Homework 7

36-467, Fall 2020

Due at 6 pm on Thursday, 29 October 2020

NOTE: This was supposed to have been due on the 22nd, but I (i) screwed up the release, and then (ii) left town for family business until it was too late to ask you to do this. This will be extra credit, meaning your grade on it will replace your lowest grade on the other homeworks (unless your grade on this is lower than on any of the other homeworks; attempting this can only improve your grade).

1. *The "sandwich covariance" for linear regression* In this problem, and this problem only, suppose our data consists of IID variables $X_1, \ldots X_n$, where each $X_i = (Y_i, Z_i)$; where both Y_i and Z_i are centered, so $\mathbb{E}[Y_i] = \mathbb{E}[Z_i] = 0$; and that we want to estimate a linear regression of Y on Z by least squares, so we would ideally like to find the b_0 which minimizes $m(b) = \mathbb{E}\left[(Y - bZ)^2\right]$. We do *not* assume that the true relationship between Y and Z is linear.

   (a) (5) Show that $m(b) = \mathrm{Var}[Y] + b^2\mathrm{Var}[Z] - 2b\mathrm{Cov}[Y, Z]$.

   (b) (5) Show that the second derivative of $m(b)$ (with respect to b) is $m''(b) = 2\mathrm{Var}[Z]$.

   (c) (5) With finite data, we approximate $m(b)$ by $M_n(b) = n^{-1}\sum_{i=1}^{n}(Y_i - bZ_i)^2$. Define the residual for the i^{th} observation as $R_i(b) = Y_i - bZ_i$. Show that the first derivative of $M_n(b)$ is $M_n'(b) = \frac{-2}{n}\sum_{i=1}^{n} R_i(b)Z_i$.

   (d) (8) Explain why it's reasonable, under our assumptions, to estimate $\mathrm{Var}[M_n'(b_0)]$ by

$$\hat{J}_n = \frac{4}{n^2}\sum_{i=1}^{n} R_i^2(\hat{b}_n)Z_i^2$$

   "Reasonable" here means you don't need to give a formal proof, but you should give reasons to explain why $\hat{J}_n$ is connected to $\mathrm{Var}[M_n'(b_0)]$. *Hint:* What're the expectations of the summands?

   (e) (8 Find an expression for the standard error of $\hat{b}_n$, the minizer of $M_n(b)$. Your answer should involve both $\hat{J}_n$ and the sample variance of Z (and possibly other things).

   (f) (4) Now assume that $Y = b_0 Z + \epsilon$ where ϵ is IID with mean 0 and variance σ^2. (That is, the usual linear-model assumptions hold.)

Show that your expression for the standard error from the last sub-problem will converge on $\sigma/\sqrt{\mathrm{Var}\left[Z\right]}$ for large n.

Notice that in this problem we did not assume that the linear regression model is right, or, if the relationship between Y and Z *is* linear, assume that the noise around the regression line has constant variance. What we've just done, in the next-to-last sub-problem, is the calculation of a "robust standard error" (because it's still valid if the usual assumptions are broken). In particular, this is a "heteroskedasticity-consistent" (HC) robust standard error (because it works even if the noise is "heteroskedastic", i.e., does not have constant variance).

2. *Estimating an AR(1) by optimizing* Suppose we're dealing with a stationary time series $X(t)$ which is centered, so $\mathbb{E}\left[X(t)\right] = 0$, and has autocovariance function $\mathrm{Cov}\left[X(t), X(t+h)\right] = \gamma(h)$. We want to estimate an AR(1) model by least squares, so we minimize the function $M_n(b) = (n-1)^{-1}\sum_{t=1}^{n-1}\left(X(t+1) - bX(t)\right)^2$. We call this minimizer $\hat{b}_n$. Take it on trust that $M_n(b) \to \mathbb{E}\left[\left(X(t+1) - bX(t)\right)^2\right] \equiv m(b)$ as $n \to \infty$. Define b_0 to be the minimizer of $m(b)$.

   Unless the sub-problem explicitly says otherwise, do *not* assume that the AR(1) model is correct.

   (a) (5) Show that $m(b) = \gamma(0)(1+b^2) - 2b\gamma(1)$.

   (b) (5) Show that $m''(b) = 2\gamma(0)$.

   (c) (5) Define the residuals $R(t;b)$ as $R(t;b) = X(t) - bX(t-1)$. Show that
   $$M_n'(b) = -\frac{2}{n-1}\sum_{t=1}^{n-1} R(t+1;b)X(t)$$

   (d) (5 Explain why $\hat{J}_n = \frac{4}{(n-1)^2}\sum_{t=1}^{n} R^2(t+1;\hat{b}_n)X^2(t)$ might not be a good estimate of $\mathrm{Var}\left[M_n'(b_0)\right]$.

   Note: There are techniques for calculating heteroskedastic-autocorrelation-consistent (HAC) robust standard errors, based on smoothing terms like R^2X^2; we'll revisit this topic towards the end of the course when we look at fitting regression models with autocorrelated noise.

3. *Estimating an AR(1) by optimizing, continued* Now suppose that the AR(1) model *is* right, so that $X(t+1) = b_0X(t) + \epsilon(t+1)$, where the ϵs all have mean 0, variance $\tau^2 > 0$ and are uncorrelated with each other and with earlier Xs.

   (a) (6 Express $m(b)$ in terms of b_0 (the true autoregressive coefficient), τ^2 and b.

   (b) (5) Show that $\mathrm{Cov}\left[X(t), \epsilon(t+1)\right] = 0$. *Hints:* Use the law of total expectation, and the facts that the innovations have expectation 0 and are uncorrelated with earlier Xs.

(c) (5) Show that $\mathrm{Cov}\left[X(t)\epsilon(t+1), X(t+h)\epsilon(t+h+1)\right] = 0$. *Hints:* Use the law of total expectation again (and the previous sub-problem).

(d) (10) Show that

$$\mathrm{Var}\left[M_n'(b_0)\right] = \frac{4}{n-1}\frac{\tau^4}{1-b_0^2}$$

Hint: Use the previous sub-problem.

(e) (8) Show that $\mathrm{Var}\left[\hat{b}_n\right] \approx \frac{1-b_0^2}{n-1}$ for large n.

4. (1 How much time did you spend on this problem set?

Rubric (10): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. All plots, tables, etc., are generated automatically by code embedded in the R Markdown file. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision. Code is properly integrated with a tool like R Markdown or knitr, and both the knitted file and the source file are submitted. The code is indented, commented, and uses meaningful names. All code is relevant, without dangling or useless commands. All parts of all problems are answered with coherent sentences, and raw computer code or output are only shown when explicitly asked for.