Economics 675: Applied Microeconometrics – Fall 2018

Assignment 2 – Due date: Mon 8-Oct

Last updated: June 4, 2018

Contents

Guidelines:

- You may work in (small) groups while solving this assignment.
- Submit individual solutions via `http://canvas.umich.edu` in one PDF file collecting everything (e.g., derivations, figures, tables, computer code).
- Computer code should be done in both `Stata` and `R`. If the numerical results do not agree across the two statistical software platforms, you must explain why that is the case.
- Start each question on a separate page. Always add a reference section if you cite other sources.
- Clearly label all tables and figures, and always include a brief footnote with useful information.
- Always attach your computer code as an appendix, with annotations/comments as appropriate.
- Please provide as much detail as possible in your answers, both analytical and empirical.

1 Question 1: Kernel Density Estimation

Let $\{x_i \in \mathbb{R} : 1 \le i \le n\}$ be a random sample. Consider the kernel density (derivative) estimator

$$\hat{f}^{(s)}(x) = \frac{\partial^s}{\partial x^s}\hat{f}(x), \quad \hat{f}(x) = \hat{f}(x; h_n), \quad \hat{f}(x;h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right), \quad s = 0, 1, 2, \cdots, S,$$

where h_n is a positive bandwidth sequence, and $K(\cdot)$ is a P-order and s-times differentiable kernel, with $P > s$.

1. Providing regularity conditions, show that

$$\mathbb{E}[\hat{f}^{(s)}(x)] = f^{(s)}(x) + h_n^P \cdot \mu_P(K) \cdot \frac{f^{(P+s)}(x)}{P!} + o(h_n^P),$$

$$\mathbb{V}[\hat{f}^{(s)}(x)] = \frac{1}{nh_n^{1+2s}} \cdot \vartheta_s(K) \cdot f(x) + o\left(\frac{1}{nh_n^{2s+1}}\right),$$

   where

$$\mu_\ell(K) = \int_{\mathbb{R}} u^\ell K(u) du, \qquad \vartheta_\ell(K) = \int_{\mathbb{R}} \left[K^{(\ell)}(u)\right]^2 du.$$

2. Recall that for the kernel density estimator, the integrated mean squared error (IMSE) is defined as

$$\mathsf{IMSE}\,[h] = \int \mathbb{E}\left[(\hat{f}(x) - f(x))^2\right] dx,$$

   and the asymptotic IMSE (AIMSE) is defined by only keeping the first order bias and variance

$$\mathsf{AIMSE}\,[h] = \int \left[\left(h_n^P \cdot \mu_P(K) \cdot \frac{f^{(P+s)}(x)}{P!}\right)^2 + \frac{1}{nh_n^{1+2s}} \cdot \vartheta_s(K) \cdot f(x)\right] dx.$$

   Show that AIMSE-optimal bandwidth choice (for estimating the s-th derivative) is

$$h_{\mathsf{AIMSE},s} = \left[\frac{(2s+1)(P!)^2}{2P} \frac{\vartheta_s(K)}{\vartheta_{s+P}(f) \cdot \mu_P(K)^2} \frac{1}{n}\right]^{\frac{1}{2s+2P+1}}$$

   Propose a consistent, fully data-driven bandwidth selection procedure. That is, propose a bandwidth estimator $\hat{h}_{\texttt{IMSE},s}$ such that $\hat{h}_{\texttt{IMSE},s}/h_{\texttt{IMSE},s} \to_p 1$.

3. Conduct the following Monte Carlo experiment.

   (a) Let the DGP be the mixture Gaussian $x_i \sim 0.5 \cdot \mathcal{N}(-1.5, 1.5) + 0.5 \cdot \mathcal{N}(1, 1)$, compute theoretically the AIMSE-optimal bandwidth for $s = 0$, sample size $n = 1,000$, with the Epanechnikov kernel $K(u) = 0.75(1 - u^2) \cdot \mathbf{1}_{\{|u| \le 1\}}$ (note that this is a second order kernel hence $P = 2$).

   (b) Keep the previous setup, and simulate the model for $M = 1,000$ times. For each simulation $1 \le m \le M$, denote by $\hat{f}_m(\cdot)$ the density estimator and $\hat{f}_{m,(i)}(\cdot)$ the estimator without using the

i-th observation. Compute the following quantities:

$$\widehat{\mathsf{IMSE}}^{\mathtt{LI}}[h] = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{n}\sum_{i=1}^{n}(\hat{f}_m(x_i;h) - f(x_i))^2, \qquad h \in \mathcal{H} = h_{\mathsf{AIMSE}} \times \{0.5, 0.6, \cdots, 1.4, 1.5\},$$

$$\widehat{\mathsf{IMSE}}^{\mathtt{LO}}[h] = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{n}\sum_{i=1}^{n}(\hat{f}_{m,(i)}(x_i;h) - f(x_i))^2, \qquad h \in \mathcal{H} = h_{\mathsf{AIMSE}} \times \{0.5, 0.6, \cdots, 1.4, 1.5\},$$

Plot $\widehat{\mathsf{IMSE}}^{\mathtt{LI}}[h]$ as a function of $h \in \mathcal{H}$, and compute $h_{\widehat{\mathsf{IMSE}},\mathtt{LI}} = \min_{h\in\mathcal{H}} \widehat{\mathsf{IMSE}}^{\mathtt{LI}}[h]$.

Plot $\widehat{\mathsf{IMSE}}^{\mathtt{LO}}[h]$ as a function of $h \in \mathcal{H}$, and compute $h_{\widehat{\mathsf{IMSE}},\mathtt{LO}} = \min_{h\in\mathcal{H}} \widehat{\mathsf{IMSE}}^{\mathtt{LO}}[h]$.

(c) True or False: $h_{\widehat{\mathsf{IMSE}},\mathtt{LI}}/h_{\mathsf{AIMSE}} \to_p 1$. Explain intuitively.

(d) Consider the the following inconsistent bandwidth selection procedure:

- Assume the data has Gaussian distribution and let $\hat{\mu}$ and $\hat{\sigma}$ be the estimated mean and standard deviation.
- Let $\phi_{\hat{\mu},\hat{\sigma}}$ be the normal density with the estimated mean and standard deviation, and

$$\hat{h}_{\mathsf{AIMSE}} = \left[\frac{\vartheta_0(K)}{\vartheta_2(\phi_{\hat{\mu},\hat{\sigma}})\cdot\mu_2(K)^2}\frac{1}{n}\right]^{\frac{1}{5}}.$$

Implement this procedure, and for each simulation $1 \le m \le M$, compute the quantity $\hat{h}_{\mathsf{AIMSE},m}$ and the average $\bar{h}_{\mathsf{AIMSE}} = M^{-1}\sum_{m=1}^{M}\hat{h}_{\mathsf{AIMSE},m}$.

2 Question 2: Linear Smoothers, Cross-validation and Series

Let $\{(y_i, x_i) \in \mathbb{R}^2 : 1 \leq i \leq n\}$ be a random sample. A linear smoother is given by

$$\hat{e}(x) = \sum_{i=1}^{n} w_{n,i}(x) y_i, \qquad w_{n,i}(x) = w_{n,i}(x_1, x_2, \cdots, x_n; x),$$

where $w_{n,i}(x)$ is only a function of $\{x_i : 1 \leq i \leq n\}$, but not of $\{y_i : 1 \leq i \leq n\}$.

1. Show that local polynomial regression and series estimators can be written as linear smoothers, and give the exact form of the "smoothing weights" $w_{n,i}(x)$ in each case.

2. Show the following simplified cross-validation formula holds for local polynomial regression and the series estimation[1]:

$$\mathsf{CV}(c) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{e}_{(i)}(x_i)\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{e}(x_i)}{1 - w_{n,i}(x_i)} \right)^2,$$

   where c denotes a tuning parameter (i.e., a bandwidth h_n for local polynomials, or a series truncation K for series estimators).

3. Providing regularity conditions, show that

$$\frac{\hat{e}(x) - e(x)}{\sqrt{\mathbb{V}[\hat{e}(x)|x_1, x_2, \cdots, x_n]}} \to_d \mathcal{N}(0, 1), \qquad \text{almost surely } \{x_i\}$$

   and derive the formula for $\mathbb{V}[\hat{e}(x)|x_1, x_2, \cdots, x_n]$.
   Propose a consistent standard error estimator.

4. Propose an asymptotically valid 95% confidence interval for $e(x)$, with x fixed.
   Explain the difference between pointwise valid and uniformly (over x) valid inference.

5. Conduct the following Monte Carlo experiment.

   (a) Consider the following DGP

   - $x_i \sim \mathsf{Uniform}(-1, 1)$;
   - $y_i = \exp(-0.1 \cdot (4x_i - 1)^2) \cdot \sin(5x_i) + \varepsilon_i$;
   - $\varepsilon_i \sim x_i^2 \cdot (\chi_5^2 - 5)$, where $x_i \perp\!\!\!\perp \chi_5^2$ and χ_5^2 denotes a random variable with a chi-squared distribution with 5 degrees of freedom.

   Set $n = 1000$ and generate $M = 1,000$ replications for this model.

   (b) Consider a power series estimator of $\mu(x_i) = \mathbb{E}[y_i|x_i]$, that is, $\hat{\mu}(x_i) = \mathbf{p}^K(x_i)'\hat{\boldsymbol{\gamma}}_K$ with $\mathbf{p}^K(x_i) = (1, x_i, x_i^2, \cdots, x_i^K)'$ and $\hat{\boldsymbol{\gamma}}_K$ the appropriate OLS estimator. Plot the average $\mathsf{CV}(K)$, across the M simulations, as a function of K and compute the CV estimator, denoted $\hat{K}_{\mathsf{CV}}$. (Use K from 1 to 20, and do not forget to add the intercept.)

[1] The following result is useful: for an invertible matrix $\mathbf{A}$ and a column vector $\mathbf{v}$, the following holds

$$(\mathbf{A} + \mathbf{v}\mathbf{v}')^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{v}\mathbf{v}'\mathbf{A}^{-1}}{1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{v}},$$

provided that $1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{v} \neq 0$.

(c) Using the data-driven tuning parameter choice $\hat{K}_{\texttt{CV}}$, plot the following functions of $x \in [-1, 1]$ (in one single graph): (i) the true regression function $\mu(x)$; (ii) the average estimated regression function $\hat{\mu}(x)$ across the M simulations; (iii) the average of confidence intervals for $\mu(x)$ across the M simulations. (For parts (ii) and (iii), using a grid of 10 evaluation points should be enough.)

(d) Repeat part (c) for the derivative of $\mu(x)$, that is, for $\mu^{(1)}(x) = \partial\mu(x)/\partial x$, with $x \in [-1, 1]$

3 Question 3: Semiparametric Semi-Linear Model

A very popular model in empirical work is the partially linear model, a simplified version of which is given as follows. Let $\mathbf{z}_i = (y_i, t_i, \mathbf{x}_i)'$, $i = 1, 2, \cdots, n$, be a random sample from the random vector $\mathbf{z} = (y, t, \mathbf{x})'$, where $y \in \mathbb{R}$ is a response variable, $t_i \in \{0, 1\}$ is a treatment indicator, and $\mathbf{x}_i \in \mathbb{R}^d$ is a set of observed covariates, regressors or confounders. The model is:

$$y_i = t_i\theta_0 + g_0(\mathbf{x}_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i|t_i, \mathbf{x}_i] = 0 \qquad \text{and} \qquad \mathbb{E}[\varepsilon_i^2|t_i, \mathbf{x}_i] = \sigma_0^2(t_i, \mathbf{x}_i),$$

where usually the main parameter of interest is the treatment effect θ_0, being the functions $g_0(\cdot)$ and $\sigma_0^2(\cdot)$ just unknown infinite dimensional nuisance parameters. (In some applications, the function $g_0(\cdot)$ is actually the parameter of interest!)

1. Provide sufficient conditions so that θ_0 is identifiable.
   Show that θ_0 solves the moment condition $\mathbb{E}[(t_i - h_0(\mathbf{x}_i))(y_i - t_i\theta_0)] = 0$, where $h_0(\mathbf{x}_i) = \mathbb{E}[t_i|\mathbf{x}_i]$; $h_0(\mathbf{x})$ is sometimes called the propensity score. Find a closed form expression of θ_0, and provide an instrumental variables interpretation.

2. This model can be easily implemented using series estimation techniques. For simplicity, consider a power series approximation with basis expansion $\mathbf{p}^{K_n}(\mathbf{x}_i)$ and observe that replacing $g_0(\mathbf{x}) \approx \mathbf{p}^{K_n}(\mathbf{x})'\boldsymbol{\gamma}_K$ leads to $\mathbb{E}[y_i|\mathbf{x}_i] \approx t_i\theta_0 + \mathbf{p}^{K_n}(\mathbf{x}_i)'\boldsymbol{\gamma}_K$. Thus, for given K, the model can estimated by running the OLS regression of y_i on $(t_i, \mathbf{p}^{K_n}(\mathbf{x}_i)')'$ and extracting the first element of the estimated vector[2].

   (a) For fixed $K_n = K$, and using the usual least-squares partial-out formulas, give a closed form expression for the OLS estimator of θ_0, denoted $\hat{\theta}(K)$.

   (b) Show that if, instead, the approximation $h_0(\mathbf{x}) \approx \mathbf{p}^K(\mathbf{x}_i)'\boldsymbol{\delta}_K$ is used in the moment condition given in part 1, then the resulting M-estimator of θ_0 is numerically equal to $\hat{\theta}(K)$.

3. The asymptotic theory for the semiparametric estimator $\hat{\theta}(K)$ is very easy for fixed K (why?), and not very difficult for $K/n \to 0$. It can be shown that, under appropriate regularity conditions,

$$\frac{\hat{\theta}(K) - \theta_0}{\sqrt{\hat{V}_{\texttt{HCO}}}} \to_d \mathcal{N}(0, 1), \qquad \hat{V}_{\texttt{HCO}} = (\mathbf{T}'\mathbf{M_P}\mathbf{T})^{-1}(\mathbf{T}'\mathbf{M_P}\widehat{\boldsymbol{\Sigma}}\mathbf{M_P}\mathbf{T})(\mathbf{T}'\mathbf{M}\mathbf{T})^{-1},$$

   with $\mathbf{T} = (t_1, t_2, \cdots, t_n)'$, $\mathbf{M_P} = \mathbf{I}_n - \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}$, $\mathbf{P} = (\mathbf{p}^K(\mathbf{x}_1), \mathbf{p}^K(\mathbf{x}_2), \cdots, \mathbf{p}^K(\mathbf{x}_n))'$, and $\widehat{\boldsymbol{\Sigma}} = \operatorname{diag}(\hat{\varepsilon}_1^2, \hat{\varepsilon}_2^2, \cdots, \hat{\varepsilon}_n^2)$ with $(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n)' = \mathbf{M_P}(\mathbf{Y} - \mathbf{T}\hat{\theta}(K))$ and $\mathbf{Y} = (y_1, y_2, \cdots, y_n)'$.

   (a) Fixing K, give a formal justification to the results above; this approach is sometimes referred to as "flexible parametric" estimation and inference (why?). Letting $K \to \infty$ at an appropriate rate does not invalidate the "fixed-K" calculations, although it may not give a good enough finite-sample approximation!

   (b) Using the results above, propose an asymptotically valid 95% confidence interval for the treatment effect parameter θ_0.

[2]Note that here $g : \mathbb{R}^{d_x} \to \mathbb{R}$, so that one needs multivariate Taylor expansion. For simplicity assume $\mathbf{x}_i = (x_{1,i}, x_{2,i})'$ is bivariate, then a second order polynomial basis of $\mathbf{x}$ is

$$\mathbf{p}(\mathbf{x}_i) = (1,\ x_{1,i},\ x_{2,i},\ x_{1,i}x_{2,i},\ x_{1,i}^2,\ x_{2,i}^2)' = (1,\ x_{1,i},\ x_{1,i}^2)' \otimes (1,\ x_{2,i},\ x_{2,i}^2)',$$

where the $\otimes$ is called tensor product.

4. Conduct the following Monte Carlo experiment.

   (a) Set $\theta_0 = 1$, $d = 5$, $\mathbf{x}_i = (x_{1,i}, x_{2,i}, \cdots, x_{d,i})'$ with

   - $x_{\ell,i} \sim$ i.i.d. $\mathsf{Uniform}(-1, 1)$, $\ell = 1, \cdots, d$ and $i = 1, 2, \cdots, n$;
   - $\varepsilon_i = 0.3637899(1 + \|\mathbf{x}_i\|^2) \cdot v_i$ with $v_i \sim$ i.i.d. $\mathcal{N}(0, 1)$;
   - $g_0(\mathbf{x}_i) = \exp(\|\mathbf{x}_i\|^2)$;
   - $t_i = \mathbf{1}(\|\mathbf{x}_i\| + u_i \geq 0)$ with $u_i \sim$ i.i.d. $\mathcal{N}(0, 1)$;

   Note that the unknown functions ($g_0(\cdot)$, $h_0(\cdot)$, etc.) are neither linear not additive separable in the observed covariates. Consider $M = 1,000$ replications, with each replication taking a random sample of size $n = 500$ with all random variables $(\mathbf{x}_i', v_i, u_i)$ statistically independent.
   For implementation, consider the following power series basis expansion (i.e., $\mathbf{p}^K(\mathbf{x})$):

   Polynomial Basis Expansion: $d = 5$ and $n = 500$

K	$\mathbf{r}_K(\mathbf{x}_i)$	K/n
6	$(1, \mathbf{x}_{1i}, \mathbf{x}_{2i}, \mathbf{x}_{3i}, \mathbf{x}_{4i}, \mathbf{x}_{5i})'$	0.012
11	$(\mathbf{r}_6(\mathbf{x}_i)', \mathbf{x}_{1i}^2, \mathbf{x}_{2i}^2, \mathbf{x}_{3i}^2, \mathbf{x}_{4i}^2, \mathbf{x}_{5i}^2)'$	0.022
21	$\mathbf{r}_{11}(\mathbf{x}_i)$ + first-order interactions	0.042
26	$(\mathbf{r}_{21}(\mathbf{x}_i)', \mathbf{x}_{1i}^3, \mathbf{x}_{2i}^3, \mathbf{x}_{3i}^3, \mathbf{x}_{4i}^3, \mathbf{x}_{5i}^3)'$	0.052
56	$\mathbf{r}_{26}(\mathbf{x}_i)$ + second-order interactions	0.112
61	$(\mathbf{r}_{56}(\mathbf{x}_i)', \mathbf{x}_{1i}^4, \mathbf{x}_{2i}^4, \mathbf{x}_{3i}^4, \mathbf{x}_{4i}^4, \mathbf{x}_{5i}^4)'$	0.122
126	$\mathbf{r}_{61}(\mathbf{x}_i)$ + third-order interactions	0.252
131	$(\mathbf{r}_{126}(\mathbf{x}_i)', \mathbf{x}_{1i}^5, \mathbf{x}_{2i}^5, \mathbf{x}_{3i}^5, \mathbf{x}_{4i}^5, \mathbf{x}_{5i}^5)'$	0.262
252	$\mathbf{r}_{131}(\mathbf{x}_i)$ + fourth-order interactions	0.504
257	$(\mathbf{r}_{252}(\mathbf{x}_i)', \mathbf{x}_{1i}^6, \mathbf{x}_{2i}^6, \mathbf{x}_{3i}^6, \mathbf{x}_{4i}^6, \mathbf{x}_{5i}^6)'$	0.514
262	$(\mathbf{r}_{257}(\mathbf{x}_i)', \mathbf{x}_{1i}^7, \mathbf{x}_{2i}^7, \mathbf{x}_{3i}^7, \mathbf{x}_{4i}^7, \mathbf{x}_{5i}^7)'$	0.524
267	$(\mathbf{r}_{262}(\mathbf{x}_i)', \mathbf{x}_{1i}^8, \mathbf{x}_{2i}^8, \mathbf{x}_{3i}^8, \mathbf{x}_{4i}^8, \mathbf{x}_{5i}^8)'$	0.534
272	$(\mathbf{r}_{267}(\mathbf{x}_i)', \mathbf{x}_{1i}^9, \mathbf{x}_{2i}^9, \mathbf{x}_{3i}^9, \mathbf{x}_{4i}^9, \mathbf{x}_{5i}^9)'$	0.544
277	$(\mathbf{r}_{272}(\mathbf{x}_i)', \mathbf{x}_{1i}^{10}, \mathbf{x}_{2i}^{10}, \mathbf{x}_{3i}^{10}, \mathbf{x}_{4i}^{10}, \mathbf{x}_{5i}^{10})'$	0.554

   (b) For each value of K in the table above, compute across the M simulations the following: (i) average of $\hat{\theta}(K)$, (ii) average bias of $\hat{\theta}(K)$, (iii) sample variance of $\hat{\theta}(K)$, (iv) average of $\hat{V}_{\mathtt{HCO}}$, and (v) average coverage rate of the confidence intervals proposed in part 3.
   For each value of K, discuss and compare (i)–(v).
   Discuss the implications of choosing a K "too small" versus a K "too large".

   (c) Employing a CV approach to selecting K, report the same quantities as in part (b), and discuss the corresponding numerical findings.

References