Paul Le Tran
plt377
20/03/22
ECO388E

Problem Set 4 (Problem Set 1 for spring 2022)

1i).

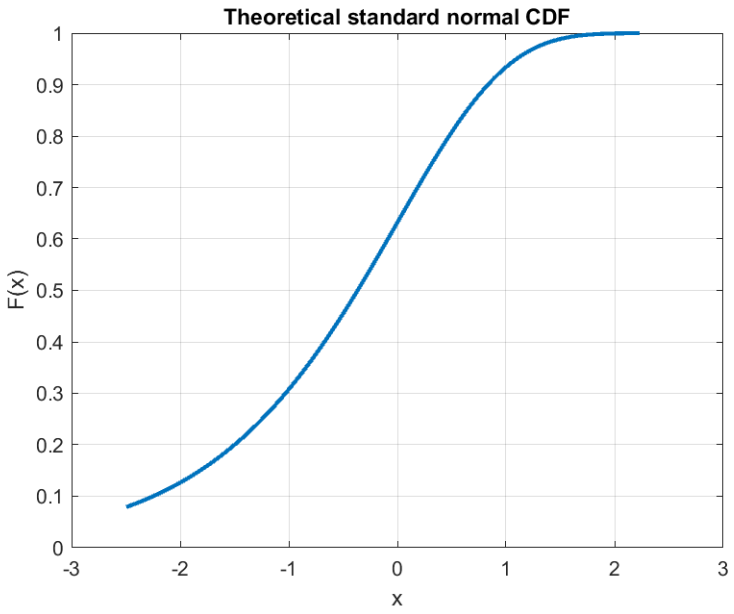

Figure 1: Theoretical standard normal cumulative distribution function (CDF).

1ii).

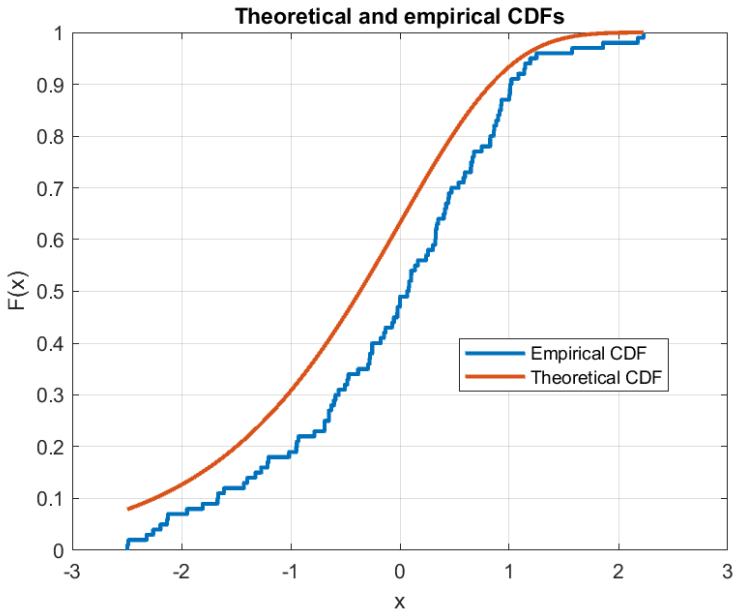

Figure 2: Theoretical standard normal CDF with the empirical CDF created using 100 iid observations drawn from the standard normal distribution.

1iii).

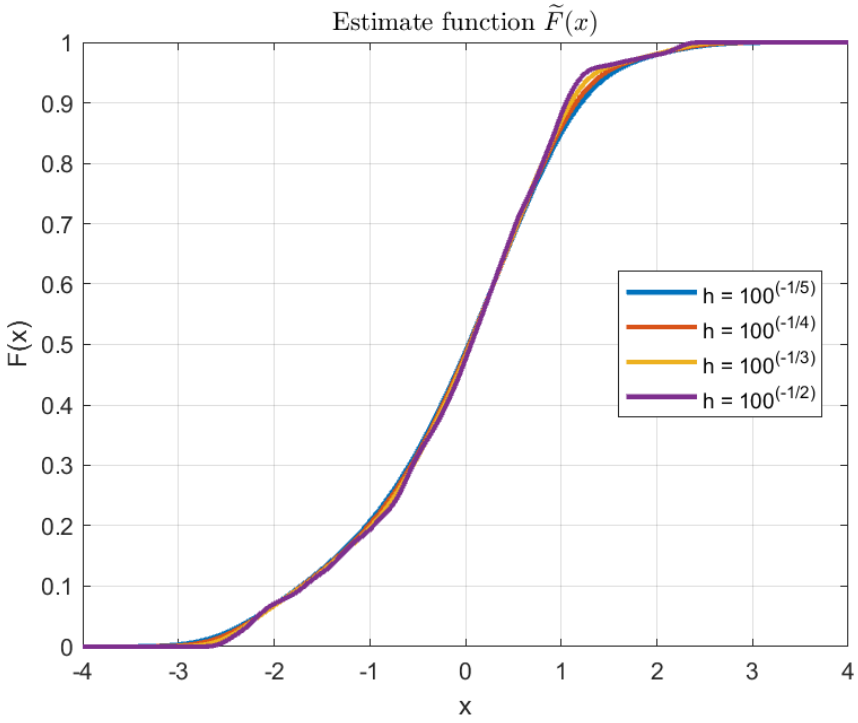

Figure 3: Empirical CDF created by the estimator function $\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x-X_i}{h}\right)$ for different bandwidth values, h.

2i).

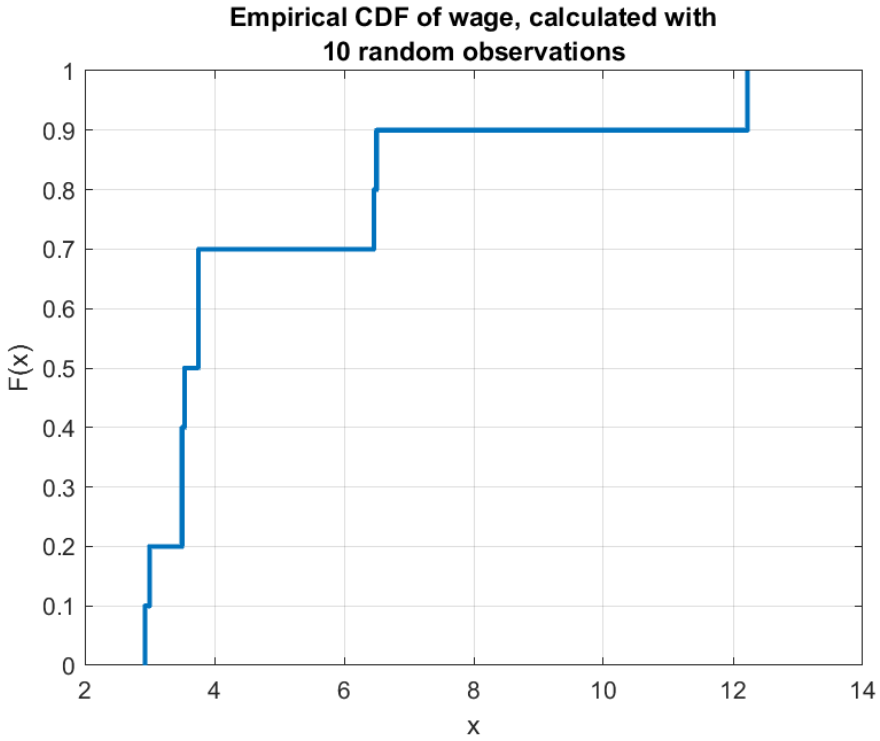

Figure 4: Empirical CDF of wages created with ten randomly chosen observations from the sample.

2ii).

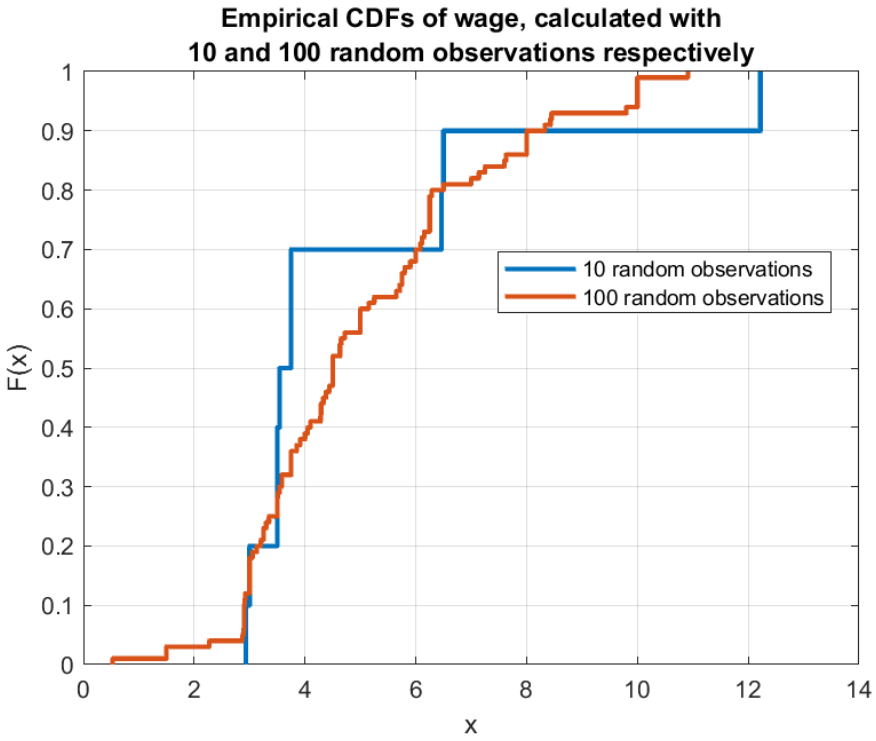

Figure 5: Empirical CDFs of wages created with ten and 100 randomly chosen observations from the sample.

2iii).

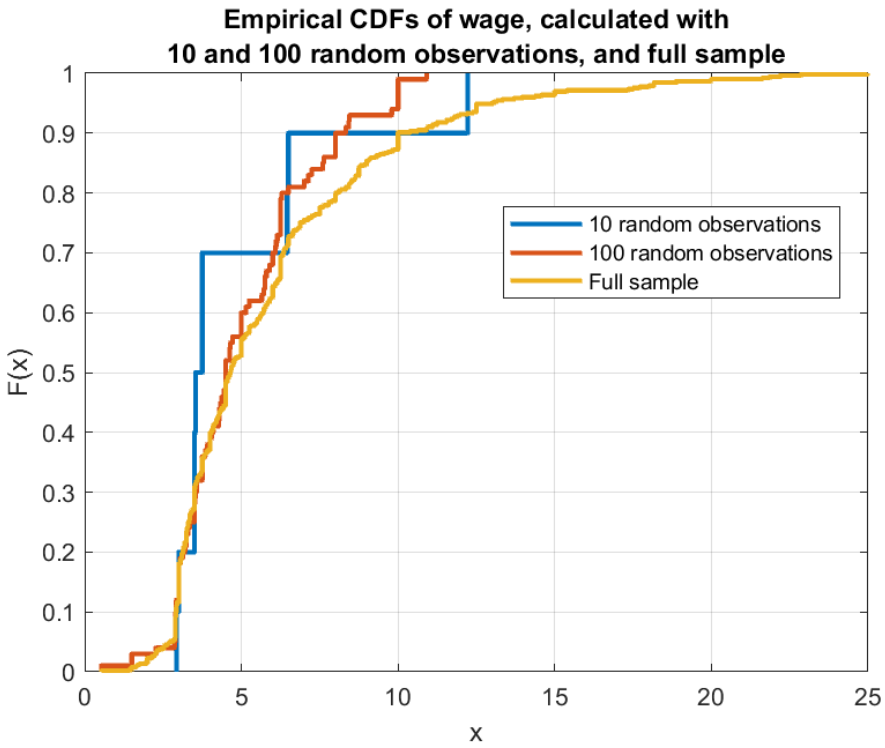

Figure 6: Empirical CDFs of wages created with ten and 100 randomly chosen observations from the sample, then using the entire sample itself.

2iv).

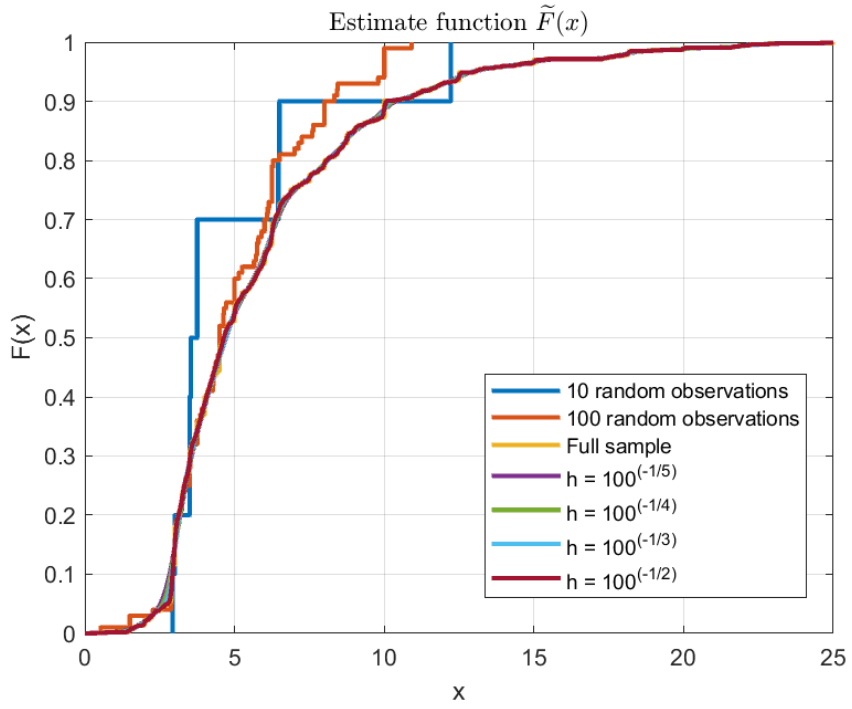

Figure 7: Empirical CDFs of wages created with ten and 100 randomly chosen observations from the sample, using the entire sample itself, and using the estimator function $\tilde{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{x-X_i}{h}\right)$ for different bandwidth values, h.

3i).

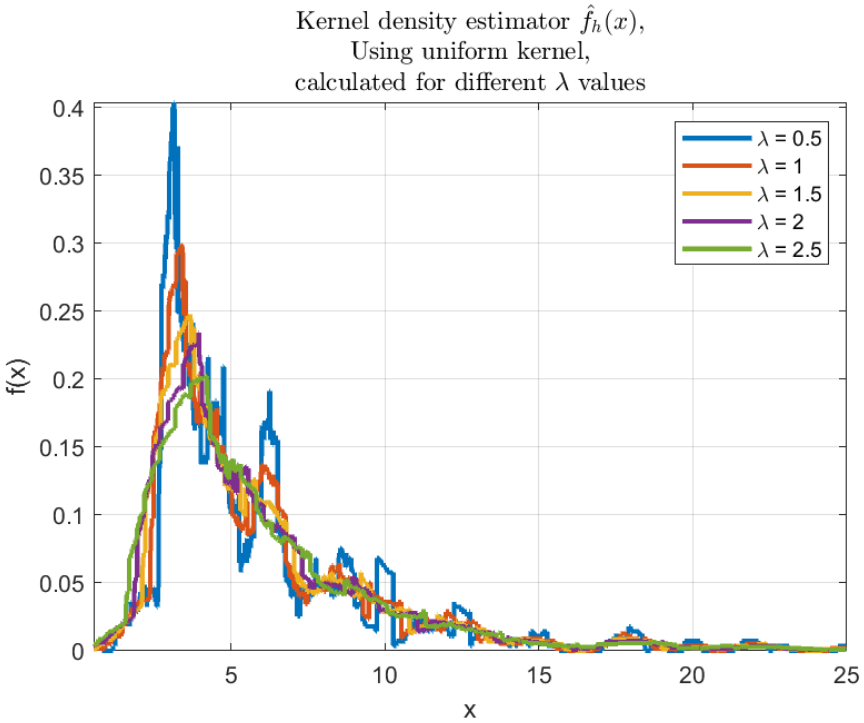

Figure 8: Density of wages estimated using the uniform kernel, bandwidth $h = \lambda\hat{\sigma}n^{-1/5}$, and for different values of λ.

3ii).

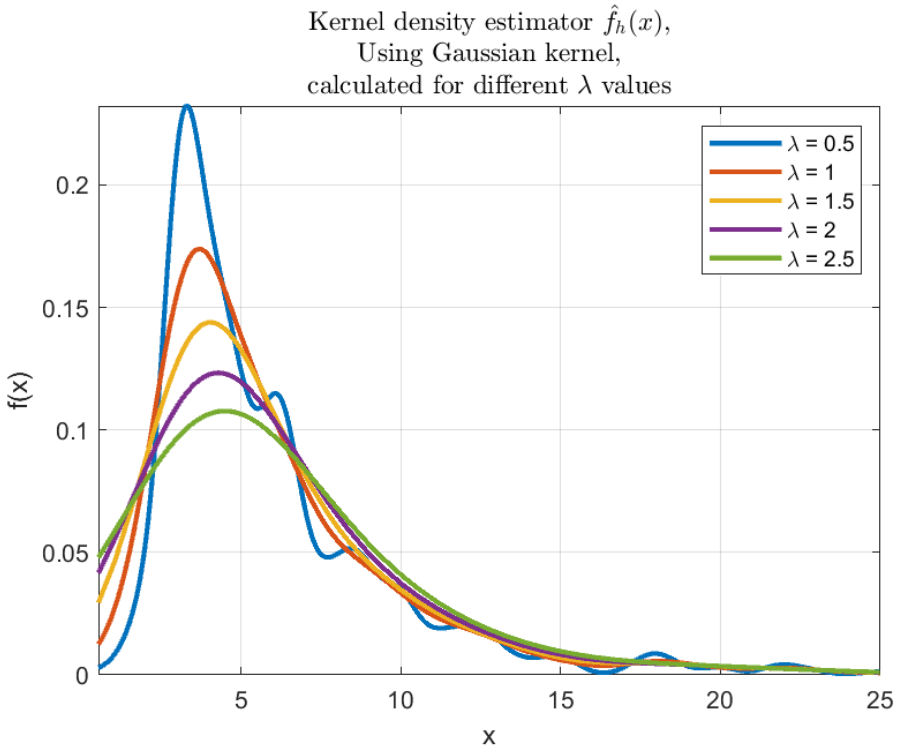

Figure 9: Density of wages estimated using the Gaussian kernel, $h = \lambda\hat{\sigma}n^{-1/5}$, and for different values of λ.

3iii).

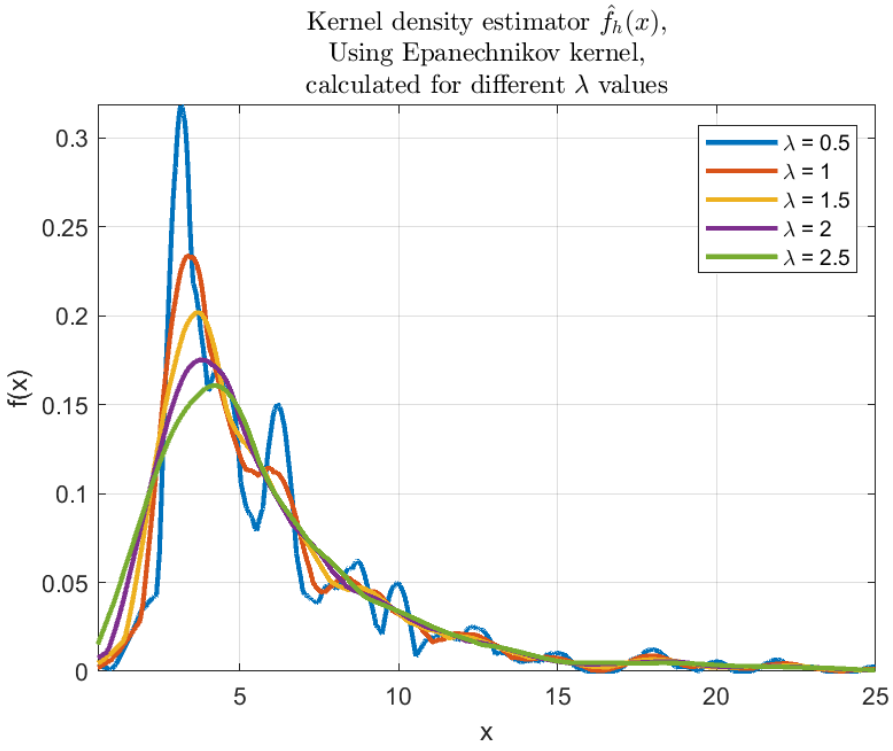

Figure 10: Density of wages estimated using the Epanechnikov kernel, $h = \lambda\hat{\sigma}n^{-1/5}$, and for different values of λ.

3iv).

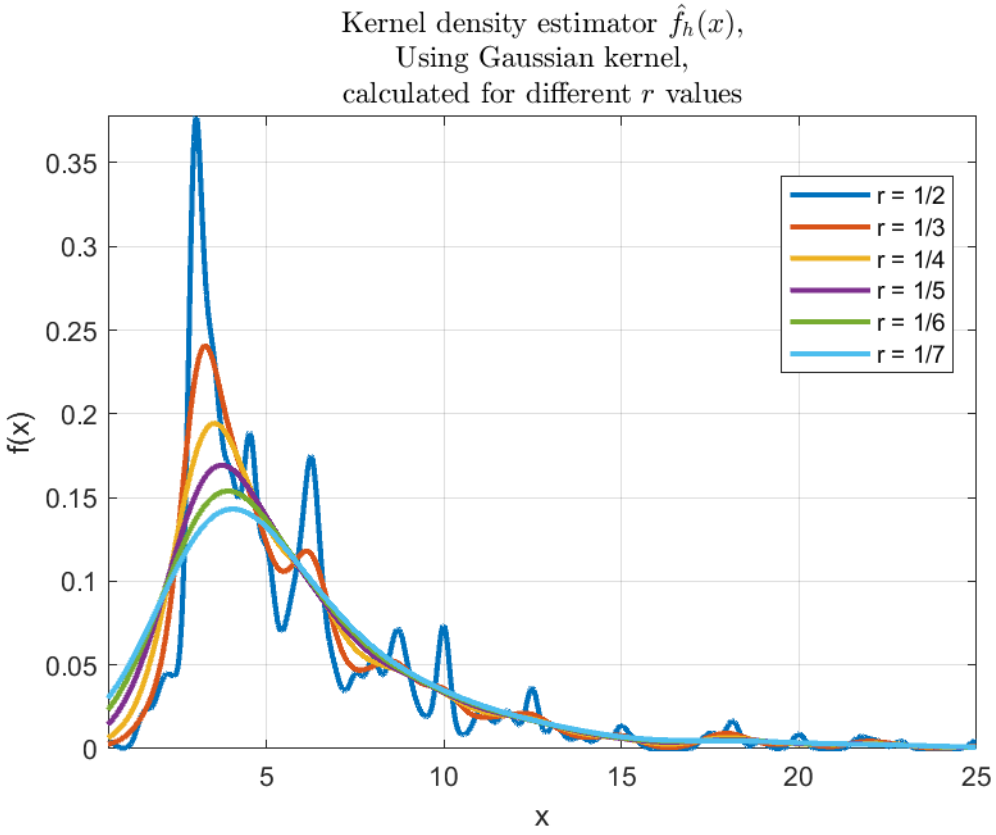

Figure 11: Density of wages estimated using the Gaussian kernel, $h = 1.06\hat{\sigma}n^{-\mathrm{r}}$, and for different values of r.

3v). The cross-validation method chosen to perform kernel density estimation for the density of wages was splitting the sample of wages. Specifically, the entire sample of wages was split into two equally sized subsamples, S_1 and S_2. The observations in both subsamples were randomly assigned. Using the Gaussian kernel, we then created the kernel density equation with S_1 as the training set, and performed maximum likelihood estimation (MLE) for the testing set, S_2 to find optimal bandwidth $h_{S_2} = 0.5227$. Similarly, we performed the same exercise with S_2 as the training set and S_1 as the testing set, giving $h_{S_1} = 0.6022$.

For the actual kernel density estimate for the distribution of wages in our sample, we took the average of the two bandwidths ($h = 0.5625$) as the chosen bandwidth for the estimate. Figure 11 below displays the result.

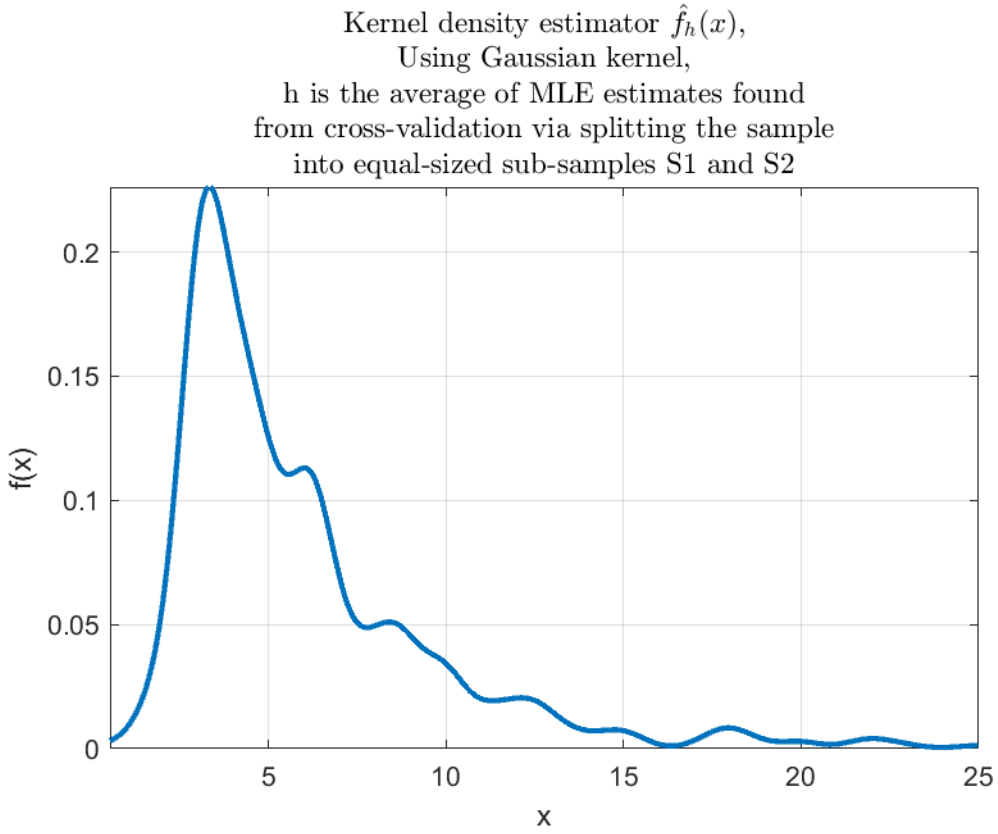

Figure 11: Density of wages estimated using the Gaussian kernel. Bandwidth h is the average of the MLE estimates found from cross-validation via splitting the sample into equally sized sub-samples S_1 and S_2.