Nonparametric Method: Take-Home Exam 1

Mai-Anh Dang | Student ID: 21608631 | TSE M2 EEE 2017-18

September 29, 2017

1 Exercise 1

The derivative of density function, $f(x)$, is denoted as $f'(x) = \dfrac{df(x)}{dx}$. The estimator of $f(x)$ is given by:

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)$$

1.1 Suggest the Kernel Estimator for f'(x)

If $\hat{f}(x)$ is the kernel estimator of $f(x)$, then it follows that $\hat{f}'(x)$ is the kernel estimator of $f'(x)$. Assume that the Kernel K is differentiable at the first order, denoted as $K'\left(\frac{x-X_i}{h_n}\right)$.

Get the first derivative of $\hat{f}(x)$ with respect to x:

$$\begin{aligned}
\hat{f}'(x) &= \frac{d}{dx}\left[\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)\right] \\
&= \frac{1}{nh_n}\sum_{i=1}^{n} K'\left(\frac{x-X_i}{h_n}\right)\left(\frac{x-X_i}{h_n}\right)' \\
&= \frac{1}{nh_n^2}\sum_{i=1}^{n} K'\left(\frac{x-X_i}{h_n}\right)
\end{aligned}$$

1.2 Bias and Variance of $\hat{\mathbf{f}}'(\mathbf{x})$

It is assumed that K is ν–order kernel ($\nu > 2$), mean:

1. $\kappa_0(K) = \int K(u)du = 1$,
2. $\kappa_j(K) = \int u^j K(u)du = 0$, when $1 \leq j < \nu$
3. $\kappa_\nu(K) = \int u^\nu K(u)du < \infty$

In additional, we assume that the sample $\{X_i\}_{i=1}^n$ is i.i.d. Also, assuming that f is ν times differentiable in each argument.

Bias of $\hat{\mathbf{f}}'(\mathbf{x})$:

$$\begin{aligned}
E(\hat{f}'(x)) &= E\left\{\frac{d}{dx}\left[\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)\right]\right\} \\
&= \frac{d}{dx}E\left[\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)\right] \\
&= \frac{d}{dx}\left\{\frac{1}{nh_n}\sum_{i=1}^{n}E\left[K\left(\frac{x-X_i}{h_n}\right)\right]\right\} \\
&= \frac{d}{dx}\left\{\frac{1}{nh_n}nE\left[K\left(\frac{x-X_i}{h_n}\right)\right]\right\} \\
&= \frac{d}{dx}\left\{\frac{1}{h_n}E\left[K\left(\frac{x-X_i}{h_n}\right)\right]\right\} \\
&= \frac{d}{dx}\left[\frac{1}{h_n}\int_{-\infty}^{+\infty}K\left(\frac{x-z}{h_n}\right)f(z)dz\right]
\end{aligned}$$

The third and fourth equality, respectively, follow the assumption about the independent and identical distribution of $\{X_i\}_{i=1}^{n}$. Next, take $u=\frac{x-z}{h_n}$. Then, $z=x-h_nu$ and $dz=-h_ndu$. Also, $u=\pm\infty$ when $z=\mp\infty$

$$\begin{aligned}
E[\hat{f}'(x)] &= \frac{d}{dx}\left[\frac{1}{h_n}\int_{+\infty}^{-\infty}K(u)f(x-h_nu)du(-h_n)\right] \\
&= \frac{d}{dx}\int_{-\infty}^{+\infty}K(u)f(x-h_nu)du
\end{aligned}$$

Substitute $f(x-h_nu)$ by its Taylor Series Expansion of $f(.)$ around x:

$$\begin{aligned}
f(x-h_nu) &= f(x)+h_nuf'(x)+\frac{h_n^2u^2}{2}f''(x)+\frac{h_n^3u^3}{3!}f^{(3)}(x)+...+\frac{h_n^\nu u^\nu}{\nu!}f^{(\nu)}(x)+o(h_n^\nu) \\
&= f(x)+\sum_{j=1}^{\nu-1}\frac{h_n^ju^j}{j!}f^{(j)}(x)+\frac{h_n^\nu u^\nu}{\nu!}f^{(\nu)}(x)+o(h_n^\nu)
\end{aligned}$$

For $f^{(j)}(x)$ is the j-th order derivative of $f(x)$.
After the substitution, we have:

$$\begin{aligned}
E[\hat{f}'(x)] &= \frac{d}{dx}\left[f(x)\int_{-\infty}^{+\infty}K(u)du+\sum_{j=1}^{\nu-1}\frac{h_n^j}{j!}f^{(j)}(x)\int_{-\infty}^{+\infty}u^jK(u)du+\frac{h_n^\nu}{\nu!}f^{(\nu)}(x)\int_{-\infty}^{+\infty}u^\nu K(u)du+o(h_n^\nu)\right] \\
&= \frac{d}{dx}\left[f(x)\kappa_0(K)+\sum_{j=1}^{\nu-1}\frac{h_n^j}{j!}f^{(j)}(x)\kappa_j(K)+\frac{h_n^\nu}{\nu!}f^{(\nu)}(x)\kappa_\nu(K)+o(h_n^\nu)\right] \\
&= \frac{d}{dx}\left[f(x)+\frac{h_n^\nu}{\nu!}f^{(\nu)}(x)\kappa_\nu(K)+o(h_n^\nu)\right]
\end{aligned}$$

By the assumption about K ν–order kernel ($\nu>2$), we have $\kappa_0(K)=1$, $\kappa_\nu(K)=c<\infty$. Also, we obtain the last equality as $\sum_{j=1}^{\nu-1}\frac{h_n^j}{j!}f^{(j)}\kappa_j(K)=0$, when $\kappa_j(K)=0$, for all $1\leq j<\nu$. Next, taking the derivative with respect to x, each $f^{(j)}(x)$ terms increase by one order of derivative. The equation becomes:

$$\begin{aligned}
E[\hat{f}'(x)] &= f'(x)+\frac{h_n^\nu}{\nu!}f^{(\nu+1)}(x)\kappa_\nu(K)+o(h_n^\nu) \\
&= f'(x)+Bias[\hat{f}'(x)]
\end{aligned}$$

where:

$$\begin{aligned} Bias[\hat{f}'(x)] &= \frac{h_n^{\nu}}{\nu!} f^{(\nu+1)}(x)\kappa_\nu(K) + o(h_n^\nu) \\ &= O(h_n^\nu) \end{aligned}$$

Variance of $\hat{\mathbf{f}}'(\mathbf{x})$:

$$\begin{aligned} Var[\hat{f}'(x)] &= Var\left[\frac{1}{nh_n^2}\sum_{i=1}^{n} K'\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{n^2h_n^4}\sum_{i=1}^{n} Var\left[K'\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{n^2h_n^4} nVar\left[K'\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{nh_n^4} Var\left[K'\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{nh_n^4}\frac{d}{du} Var[K(u)] \end{aligned}$$

The second and third equality, respectively, follow the assumption about the independent and identical distribution of $\{X_i\}_{i=1}^n$. For the last line, $K'(u)$ is the first derivative with respect to u. Take $u = \frac{x-z}{h_n}$. Then, $z = x - h_n u$ and $dz = -h_n du$; $\frac{dx}{du} = h_n$. Also, $u = \pm\infty$ when $z = \mp\infty$.

Compute the $Var[K(u)]$:

$$Var[K(u)] = Var\left[K\left(\frac{x-X_i}{h_n}\right)\right] = \underbrace{E\left[K\left(\frac{x-X_i}{h_n}\right)^2\right]}_{(1)} - \underbrace{E\left[K\left(\frac{x-X_i}{h_n}\right)\right]^2}_{(2)}$$

The second and third equality, respectively, follow the assumption about the independent and identical distribution of $\{X_i\}_{i=1}^n$. Now, we calculate term (1) and (2).

For term (1), with similar procedure:

$$\begin{aligned} E\left[K\left(\frac{x-X_i}{h_n}\right)^2\right] &= \int_{-\infty}^{+\infty} K\left(\frac{x-z}{h_n}\right)^2 f(z)dz \\ &= \int_{+\infty}^{-\infty} K(u)^2 f(x-h_nu)(-h_n)du \\ &= \int_{-\infty}^{+\infty} K(u)^2 f(x-h_nu)h_n du \\ &= \int_{-\infty}^{+\infty} K(u)^2 \left[f(x) + \sum_{j=1}^{\nu-1}\frac{h_n^j u^j}{j!} f^{(j)}(x) + \frac{h_n^\nu u^\nu}{\nu!} f^{(\nu)}(x) + o(h_n^\nu)\right] h_n du \\ &= h_n f(x)\int_{-\infty}^{+\infty} K(u)^2 du + \sum_{j=1}^{\nu-1}\frac{h_n^{j+1}}{j!} f^{(j)}(x)\int_{-\infty}^{+\infty} u^j K(u)^2 du \\ &\quad + \frac{h_n^{\nu+1}}{\nu!} f^{(\nu)}(x)\int_{-\infty}^{+\infty} u^\nu K(u)^2 du \\ &= h_n f(x)\psi(K) + O(h_n^2) \end{aligned}$$

where $\psi(K) = \int_{-\infty}^{+\infty} K(u)^2 du$. In the last equation, for the remaining terms, the lower order of h_n dominates.

For term (2),

$$
\begin{aligned}
E\left[K\left(\frac{x-X_i}{h_n}\right)\right]^2 &= \left[\int_{-\infty}^{+\infty} K(u)f(x-h_nu)h_ndu\right]^2 \\
&= \left[h_n\int_{-\infty}^{+\infty} K(u)f(x-h_nu)du\right]^2 \\
&= \left\{h_n\left[f(x)+\frac{h_n^{\nu}}{\nu!}f^{(\nu)}(x)\kappa_{\nu}(K)+o(h_n^{\nu})\right]\right\}^2 \\
&= \left[h_nf(x)+\frac{h_n^{\nu+1}}{\nu!}f^{(\nu)}(x)\kappa_{\nu}(K)+o(h_n^{\nu+1})\right]^2 \\
&= \left[h_nf(x)+O(h_n^{\nu+1})\right]^2 \\
&= [O(h_n)]^2
\end{aligned}
$$

Put all terms into the original equation, we have:

$$
\begin{aligned}
Var[\hat{f}'(x)] &= \frac{1}{nh_n^4}\frac{d}{du}Var[K(u)] \\
&= \frac{1}{nh_n^4}\frac{d}{du}\left[h_nf(x)\psi(K)+O(h_n^2)+[O(h_n)]^2\right] \\
&= \frac{1}{nh_n^3}\frac{d}{du}\left[f(x)\psi(K)+O(h_n)\right] \\
&= \frac{1}{nh_n^3}\left[\frac{d}{du}f(x)\psi(K)+f(x)\frac{d}{du}\psi(K)+O(h_n)\right] \\
&= \frac{1}{nh_n^3}\left[f'(x)\frac{dx}{du}\psi(K)+f(x)\frac{d}{du}\int_{-\infty}^{+\infty}K(u)^2+O(h_n)\right] \\
&= \frac{1}{nh_n^3}\left[f'(x)h_n\psi(K)+f(x)K(u)^2+O(h_n)\right] \\
&= \frac{1}{nh_n^2}f'(x)\psi(K)+\frac{1}{nh_n^3}f(x)K(u)^2+O\left(\frac{1}{h_n^2}\right) \\
&= O\left(\frac{1}{nh_n^3}\right)
\end{aligned}
$$

In which, we have defined before $\frac{dx}{du}=h_n$ and the lower order of h_n dominates.

For $Bias[\hat{f}'(x)=O(h_n^{\nu})$ and $Var[\hat{f}'(x)]=O(\frac{1}{nh_n^3})$, when h_n decreases, Bias decrease but Var increases. For consitent estimation, we need: $n\to\infty$, $h_n\to 0$, and $nh_n^3\to\infty$.

1.3 MSE-optimal Bandwidth

$$
\begin{aligned}
MSE[\hat{f}'(x)] &= E\left[(\hat{f}'(x)-f'(x))^2\right] \\
&= Var[\hat{f}'(x)]+\{Bias[\hat{f}'(x)]\}^2 \\
&= O\left(\frac{1}{nh_n^3}\right)+O(h_n^{2\nu})
\end{aligned}
$$

h_n^{opt} will minimize MSE and balance the effects on squared bias and variance.

$$
h_n^{opt}=\underset{h_n}{\operatorname{argmin}}MSE[\hat{f}'(x)]
$$

h_n^{opt} will satisfy:

$$\frac{dMSE[\hat{f}'(x)]}{dh_n} = 0$$
$$\implies O\left(\frac{1}{nh_n^4}\right) = O\left(h_n^{2\nu-1}\right)$$
$$\implies h_n^{2\nu-1} \propto n^{-1}h_n^{-4}$$
$$\implies h_n^{2\nu+3} \propto n^{-1}$$
$$\implies h_n^{opt} \propto n^{\frac{-1}{2\nu+3}}$$

The rate of convergence of $\hat{f}'(x)$ to $f'(x)$, which is also the rate that the bias vanishes, is of order:

$$O[(h_n^{opt})^\nu] = O(n^{\frac{-\nu}{2\nu+3}})$$

We can see that ν increase, this rate increases.

2 Exercise 2

2.1 Explore the Data and Discuss the Context

Theoretically, we could predict that there is a positive relationship between *Income* and *Food–Expenditure*, i.e. when the income increases the household is enabled to spend more in food. It is also expected that for higher-level of income this effect will be weaken (i.e. $g'(x) > 0; g''(x) < 0$).

However, contextually, the dataset is limited to the range of relatively low- and middle- income. The observations with higher income-level is rare, and being treated as the outliers in this analysis. The summary of raw data in variables $Income$ and $FoodExpenditure$ is in the table below:

Variable	Min	1st Quant.	Mean	Median	3rd Quant.	Max
Income	377.1	638.9	884.0	982.5	1164.0	4958.0
Food Expense	242.3	429.7	582.5	623.0	743.9	2033.0

Initial plot with least-squared reference in **Figure 1**.

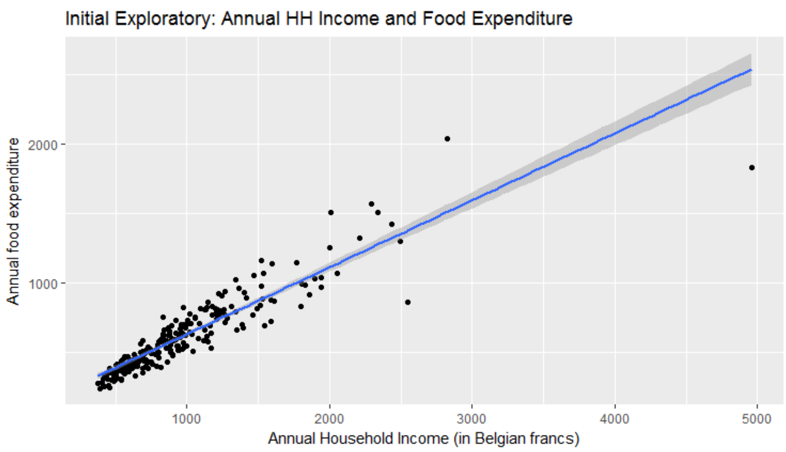


Figure 1: Initial Plot Food Expenditure conditional on Household Income

In the considered range of observations in this context, we expect the $g(x)$ **to be flat, closely to be linear**. The range of observations in both *Income* and *Food Expenditure* is strictly positive. It means that we do not have the *"boundary issue"* in this context.

Consequently, in this context, we expect *Local Constant (NW) Estimates* to perform **relatively better** than *Local Linear* (which is more preferred for curvier $g(x)$.

2.2 Compute and Plot

For the data-set *engel.dta*, with 235 observations on *Annual household income in belgian francs* (X), and *Annual household food expenditure* (Y). Assume that we have the estimate model:

$$E[Y|X = x] = g(x)$$

The applied kernel function is *second-order Epanechnikov*: $K(u) = \frac{3}{4}(1 - u^2)1_{(|u|\leq 1)}$. The *Least-Squared Cross-Validation* bandwidths is chosen, respectively for *Local Constant (Nadaraya-Watson) estimates* and *Local Linear estimates*: $h_n^{lc} = 183.016$; $h_n^{ll} = 130.91$.

The **Local Constant Estimator** is defined as:

$$\hat{g}^{lc}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}$$

Assume that $Y_i = g(X_i) + \epsilon_i \approx g(x) + (X_i - x)\beta + \epsilon_i$, the **Local Linear Estimator** is defined as:

$$\begin{bmatrix} \hat{\alpha}(x) \\ \hat{\beta}(x) \end{bmatrix} = \left(\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n} Z_i Z_i'\right)\right)^{-1} \left(\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) Z_i Y_i\right)$$

where $\hat{g}^{ll}(x) = \hat{\alpha}(x)$

In addition, the graphs of Local Linear and Local Constant Estimates would be compared with the linear mean regression (OLS). The graphs are illustrated in **Figure 2**. To have a well-behaved sample, we limit the range of consideration, for HH Income below 2,000, and HH food expenditure below 1,500 Belgian francs, without losing too many observations.

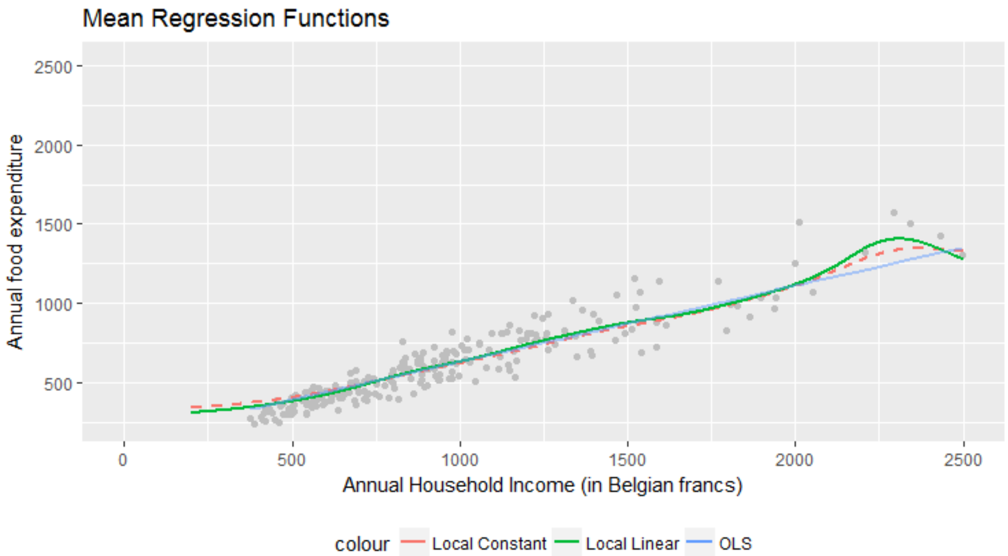


Figure 2: Mean Regression Function of Food Expenditure conditional on Household Income

2.3 Analyze the Plots and Compare to Linear Mean Regression

Discussing about the plots in **Figure 2**, as the predictions, the relationship between household Income and household Food expenditure is positive, and quite flat in the chosen range of income. Both *Local Constant estimates and Local Linear estimates* are *very close to the OLS estimates.* As the effect of smoothing kernel function, the Local Constant and Local Linear do not produce perfect linear relationship. But, the results are acceptably similar.

When the range of income go beyond the threshold of *appx.* 2,000 Belgian francs, the linear relationship is less strong (the non-linearity could be expected as the previous discussion). Food is necessity goods, that over a certain threshold, the marginal utility of Food will be declined. Thus, the higher income does not encourage further consumption and expenditure in food. Then, for higher range of income, the *Local Constant and Local Linear estimates are apart from the OLS.* However, the data above this range of income is not sufficiently available to investigate any further.

3 Exercise 3

The **Monte-Carlo Simulation** would be conducted, with **100 times** of simulations, for sample sizes $\mathbf{n = 100}$ and $\mathbf{n = 1000}$. The univariate density of X distribution and joint bivariate density distribution of Y would be estimated, using the *Gaussian kernel* $K(u) = \frac{1}{\sqrt[2]{2\pi}} exp\left(-\frac{u^2}{2}\right)$ and the *Least-Squared Cross-Validation bandwidth* to chose h. The average of chosen bandwidths in 100 monte-carlos are report in the below tables:

	Sample	h	h1 = h2 = h
1	n = 100	1.1937	1.1937
2	n = 1000	0.7327	0.7327

Notice that for the bigger sample, the bandwidth becomes smaller. It means that the bias of estimates will converge faster for the large sample. The error will be smaller, the estimates is expected be closer to the truth, which will be illustrated in the next parts.

3.1 The estimate $\hat{\mathbf{f}}(\mathbf{X})$ for $\mathbf{X} \sim \mathbf{N(4, 9)}$

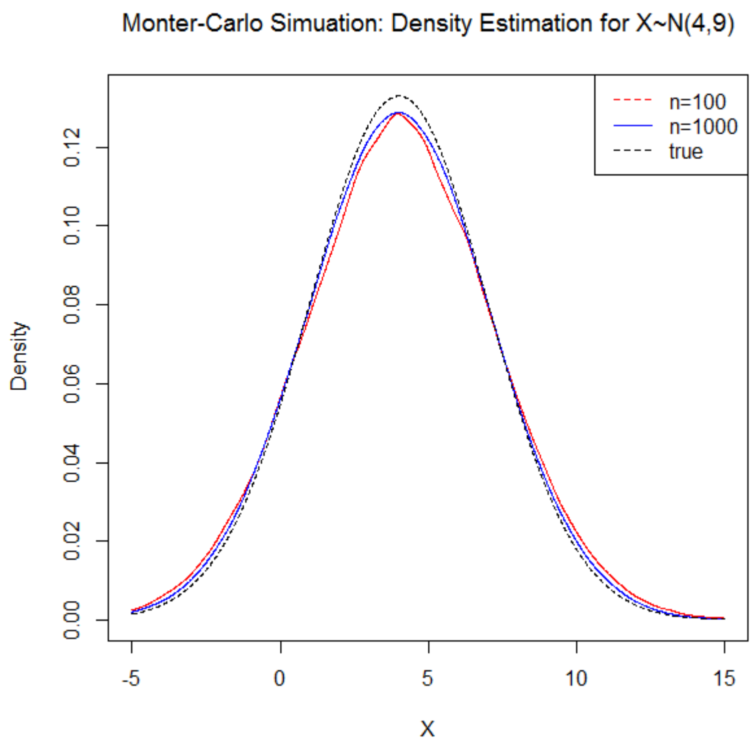


Figure 3: Monte-Carlo Simulation: Density Estimation of $X \sim N(4, 9)$

Randomly generate n draws of a random variable $X \sim N(4, 9)$. Then, the **true density function** of X is that:

$$f(x) = \frac{1}{\sqrt[2]{2\pi}\sigma_X} \exp\left(\frac{-(x-\mu_X)^2}{2\sigma_X^2}\right) \tag{1}$$

where: $\mu_X = 4$ and $\sigma_X = 3$. The chosen grid of points x is $[-5; 15]$. Putting each values in this considered range of x into *eq.(1)*, we will obtain the true curve of density function *(which is the black line in* ***Figure 3****)*.

Assuming that the data X is given, we use the nonparametric kernel estimator for the density of X:

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{2}$$

The estimates in *eq.(2)* will be consistent if $n \to \infty$, $h_n \to \infty$, $nh_n \to \infty$.

On the same range of x, estimating the density of X for both the sample of $n = 100$ and $n = 1000$, taking the average results of 100 monte-carlos, we obtain the graphs of $\hat{f}(x)$ for $n = 100$ *(red line)* and $n = 1000$ *(blue line)*, in **Figure 3**.

The nonparametric kernel estimators do not fit the true density perfectly, but performing quite well, even for the small sample of $n = 100$. When n increase, order of error decrease. As the previous discussion and our expectation, for the large sample $n = 1000$, the estimated density is closer to the true curve.

3.2 The Estimate $\hat{\mathbf{f}}(\mathbf{Y})$ for Two-dimensional Random Vector Y

Randomly generate n draws of a random variable Y:

$$Y \sim N\left[\begin{bmatrix}4\\2\end{bmatrix}, \begin{bmatrix}9 & 1\\1 & 4\end{bmatrix}\right]$$

Then, the **true density function** of $Y = (y_{01}, y_{02})$ is that:

$$f(\mathbf{Y_0}) = \frac{\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{y_{01}-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2 - 2\rho\left(\frac{y_{02}-\mu_{Y_2}}{\sigma_{Y_2}}\right)\left(\frac{y_{01}-\mu_{Y_1}}{\sigma_{Y_1}}\right) + \left(\frac{y_{02}-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right]\right\}}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt[2]{(1-\rho^2)}} \tag{3}$$

where: $\mu_{Y_1} = 4$; $\mu_{Y_2} = 2$; $\sigma_{Y_1} = 3$; $\sigma_{Y_2} = 2$; $\rho\sigma_{Y_1}\sigma_{Y_2} = 1 \Rightarrow \rho = \frac{1}{6}$. The chosen grid of points $\mathbf{Y_0}$ is $[-5; 15] \times [-5; 15]$. Putting each pairs of values $(y_{01}; y_{02}$ in this considered range into eq.(3), we will obtain the true density *(which is the green line in* ***Figure 4*** *)*.

Assuming that the data of $\mathbf{Y} = (Y_1, Y_2)$ is given, we use the nonparametric kernel estimator for the joint density of $\mathbf{Y}$ in *eq.(4)*. Taking $h_{1n} = h_{2n} = h$ (from the previous part) and choose multivariate kernel to be the product of univariate kernel, we obtain *eq.(5)*:

$$\begin{aligned} \hat{f}(Y_0) &= \frac{1}{nh_{1n}h_{2n}} \sum_{i=1}^{n} \left(\frac{Y_{i1} - Y_{01}}{h_{1n}}, \frac{Y_{i2} - Y_{02}}{h_{2n}}\right) & (4) \\ &= \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{Y_{i1} - Y_{01}}{h}\right) K\left(\frac{Y_{i2} - Y_{02}}{h}\right) & (5) \end{aligned}$$

The estimates of the joint density will be consistent if $n \to \infty$, $h \to \infty$, $nh^2 \to \infty$

On the same chosen grids of points $\mathbf{Y_0}$, estimating the joint density of $\mathbf{Y}$ for both the sample $n = 100$ and $n = 1000$, taking the average results of 100 monte-carlos, we obtain the graphs of $\hat{f}(\mathbf{Y})$, for $n = 100$ *(in red)* and $n = 1000$ *(in blue)*, in **Figure 4**.

The nonparametric kernel estimators perform quite well. As the expectation, for the larger sample $n = 1000$, the estimated joint density gets closer to the true density. However, the improvement comparing to $n = 100$ is not large. The bias in the large sample $n = 1000$ is still big.

3.3 Discussion about the Rate of Convergence of $\hat{\mathbf{f}}(\mathbf{X})$ and $\hat{\mathbf{f}}(\mathbf{X})$

Visually through the Monte-Carlo Stimulation, we can see that the bivariate joint density $\hat{f}(\mathbf{Y})$ converges more slowly than the univariate density. When taking to account more dimensions, the rate of converge will be slower *(Curse of Dimensionality)*. It will require a larger sample size to obtain the same accurate estimate.

Univariate Density Estimate Considering the univariate density estimation $\hat{f}(x)$ in *equ.(2)*, for MSE (similar for IMSE or ISE):

$$
\begin{aligned}
MSE[\hat{f}(x)] = [Bias[\hat{f}(x)]]^2 - Var[\hat{f}(x)] &= O(h_n^4) + O\left(\frac{1}{nh_n}\right) \\
\frac{dMSE[\hat{f}(x)]}{dh_n} = 0 &\Rightarrow O(h_n^3) = O\left(\frac{1}{nh_n^2}\right) \\
&\Rightarrow h_n^{opt} \propto n^{-1/5}
\end{aligned}
$$

In this case, the rate of convergence of $\hat{f}$ to f, or the rate of convergence of Bias to 0, is of order $O(n^{-2/5})$.

Bivariate Density Estimate Considering the univariate density estimation $\hat{f}(\mathbf{Y})$ in *equ.(5)*, for MSE (similar for IMSE or ISE):

$$
\begin{aligned}
MSE[\hat{f}(Y)] &= [Bias[\hat{f}(Y)]]^2 - Var[\hat{f}(Y)] \\
&= O[(h_{1n}^2 + h_{2n}^2)^2] + O\left(\frac{1}{nh_{1n}h_{2n}}\right) = O(h^4) + O\left(\frac{1}{nh_n^2}\right) \\
\frac{dMSE[\hat{f}(x)]}{dh_n} &= 0 \\
\Rightarrow O(h_n^3) &= O\left(\frac{1}{nh_n^3}\right) \\
\Rightarrow h_n^{opt} &\propto n^{-1/6}
\end{aligned}
$$

In this case, the rate of convergence of $\hat{f}$ to f, or the rate of convergence of Bias to 0, is of order $O(n^{-2/6})$.

Comparing the rate of convergence, the univariate density estimates $\hat{f}(X)$ will converge faster to the truth (as in the simulation).

4 Exercise 4

4.1 Exploring Data

The summary of GDP values in 2005 and 2016 is reported in the following table:

GDP Year	Min	1st Quant.	Mean	Median	3rd Quant.	Max	Std. dev
2005	0.1	2.2	12.4	244.1	101.4	13090.0	1075.8
2016	0.2	6.7	27.4	389.2	192.5	18570.0	1652.9

The difference between the highest and lowest GDP values is high. The standard deviation is large. Apparently, the sample is affected by outliers. Therefore, we focus on the range from *0-1500* billions of dollars in this estimate *(eliminate extreme values)*. Even in this range, the difference between the highest and lowest is far, yet due to the limited number of observations, we cannot narrow the range. Therefore, the similar procedure will be applied to obtain the density estimation for the **log GDP** values, which transformed the data to be more well-behaved and appropriate for investigating.

4.2 Evidence from the Density Estimation of GDP values

The estimator of density function is defined as:

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{h_n}^{n} K\left(\frac{x - X_i}{h_n}\right)$$

where $n = 191$ *(GDP values from 191 countries in 2005 and 2016)*; the applied Kernel function will be Epanechnikov second-order $K(u) = \frac{3}{4}(1 - u^2)1_{(|u|\leq 1)}$. The bandwidths h_n is calculated by different methods: *i) Likelihood Cross-validation (CV-ML); ii) Rule-of-thumb; iii) Least-squared Cross-validation (LS-CV)*. The different values of bandwidths are reported in the following table:

	Year	CV Likelihood	Rule-of-thumb	LS CV
1	2005	11.51	6.43	0.37
2	2016	91.52	14.28	0.65

The graphs of density estimation of GDP in 2005 and 2016 is in **Figure 5**. Comparing to other methods, the optimal bandwidths suggested by LS-CS is much smaller. Thus, by this method, the convergence is faster, the error is smaller, yet the graph is noisier (under-smoothed).

In all graphs, the density for the lowest part of GDP decreased, from 2005 to 2016. The density graph by *Likelihood Cross-validation bandwidths* shows a clear right-shift, which indicates the growth in GDP for most of countries. The evidence of growing is less apparent for the graphs with LS-CV and rule-of-thumb bandwidths.

However, there is not enough evidence to claim that: "Majority of the countries experiences a recession in the last decade."

4.3 Evidence from the Density Estimation of Log GDP values

The description of GDP values after transforming is in the following table:

Log GDP	Min	1st Quant.	Mean	Median	3rd Quant.	Max	Std. dev
2005	-2.303	1.193	2.518	2.790	4.619	9.480	2.3768
2016	-1.609	1.902	3.311	3.447	5.260	9.829	2.3383

In this case, the transformed data is more appropriate to analyze: the range is more concentrated, the standard deviation is much smaller. Then, the difference among optimal bandwidths by different methods is smaller. Among different method, LS-CV suggests a smallest optimal bandwidth, which means that the error of this method is smaller. The graphs of density estimation of Log GDP in 2005 and 2016 is in **Figure 6**.

By similar methods, different values of bandwidths for the density estimation of log GDP are reported in the following table:

	Year	CV Likelihood	Rule-of-thumb	LS CV
1	2005	0.85	0.88	0.61
2	2016	0.47	0.87	1.05

The density graph by *Likelihood Cross-validation bandwidths* shows a clear right-shift, which indicates the growth in GDP for most of countries. Meanwhile, the density estimation by the rule-of-thumb shows an unchanged status for 2005 and 2016. Finally, in the density estimation by LS-CV, there is a slightly right-shift, which shows that in most of countries the economies grow.

In conclusion, in most of countries, the levels of GDP are at least maintained the same in 2016 and 2005. **It rejects the claim: "Majority of the countries experiences a recession in the last decade."**

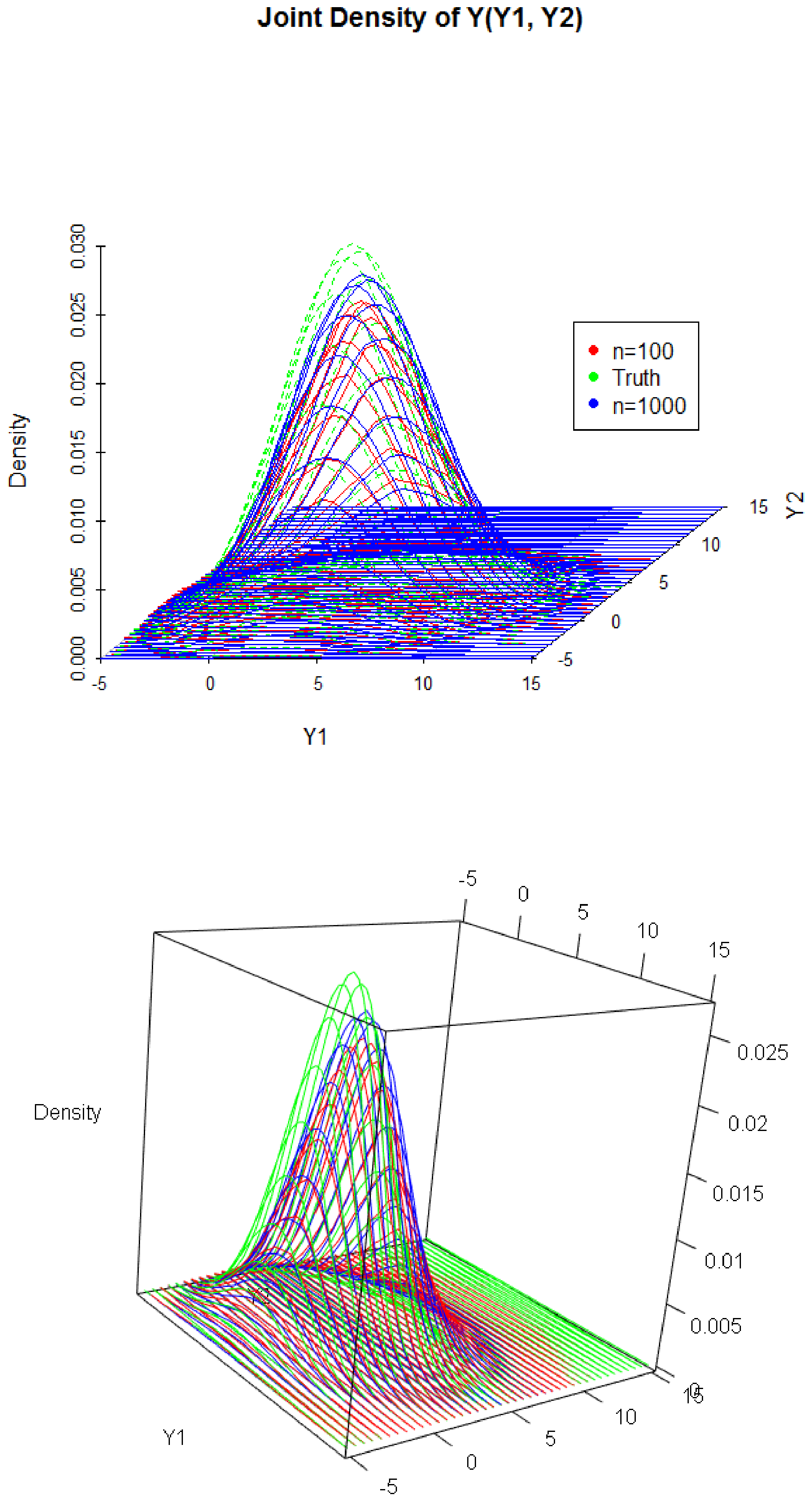
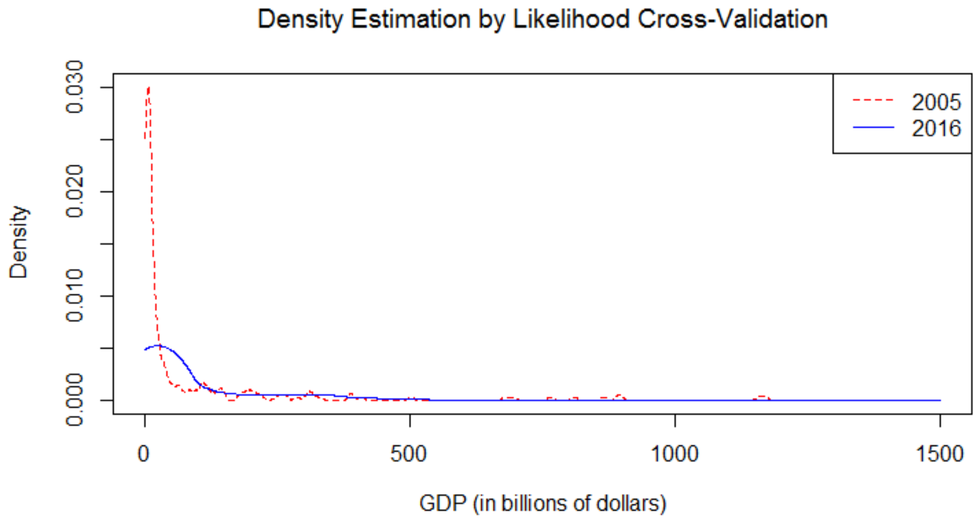
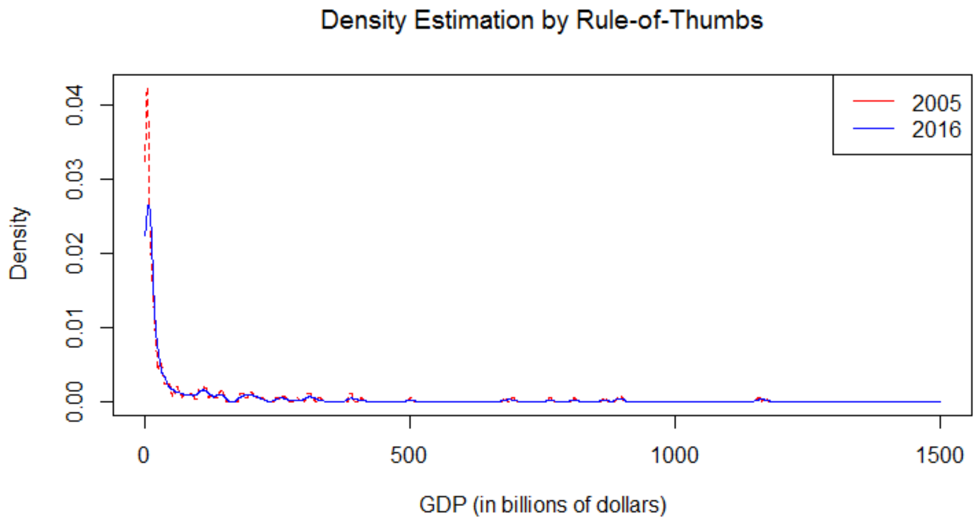


Figure 4: Monte-Carlo Simulation: Joint Density Estimation of two-dimensional Y





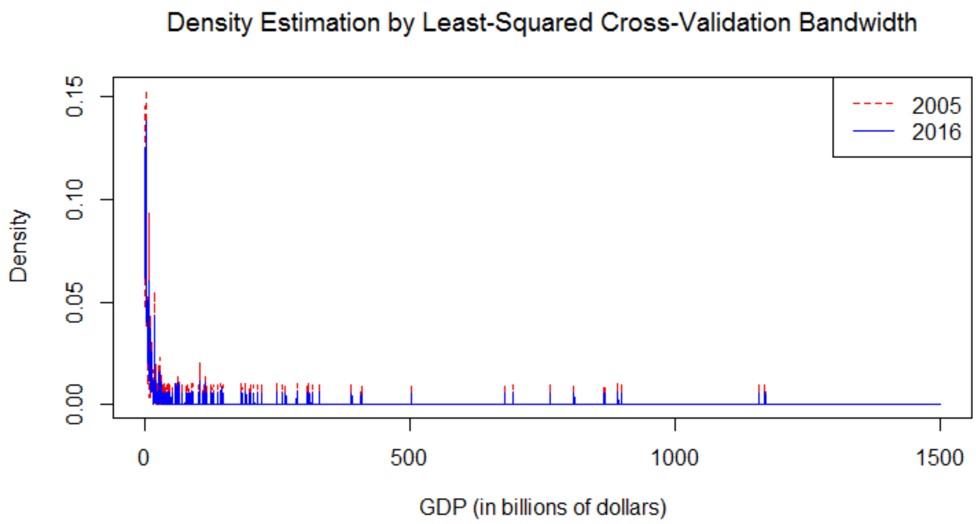


Figure 5: The graphs of density estimation of GDP for 2005 and 2016

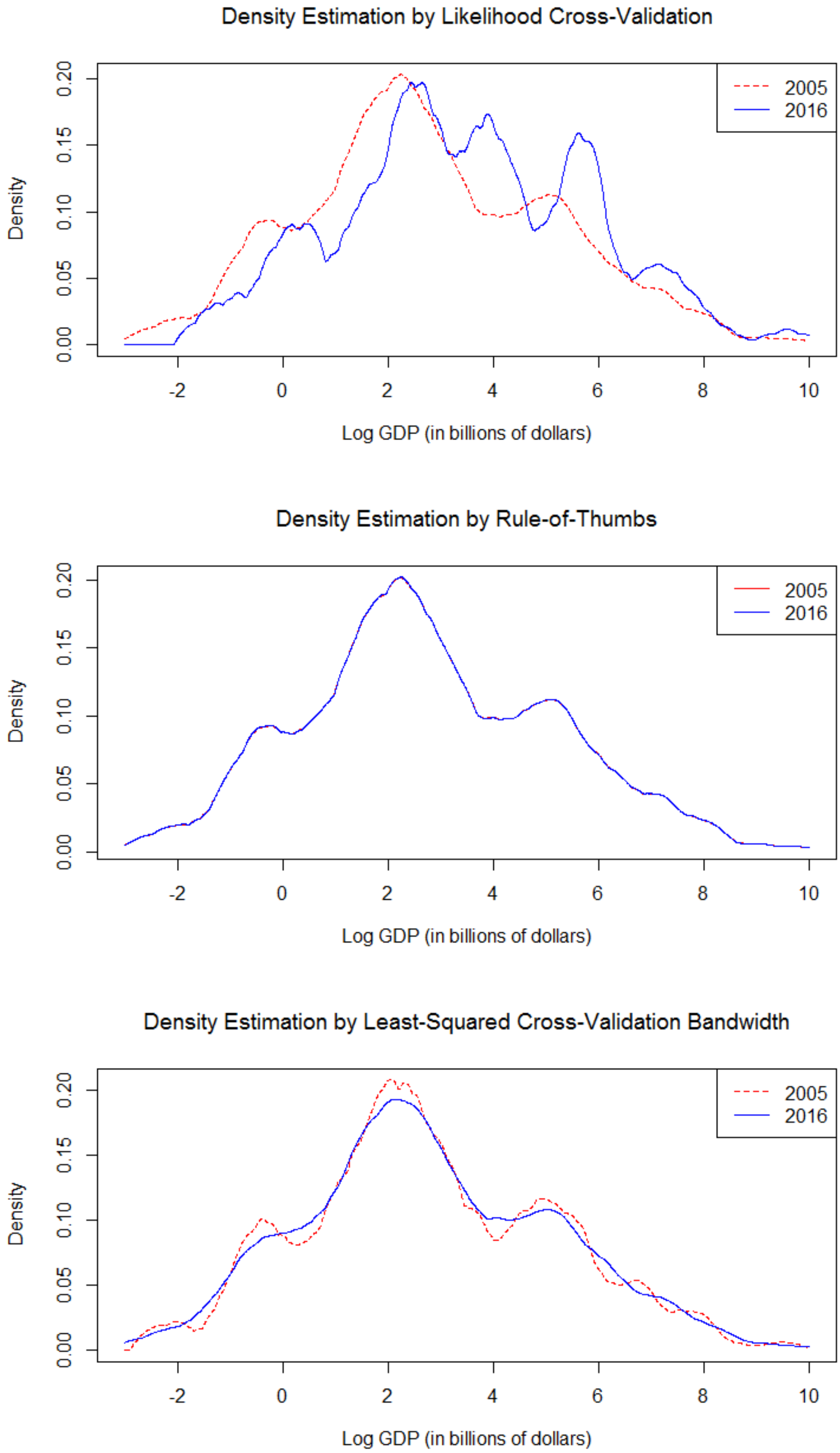


Figure 6: The graphs of density estimation of Log GDP for 2005 and 2016