–of–sample average ASE values for both DGPs were tabulated in Table 4.

For the fourth DGP, Bayesian sampling leads to a slightly better NW estimator than the CV, although the difference between any pair of competing ASE values is marginal. However, for the fifth DGP, Bayesian sampling clearly outperforms its competing counterpart. As the sample size increases, the difference between any pair of competing ASE values becomes smaller. Furthermore, no matter which bandwidth estimation/selection method is used, the NW estimator outperforms the conventional frequency estimator.

4.2 Accuracy of the error density estimator

The proposed Bayesian sampling algorithm is based on the assumption of kernel–form error density given by (6), whose bandwidth is sampled at the same time when bandwidths of the NW estimator are sampled. Upon completion of the sampling algorithm, we also obtain a kernel density estimator of the error density. However, when CV is used for selecting bandwidths for the NW estimator, one may obtain the kernel density estimator of residuals, but its bandwidth has to be selected based on residuals through an existing bandwidth selector such as the likelihood cross–validation. Thus, it requires two stages of using the cross–validation method to select bandwidths for the NW estimator and the kernel density estimator of residuals, and we call it

n	Model	In–sample ASE				Out–of–sample ASE			
		Mean	Median	SD	IQR	Mean	Median	SD	IQR
DGP4									
100	Bayesian	0.3082	0.2947	0.0884	0.1142	0.3296	0.3111	0.1220	0.1455
	CV	0.3188	0.3019	0.1018	0.1297	0.3380	0.3147	0.1413	0.1623
	Frequency	0.3192	0.3004	0.1022	0.1275	0.3846	0.3653	0.1361	0.1644
200	Bayesian	0.2098	0.2037	0.0492	0.0668	0.1990	0.1932	0.0606	0.0791
	CV	0.2114	0.2050	0.0520	0.0682	0.1987	0.1903	0.0608	0.0835
	Frequency	0.2116	0.2050	0.0515	0.0677	0.2228	0.2135	0.0646	0.0807
500	Bayesian	0.1277	0.1258	0.0236	0.0305	0.1100	0.1033	0.0642	0.0331
	CV	0.1282	0.1258	0.0243	0.0315	0.1068	0.1033	0.0269	0.0331
	Frequency	0.1283	0.1257	0.0247	0.0318	0.1165	0.1126	0.0278	0.0344
DGP5									
100	Bayesian	0.3751	0.3643	0.0920	0.1172	0.4715	0.4438	0.1866	0.2401
	CV	0.6132	0.5809	0.2045	0.2438	0.4931	0.4619	0.1925	0.2161
	Frequency	0.6136	0.5795	0.2021	0.2453	0.4917	0.4600	0.1797	0.2310
200	Bayesian	0.2707	0.2647	0.0557	0.0696	0.2829	0.2691	0.0870	0.1139
	CV	0.4218	0.4085	0.1069	0.1329	0.2953	0.2836	0.0974	0.1171
	Frequency	0.4226	0.4086	0.1091	0.1325	0.2917	0.2785	0.0847	0.1056
500	Bayesian	0.1727	0.1706	0.0289	0.0367	0.1533	0.1490	0.0357	0.0481
	CV	0.2619	0.2567	0.0495	0.0647	0.1550	0.1526	0.0348	0.0480
	Frequency	0.2619	0.2567	0.0495	0.0644	0.1557	0.1522	0.0355	0.0490

Table 4: *Average and variation measures of the ASE values derived through 1,000 samples simulated according to the fourth and fifth DGPs. The blue colored numbers represent the minimum mean ASEs.*

two–stage CV.

The performance of a kernel estimator of the error density denoted as $\widehat{f}(\cdot)$, is examined by its integrated squared errors (ISE). In our Monte Carlo simulation studies, the ISE was numerically approximated at 1001 equally spaced grid points on $[-5,5]$:

$$\text{ISE} \approx \sum_{i=1}^{1001} \left\{ f\left(-5+\frac{i-1}{100}\right) - \widehat{f}\left(-5+\frac{i-1}{100}\right)\right\}^2 \times 0.01.$$

The mean ISE (MISE) was approximated by the mean of ISE values derived from the 1,000 Monte Carlo replications for each DGP. The in–sample MISE of the kernel estimator of error density with its bandwidth chosen through Bayesian sampling or the two–stage CV for all five DGPs are presented in Table 5. For any DGP and any sample size considered, Bayesian sampling obviously outperforms the two–stage CV in estimating/selecting the bandwidth for the kernel estimator of the error density.

		Bayesian				Two–stage CV			
DGP	n	Mean	Median	SD	IQR	Mean	Median	SD	IQR
	100	0.0085	0.0055	0.0085	0.0071	0.0241	0.0090	0.0421	0.0219
1	200	0.0039	0.0030	0.0031	0.0031	0.0047	0.0028	0.0057	0.0041
	500	0.0015	0.0012	0.0012	0.0012	0.0018	0.0015	0.0012	0.0012
	100	0.0199	0.0107	0.0305	0.0161	0.0468	0.0181	0.1208	0.0394
2	200	0.0065	0.0040	0.0092	0.0044	0.0072	0.0046	0.0090	0.0062
	500	0.0023	0.0017	0.0020	0.0016	0.0024	0.0018	0.0019	0.0019
	1000	0.0013	0.0011	0.0010	0.0009	0.0018	0.0013	0.0056	0.0012
	50	0.0101	0.0068	0.0096	0.0080	0.0156	0.0101	0.0262	0.0111
3	100	0.0056	0.0040	0.0051	0.0041	0.0082	0.0064	0.0081	0.0059
	200	0.0030	0.0025	0.0022	0.0021	0.0048	0.0041	0.0028	0.0032
	100	0.0099	0.0056	0.0295	0.0073	0.0165	0.0067	0.0347	0.0107
4	200	0.0041	0.0029	0.0037	0.0031	0.0047	0.0028	0.0059	0.0032
	500	0.0020	0.0014	0.0071	0.0011	0.0033	0.0013	0.0343	0.0011
	100	0.0102	0.0059	0.0161	0.0074	0.0220	0.0086	0.0449	0.0163
5	200	0.0043	0.0030	0.0058	0.0033	0.0062	0.0036	0.0086	0.0047
	500	0.0017	0.0014	0.0014	0.0012	0.0019	0.0014	0.0017	0.0014

Table 5: *Average and variation measures of the in–sample ISE values derived through 1,000 repetitions for all five DGPs, where the bandwidths of the kernel error density estimator were chosen through Bayesian sampling and two–stage CV. The blue colored numbers represent the minimum MISE.*

With all the bandwidths chosen based on in–sample observations for each DGP, we calculated the out–of–sample MISE of the kernel estimator of the error density. All the out–of–sample MISE values were tabulated in Table 6. We found that for the second and fifth DGPs, Bayesian sampling slightly outperforms two–stage CV, and for the other three DGPs, results obtained from Bayesian sampling is comparable to those obtained from two–stage CV. Out of all 15 cases of different DGPs and different sample sizes, Bayesian sampling performs better than the two–stage CV for 9 cases, while the latter performs better than the former in 2 cases. Both perform similarly in 4 cases.

DGP	n	Bayesian Mean	Bayesian Median	Bayesian SD	Bayesian IQR	Two–stage CV Mean	Two–stage CV Median	Two–stage CV SD	Two–stage CV IQR
1	100	0.0013	0.0009	0.0013	0.0013	0.0012	0.0008	0.0013	0.0011
	200	0.0007	0.0005	0.0006	0.0006	0.0007	0.0005	0.0007	0.0006
	500	0.0004	0.0003	0.0003	0.0004	0.0006	0.0005	0.0004	0.0005
2	100	0.0011	0.0008	0.0009	0.0009	0.0012	0.0009	0.0011	0.0010
	200	0.0006	0.0005	0.0004	0.0005	0.0007	0.0006	0.0004	0.0005
	500	0.0003	0.0003	0.0002	0.0002	0.0004	0.0004	0.0002	0.0003
	1000	0.0002	0.0002	0.0001	0.0002	0.0003	0.0003	0.0002	0.0002
3	50	0.0040	0.0030	0.0032	0.0030	0.0038	0.0029	0.0041	0.0028
	100	0.0023	0.0020	0.0016	0.0019	0.0023	0.0019	0.0018	0.0018
	200	0.0014	0.0012	0.0009	0.0010	0.0014	0.0012	0.0010	0.0009
4	100	0.0021	0.0018	0.0015	0.0018	0.0023	0.0017	0.0021	0.0017
	200	0.0012	0.0010	0.0009	0.0010	0.0012	0.0010	0.0009	0.0010
	500	0.0006	0.0005	0.0004	0.0005	0.0006	0.0005	0.0004	0.0005
5	100	0.0017	0.0012	0.0021	0.0013	0.0021	0.0018	0.0015	0.0016
	200	0.0008	0.0007	0.0007	0.0007	0.0013	0.0012	0.0009	0.0010
	500	0.0004	0.0003	0.0003	0.0003	0.0008	0.0008	0.0004	0.0005

Table 6: *Average and variation measures of the out–of–sample ISE values derived through 1,000 repetitions for all five DGPs, where the bandwidths of the kernel error density estimator were chosen through Bayesian sampling and two–stage CV. The blue colored numbers represent the minimum MISE.*

To conclude the Monte Carlo simulation study presented in this section, we have found that for all five DGPs considered, the proposed Bayesian sampling approach outperforms its counterpart, two–stage CV, in choosing bandwidths for the NW estimator of the regression function and the kernel estimator of error density .

5 An application to modeling stock returns

The purpose of this empirical study is to demonstrate the motivation for our study, as well as the benefit of the proposed sampling algorithm for bandwidth estimation in comparison with the existing bandwidth selection method. We are interested in modeling the daily return of the

All Ordinaries (Aord) index in the Australian stock market, where one explanatory variable is the overnight daily return of the S&P 500 index because from the beginning of 2007 onwards, the U.S. has had a leading effect on other markets worldwide. Such a nonparametric regression model was previously studied by Zhang et al. (2011) to demonstrate their sampling algorithm for bandwidth estimation.

Although many market analysts observed the phenomenon that the Australian stock market generally followed the overnight market movement in the U.S., there were some exceptions that the Australian market moved along an opposite direction. This motivated us to look for another explanatory variable, and one such variable is an indicator of a major stock market in the European zone. The indicator was expected to suggest the market movement in Australia because the U.S. stock market was also supposed to have a leading effect on European stock markets. Consequently, we model the Aord daily return as an unknown function of the overnight S&P 500 return and a binary variable indicating whether the overnight FTSE index went up or down. This nonparametric regression model should better reveal the actual relationship between the Australian stock market and the U.S. market than the model investigated by Zhang et al. (2011).

5.1 Data

We downloaded daily closing index values of the Aord, S&P 500 and FTSE between the 3rd January 2007 and the 1st October 2012 from `Yahoo.com`. Each value of the Aord index was matched to the corresponding overnight values of the S&P 500 and FTSE indices. Whenever one market experienced a non–trading day, the trading data collected from all three markets on that day were excluded. The sample contains $n = 1,373$ observations of daily continuously compounded return of each index.

We fitted the nonparametric regression model given by

$$y_i = m(x_{1,i}, x_{2,i}) + \varepsilon_i, \qquad \text{for } i = 1, 2, \cdots, n, \tag{11}$$

to the sample data, where y_i is the Aord daily return, $x_{i,1}$ is the S&P 500 daily return, $x_{i,2}$ is a binary regressor taking the value of one if the FTSE daily return is positive and zero otherwise, and $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are assumed to be iid and follow a distribution with its density given by (6).

5.2 Results

The proposed Bayesian sampling algorithm was employed to estimate bandwidths in the NW estimator and the kernel–form error density. The first row panel of Table 7 presents the estimates of bandwidths, their 95% Bayesian credible intervals and associated SIF values. These small SIF values indicate that the sampler achieved very good mixing. In our experience, a SIF value of no more than 100 usually indicates reasonable mixing.

Parameter	Mean	95% Bayesian credible interval	Standard deviation	Batch–mean standard error	SIF
Bayesian (binary)					
$h_{\text{S\&P 500}}$	0.6816	(0.5850, 0.7887)	0.0016	0.0533	9.1
λ_{FTSE}	0.2630	(0.1221, 0.4076)	0.0033	0.0724	20.8
b	0.4932	(0.4375, 0.5517)	0.0008	0.0297	6.5
CV					
$h_{\text{S\&P 500}}$	0.3354				
λ_{FTSE}	0.8249				
b	0.3181				
Bayesian (continuous)					
$h_{\text{S\&P 500}}$	0.6174	(0.5265, 0.7163)	0.0033	0.0510	42.4
h_{FTSE}	0.6869	(0.6421, 0.7341)	0.0016	0.0252	42.0
b	0.2391	(0.1086, 0.3650)	0.0037	0.0703	28.2

Table 7: *Estimated bandwidths for the regressors and the kernel–form error density. The words "binary" and "continuous" in the parentheses refer to whether the second regressor is binary or continuous.*

The second row panel of Table 7 presents the bandwidths selected through two–stage CV. The bandwidths for the continuous regressor and the kernel–form error density derived through two–stage CV are clearly different from those derived through Bayesian sampling.

For comparison purpose, we also estimated bandwidths for the nonparametric regression model given by (11) with its binary regressor replaced by the FTSE daily return using the sampling algorithm provided by Zhang and King (2011). The estimates of bandwidths and their associated

95% Bayesian credible intervals and SIFs are presented in the third row panel of Table 7.

With the first pair of out–of–sample observations of S&P 500 and FTSE returns, we are able to make the one–day–ahead forecast of the Aord return before the Australian stock market opened its trading. In our sample, we collected a pair of observations of the S&P 500 and FTSE returns on the 1st October 2012 (local time), and then used the nonparametric regression model given in (11) to make a point forecast of the Aord return on the 2nd October 2012. Such a point forecast was made at each iteration of the MCMC sampling procedure. Upon completing the sampling, we derived a posterior point forecast of the Aord return by averaging those forecasts made during all iterations. The 95% Bayesian credible interval of the forecasted Aord return was also obtained. These results are presented in the first two rows of Table 8.

Forecast	Bayesian (binary)	Bayesian (continuous)	CV
Point forecast	0.1984%	0.2938%	0.0416%
95% credible interval	(0.1553%, 0.2382%)	(0.2873%, 0.2999%)	—
95% VaR	\$1.4861	\$1.2672	\$1.6048
99% VaR	\$2.5176	\$2.2113	\$2.5523

Table 8: *One–day–ahead forecast of the Aord percentage return and its VaRs for a \$100 investment on the Aord. The words "binary" and "continuous" in the parentheses refer to whether the second regressor is binary or continuous. The actual return on the 2nd October 2012 is 0.9842%.*

Apart from the one–day–ahead point forecast of the Aord return, we are able to make the one–day–ahead density forecast of the Aord return by plugging–in the sampled bandwidth into the kernel–form error density at each iteration during the sampling procedure. The forecasted density function was calculated at 25,000 grid points and saved in a row of a matrix with 25,000 columns. Upon finishing the sampling procedure, we computed the average of these density functions forecasted at all iterations. Figures 1 and 2 shows the forecasted cumulative density function (CDF) of the Aord daily return.

At the 95% and 99% confidence levels, the one–day–ahead value–at–risks (VaRs) were also computed through the above–derived CDF. According to the nonparametric regression model given by (11) with bandwidths estimated through Bayesian sampling, the VaRs for a \$100 investment on the Aord index are respectively, \$1.4861 and \$2.5176 at the 95% and 99% confidence levels.

With bandwidths selected through the two–stage CV for the nonparametric regression model given by (11), we also derived the one–day–ahead point forecast of the Aord return shown in the second column of Table 8, as well as the one–day–ahead density forecast of the Aord return with its CDF graph shown in Figure 2. According to the nonparametric regression model given by (11) with bandwidths selected through the two–stage CV, the 95% and 99% VaRs for every \$100 investment on the Aord index are respectively, \$1.6048 and \$2.5523, which are larger than the corresponding VaRs derived through Bayesian sampling.

Under the alternative nonparametric regression model with its second regressor being the FTSE return, we derived the one–day–ahead point forecast of the Aord return shown in the second last column of Table 8, the one–day–ahead density forecast of the Aord return with its CDF plotted in Figure 2. The 95% and 99% VaRs for every \$100 investment on the Aord index are respectively, \$1.2672 and \$2.2173, which are smaller than the corresponding VaRs derived through Bayesian sampling, but slightly larger than those derived through the two–stage CV.

It seems that when the second regressor is continuous, the resulting VaR is over–estimated in comparison to the VaRs derived when the second regressor is binary. Moreover, even if the binary rather than the continuous regressor is included in the nonparametric regression model, the two–stage CV leads to a clearly under–estimated VaR in comparison to its counterpart, the Bayesian sampling method for bandwidth estimation. However, the above results were derived based on forecasted densities of one day's Aord return only. To further justify the empirical importance of our method, we checked the relative frequency of exceedance through rolling samples.

5.3 Relative frequency of exceedance

As the nonparametric regression model with both continuous and binary regressors outperforms the same model with two continuous regressors, we only considered the former model for backtesting purposes. The performance of the one–day–ahead forecasted VaR was examined by the relative frequency of exceedance derived through rolling samples. The concept of exceedance refers to the fact that the actual daily loss exceeds the estimated daily VaR during the same period of holding the invested asset. The relative frequency of exceedance is a measure of the accuracy of a VaR estimate. Let α denote the confidence level for computing VaRs. If the relative frequency of exceedance is close to $(1-\alpha)$, the underlying method for computing the VaR can be regarded as appropriate. The closer the relative frequency of exceedance is to $(1-\alpha)$, the better the underling VaR estimation method would be.

In order to calculate the relative frequency of exceedance, the samples have a fixed size of 1,000. During the whole sample period, the first sample contains the first 1,000 observed vectors of the Aord, S&P 500 and FTSE returns, based on which we estimated bandwidths through Bayesian sampling and computed the VaRs at the 95% and 99% confidence levels. The second sample was derived by rolling the first sample forward for one day. Based on the second sample, we did the same thing as we did for the previous sample. This procedure continued until the second last observation was included in the sample for estimating bandwidths. There are a total of 373 samples for forecasting the one–day–ahead VaRs.

With the daily VaRs forecasted through rolling samples, we calculated the relative frequency of exceedance at different α values with bandwidths chosen through Bayesian sampling and the two–stage CV. With Bayesian sampling, the resulted relative frequencies are respectively, 0.80% and 4.81% at the 99% and 95% confidence levels. However, with two–stage CV method, the corresponding relative frequencies of exceedance are 1.07% and 6.95%, respectively. It shows that two–stage CV for bandwidth selection leads to under–estimated VaRs in comparison to our proposed sampling method.

6 Conditional density estimation of GDP growth rates

The sampling algorithm proposed in Section 3 can be modified for the purpose of choosing bandwidths in kernel density estimation of continuous and discrete variables. Maasoumi, Racine, and Stengos (2007) investigated kernel density estimation of the gross domestic product (GDP) growth rate among OECD and non–OECD countries across different years, where the OECD status of the country and the year of the observed growth rate are included in the data matrix. Consequently, they aimed to estimate the trivariate density of growth rates (denoted as $x^{(c)}$), OECD status and year (denoted as $\boldsymbol{x}^{(d)}$), where the last two variables are respectively, binary and ordered categorical. Their primary purpose was to explore the dynamic evolution of OECD and non–OECD countries' distributions of GDP growth rates across different years. Maasoumi et al. (2007) proposed using the kernel estimator with unordered and ordered discrete kernels assigned to the discrete variable to estimate such trivariate densities, where bandwidths were selected through the likelihood cross–validation (LCV) (see also Hall, 1987; Maasoumi et al., 2007).

We are interested in choosing bandwidths for the kernel estimator of $\boldsymbol{x} = \left(x^{(c)}, \boldsymbol{x}^{(d)}\right)'$, which is expressed as

$$\widehat{f}\left(x^{(c)}, \boldsymbol{x}^{(d)}; h, \boldsymbol{\lambda}\right) = \frac{1}{n}\sum_{j=1}^{n} K_h^{(c)}\left(x^{(c)} - x_j^{(c)}\right) \times K_{\boldsymbol{\lambda}}^{(d)}\left(\boldsymbol{x}^{(d)}, \boldsymbol{x}_j^{(d)}\right).$$

where $x_i^{(c)}$ and $\boldsymbol{x}_i^{(d)}$, for $i = 1, 2, \cdots, n$, are observations of $x^{(c)}$ and $\boldsymbol{x}^{(d)}$, respectively. The kernel function for GDP growth rates is the Gaussian kernel, while the OECD status is assigned with a kernel function given by (3), and the kernel for years is given by (4).

The likelihood of $\boldsymbol{x} = \left\{\left(x_i^{(c)}, \boldsymbol{x}_i^{(d)}\right) : i = 1, 2, \cdots, n\right\}$ for given $(h, \boldsymbol{\lambda})$ is approximately

$$\ell\left(\boldsymbol{x} \middle| h, \boldsymbol{\lambda}\right) \approx \prod_{i=1}^{n} \widehat{f}_{(-i)}\left(x_i^{(c)}, \boldsymbol{x}_i^{(d)}; h, \boldsymbol{\lambda}\right),$$

where $\widehat{f}_{(-i)}$ is the leave–one–out version of $\widehat{f}$ (see for example, Zhang, King, and Hyndman, 2006).

The priors of bandwidth parameters are the same as those discussed in Section 3.2. The posterior of $(h, \boldsymbol{\lambda})$ is proportional to the product of $\ell\left(\boldsymbol{x}\middle|h,\boldsymbol{\lambda}\right)$ and the priors of bandwidths. The random–walk Metropolis algorithm was used to implement the posterior simulation, where the proposal density is the standard trivariate Gaussian, and the tuning parameter was chosen to make the acceptance rate around 0.234. The posterior mean of each simulated chain is used as an estimate of the corresponding bandwidth.

With the estimated bandwidths, we derived the joint density of GDP growth rate, OECD status and years. The conditional density of GDP growth rates for given values of OECD status and year is the joint density estimator divided by the marginal density estimator of OECD status and year. Note that bandwidths estimated for the joint trivariate kernel density estimator can be used for the kernel conditional density estimator.

Given different values of OECD status and year, we calculated the conditional densities and CDFs of GDP growth rates with their graphs presented in Figures 3 and 4. It shows that Bayesian sampling and LCV lead to clearly different density functions of growth rates. The growth–rate distributions of an OECD country and a non–OECD country are very different from 1965 to 1995. Second, the growth–rate density of an OECD country is almost symmetrical and less dispersed than that of a non–OECD country, and this phenomenon becomes obvious over time. The growth–rate density of a non–OECD country is asymmetrical and has larger variation than that of an OECD country. It appears to manifest bimodality and indicate "polarization" within non–OECD countries. The results re–confirm the findings of Maasoumi et al. (2007). Figure 5 presents the stacked plots of OECD and non–OECD density and distribution functions of growth rates for all years, where bandwidths were estimated through Bayesian sampling.

We found the following empirical evidence. First, given the year value at either 1965 or 1970, the conditional growth–rate distribution of an OECD country (purple dotted line on the two top right graphs of Figure 3) stochastically dominates that of a non–OECD country (blue solid line). However, there has been no such stochastic dominance since 1975. Second, given the OECD status of a country, the country's growth–rate distribution in 1965 (black solid line on the top right graph of Figure 5) stochastically dominates its growth–rate distributions in the other years; and its growth–rate distribution in 1990 (pink long– and short–dashed line) stochastically dominates its growth–rate distributions in 1980 (blue dot–dash line), 1985 (brown dashed line) and 1995 (purple solid line), respectively.

7 Conclusion

We have presented a Bayesian sampling approach to the estimation of bandwidths in a nonparametric regression model with continuous and discrete regressors, where the regression function is estimated by the NW estimator and the unknown error density is approximated a kernel–form error density. Monte Carlo simulation results show that the proposed Bayesian sampling method performs better than, or at least on par in only a small proportion of occasions with, the cross–validation for choosing bandwidths for the NW estimator. The advantage of the proposed Bayesian approach over the cross–validation is its ability to estimate the error density. As measured by the MISE, the Bayesian method outperforms the two–stage CV approach for estimating the bandwidth in the kernel–form error density. Thus, the proposed sampling method is recommended for estimating bandwidths in the regression–function and kernel–form error density estimators.

The proposed Bayesian sampling algorithm is used to estimate bandwidths for the nonparametric regression of All Ordinaries (Aord) daily return on the overnight S&P 500 return and an indicator of the FTSE return. In comparison to the cross–validation for bandwidth selection, the proposed sampling method leads to a different one–step–ahead forecasted density of the Aord return. Consequently, the resulting value–at–risk measure, as well as the relative frequency of exceedance, is different from the one derived with bandwidths selected through cross–validation. In this example, Bayesian sampling for bandwidth estimation in the nonparametric regression of mixed regressors leads to better results than cross–validation. In an application that involves of kernel density estimation of a country's GDP growth rate conditional on its OECD status and the year of observations, Bayesian sampling for bandwidth estimation leads to different density estimates from those with bandwidths selected through likelihood cross–validation.

Acknowledgements

The first author extends sincere thanks to Jeffrey Racine for insightful discussion during an early stage of this paper. This research is supported under Australian Research Council's *Discovery Projects* funding scheme (project number DP1095838).

References

Ahmad, I. A., Li, Q., 1997. Testing symmetry of an unknown density function by kernel method. Journal of Nonparametric Statistics 7 (3), 279–293.

Aitchison, J., Aitken, C. G. G., 1976. Multivariate binary discrimination by the kernel method. Biometrika 63 (3), 413–420.

Akritas, M. G., Van Keilegom, I., 2001. Non-parameteric estimation of the residual distribution. Scandinavian Journal of Statistics 28 (3), 549–567.

Cheng, F., Sun, S., 2008. A goodness of fit test of the errors in nonlinear autoregressive time series models. Statistics & Probability Letters 78 (1), 50–59.

Dette, H., Kusi-Appiah, S., Neumeyer, N., 2002. Testing symmetry in nonparametric regression models. Journal Nonparametric Statistics 14 (5), 477–494.

Efromovich, S., 2005. Estimation of the density of regression errors. The Annals of Statistics 33 (5), 2194–2227.

Escanciano, J. C., Jacho-Chávez, D. T., 2012. $\sqrt{n}$ uniformly consistent density estimation in nonparametric regression models. Journal of Econometrics 167 (2), 305–316.

Garthwaite, P. H., Fan, Y., Sisson, S. A., 2010. Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. Working paper, University of New South Wales, Sydney.
URL http://arxiv.org/pdf/1006.3690v1.pdf

Geweke, J., 2009. Complete and Incomplete Econometric Models. Princeton University Press, New Jersey.

Hall, P., 1987. On Kullback-Leibler loss and density estimation. The Annals of Statistics 15 (4), 1491–1519.

Hall, P., Li, Q., Racine, J., 2007. Nonparametric estimation of regression functions in the presence of irrelevant regressors. The Review of Economics and Statistics 89 (4), 784–789.

Hayfield, T., Racine, J. S., 2008. Nonparametric econometrics: The np package. Journal of Statistical Software 27 (5), 1–32.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65 (3), 361–393.

Li, C., Ouyang, D., Racine, J. S., 2009. Nonparametric regression with weakly dependent data: The discrete and continuous regressor case. Journal of Nonparametric Statistics 21 (6), 697–711.

Li, Q., Racine, J. S., 2003. Nonparametric estimation of distributions with categorical and continuous data. Journal of Multivariate Analysis 86 (2), 266–292.

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton.

Li, Q., Zhou, J., 2005. The uniqueness of cross-validation selected smoothing parameters in kernel estimation of nonparametric models. Econometric Theory 21 (5), 1017–1025.

Loynes, R. M., 1980. The empirical distribution function of residuals from generalised regression. The Annals of Statistics 8 (2), 285–298.

Ma, S., Racine, J. S., 2013. Additive regression splines with irrelevant categorical and continuous regressors. Statistica Sinica 23 (2), 515–541.

Maasoumi, E., Racine, J., Stengos, T., 2007. Growth and convergence: A profile of distribution dynamics and mobility. Journal of Econometrics 136 (2), 483–508.

Meyer, R., Yu, J., 2000. BUGS for a Bayesian analysis of stochastic volatility models. Econometrics Journal 3 (2), 198–215.

Muhsal, B., Neumeyer, N., 2010. A note on residual-based empirical likelihood kernel density estimation. Electronic Journal of Statistics 4, 1386–1401.

Nadaraya, E. A., 1964. On estimating regression. Theory of Probability and its Applications 9 (1), 141–142.

Neumeyer, N., Dette, H., 2007. Testing for symmetric error distribution in nonparametric regression models. Statistica Sinica 17 (2), 775–795.

Racine, J., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data. Journal of Econometrics 119 (1), 99–130.

Roberts, G. O., Rosenthal, J. S., 2009. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics 18 (2), 349–367.

Samb, R., 2011. Nonparametric estimation of the density of regression errors. Comptes Rendus Mathematique 349 (23–24), 1281–1285.

Su, L., Chen, Y., Ullah, A., 2009. Functional coefficient estimation with both categorical and continuous data. In: Li, Q., Racine, J. (Eds.), Advances in Econometrics. Vol. 25. Emerald Group Publishing Limited, pp. 131–167.

Tse, Y. K., Zhang, X., Yu, J., 2004. Estimation of hyperbolic diffusion using the Markov chain Monte Carlo method. Quantitative Finance 4 (2), 158–169.

Watson, G., 1964. Smooth regression analysis. Sankhya: The Indian Journal of Statistics, Series A 26 (4), 359–372.

Yuan, A., de Gooijer, J. G., 2007. Semiparametric regression with kernel error model. Scandinavian Journal of Statistics 34 (4), 841–869.

Zhang, X., King, M. L., 2011. Bayesian semiparametric GARCH models. Working paper 11/24, Monash University.
URL http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp24-11.pdf

Zhang, X., King, M. L., Hyndman, R. J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Computational Statistics and Data Analysis 50 (11), 3009–3031.

Zhang, X., King, M. L., Shang, H. L., 2011. Bayesian estimation of bandwidths for a nonparametric regression model with a flexible error density. Working paper 11/10, Department of Econometrics & Business Statistics, Monash University.
URL http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp10-11.pdf

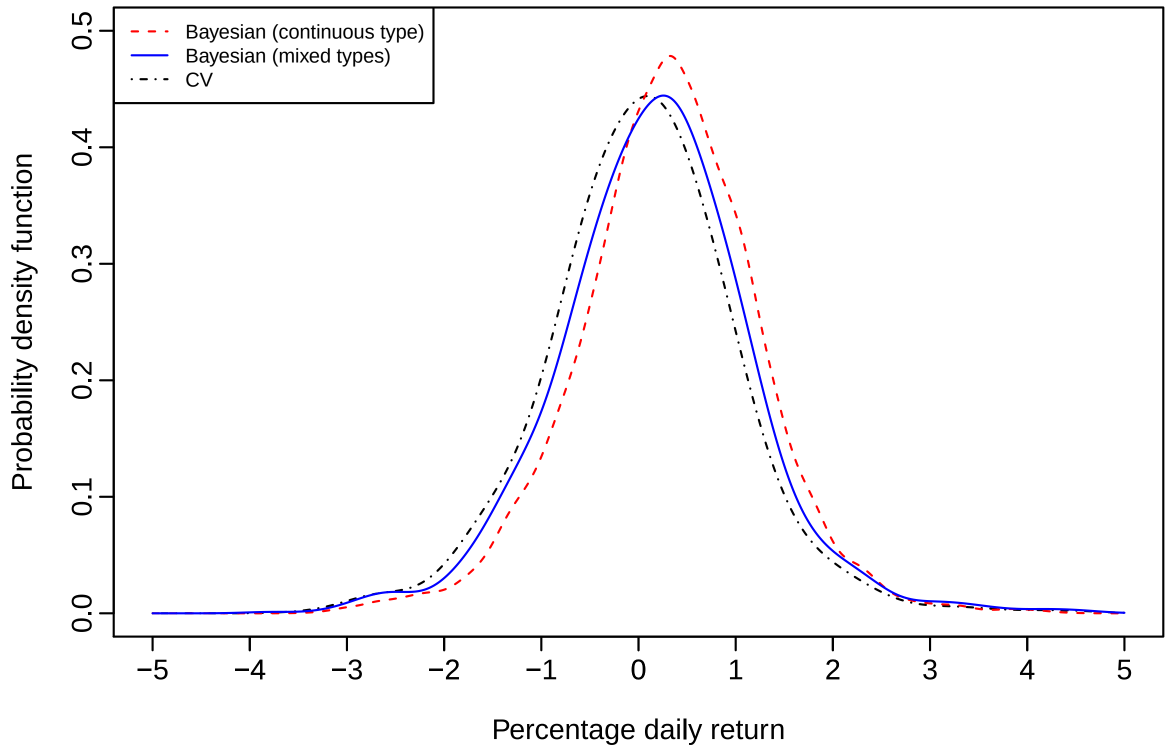

Figure 1: *Graphs of PDFs of the forecasted All Ordinaries return on the 2nd October 2012.*

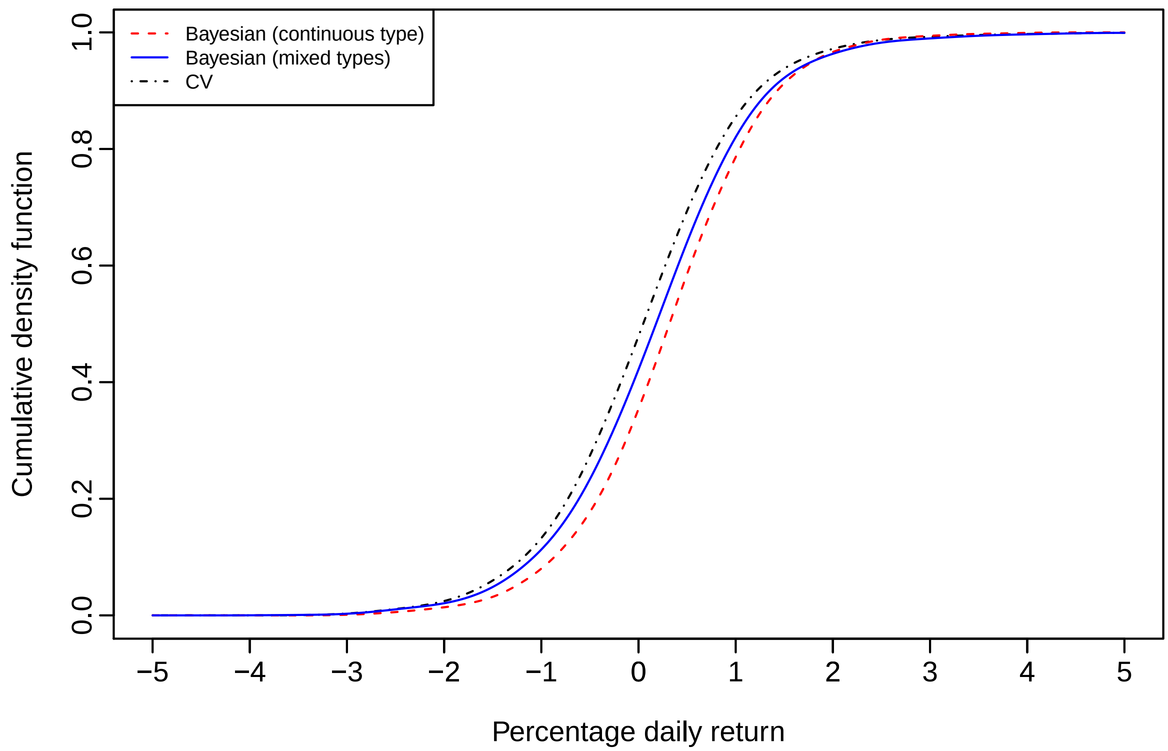

Figure 2: *Graphs of CDFs of the forecasted All Ordinaries return on the 2nd October 2012.*

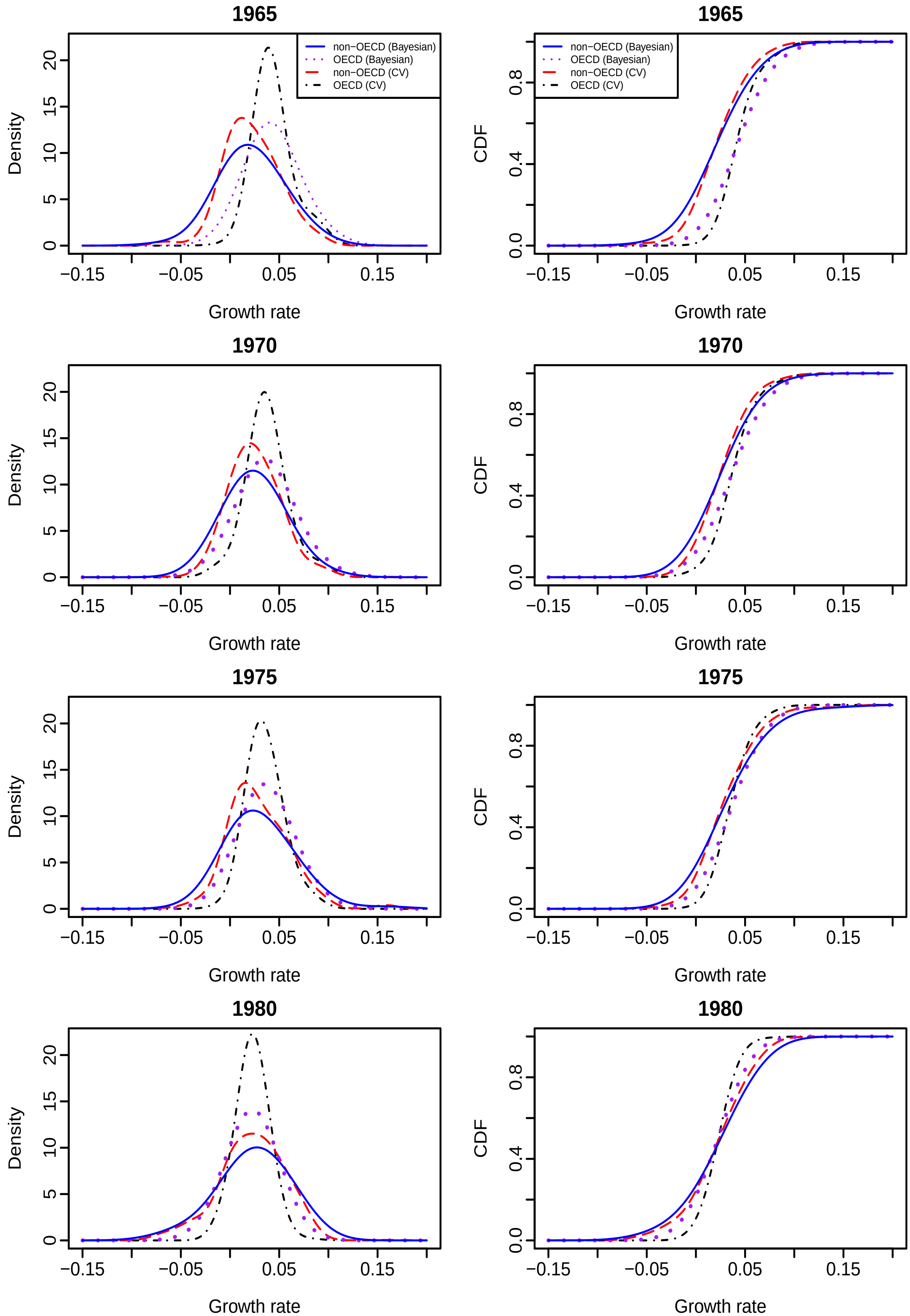

Figure 3: *Density and distribution functions of GDP growth rate by year and OECD status.*

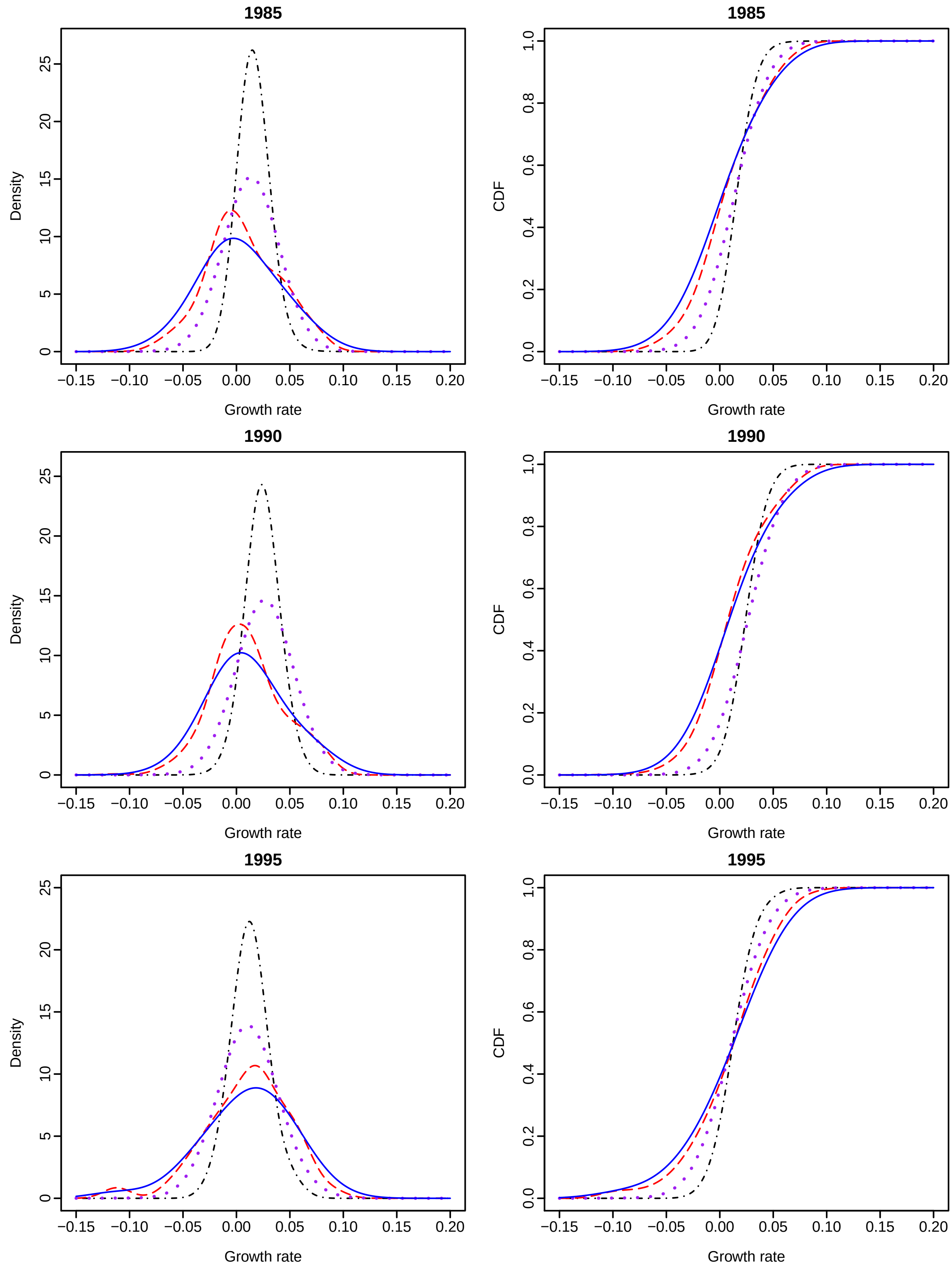

Figure 4: *Density and distribution functions of GDP growth rate by year and OECD status (continued).*

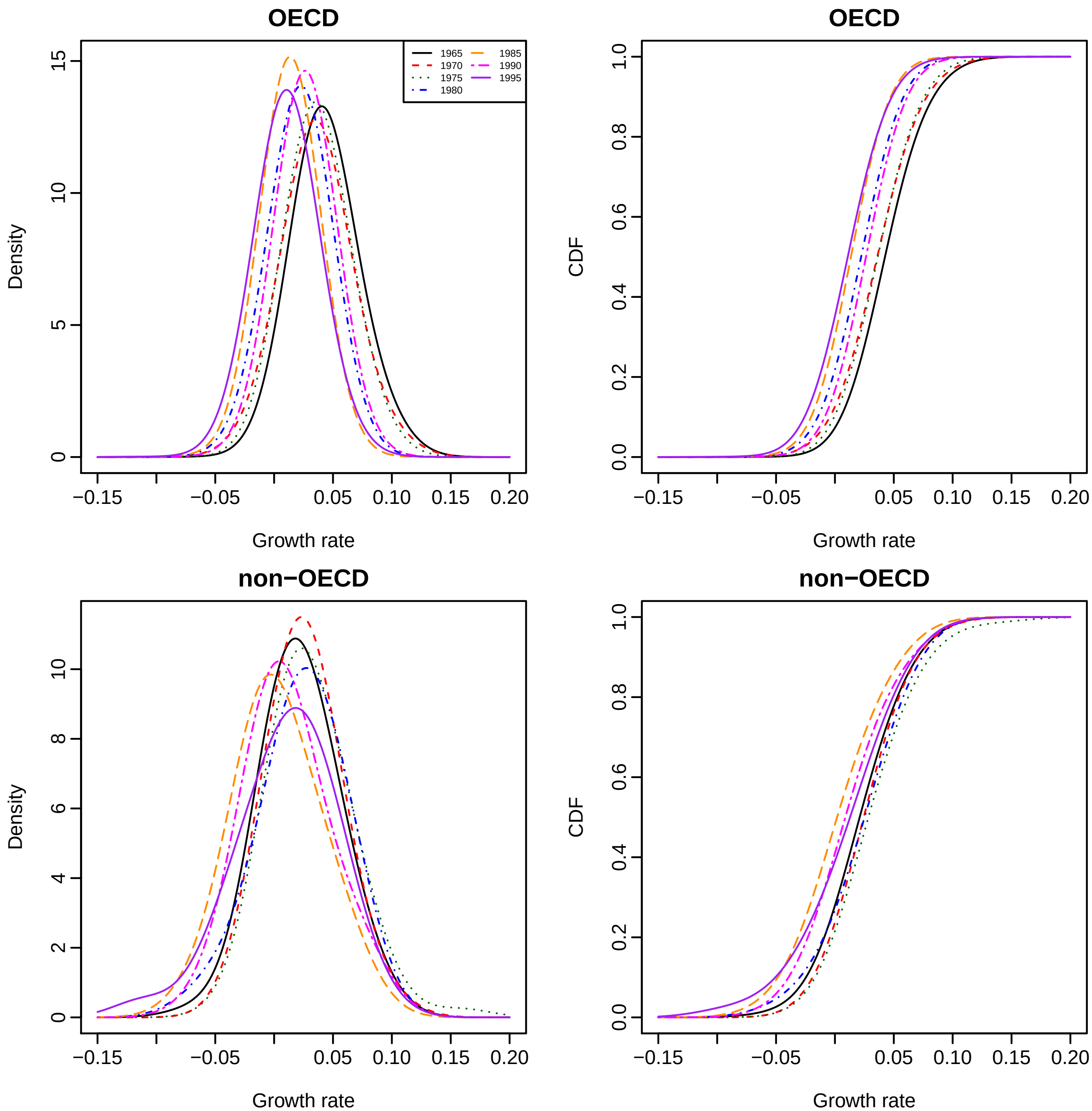

Figure 5: *Stacked plots of density and distribution functions of GDP growth rate.*