

ECON 710B - Problem Set 10

Alex von Hafften*

4/13/2021

Exercise 22.1

Take the model $Y = X'\theta + e$ where e is independent of X and has known density function $f(e)$ which is continuously differentiable.

- (a) Show that the conditional density of Y given $X = x$ is $f(y - x'\theta)$.

$$P(Y \leq y | X = x) = P(X'\theta + e \leq y | X = x) = P(x'\theta + e \leq y) = P(e \leq y - x'\theta) = F(y - x'\theta)$$

$$\implies F'(y - x'\theta) = f(y - x'\theta)$$

- (b) Find the functions $\rho_i(\theta)$ and $\psi_i(\theta)$.

Because conditional distribution of Y is known, we can use MLE to estimate θ :

$$\rho_i(\theta) = -\log f(y_i - x_i'\theta)$$

$$\implies \psi_i(\theta) = \frac{\partial}{\partial \theta} \rho_i(\theta) = -\frac{\partial}{\partial \theta} \log f(y_i - x_i'\theta) = \frac{f'(y_i - x_i'\theta)}{f(y_i - x_i'\theta)} x_i$$

- (c) Calculate the asymptotic covariance matrix.

Note that at the true value of θ ,

$$\psi_i(\theta) = \frac{f'(y_i - x_i'\theta)}{f(y_i - x_i'\theta)} x_i = \frac{f'(e_i)}{f(e_i)} x_i$$

Since we use correctly specified MLE (Hansen p. 762),

$$V = \Omega^{-1} = E[\psi_i(\theta)\psi_i(\theta)']^{-1} = E\left[\frac{f'(e_i)}{f(e_i)} x_i x_i' \frac{f'(e_i)}{f(e_i)}\right] = E\left[\left(\frac{f'(e_i)}{f(e_i)}\right)^2\right] E[x_i x_i']$$

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Sarah Bass, Emily Case, Danny Edgel, and Katherine Kwok.

Exercise 23.1

Take the model $Y = \exp(\theta) + e$ with $E[e] = 0$.

- (a) Is the conditional mean linear or nonlinear in θ ? Is this a nonlinear regression model?

The conditional mean nonlinear in θ : $E[Y|\theta] = E[\exp(\theta) + e|\theta] = \exp(\theta)$.

This is a nonlinear regression model.

- (b) Is there a way to estimate the model using linear methods? If so explain how to obtain an estimator $\hat{\theta}$ for θ .

Yes, define $\beta := \exp(\theta) \implies Y = \beta + e$. The conditional mean is linear in β : $E[Y|\beta] = E[\beta + e|\beta] = \beta$, so we can estimate $\hat{\beta}$ with OLS: $\hat{\beta} = \bar{Y}$. Then, $\hat{\theta} := \log(\hat{\beta}) = \log(\bar{Y})$.

- (c) Is your answer in part (b) the same as the NLLS estimator, or different?

They are the same. In the nonlinear regression model, $m(x, \theta) = E[Y|X = x] = \exp(\theta) \implies \bar{Y} = \exp(\hat{\theta}) \implies \hat{\theta} = \log(\bar{Y})$.

Exercise 23.2

Take the model $Y^{(\lambda)} = \beta_0 + \beta_1 X + e$ with $E[e|X] = 0$ where $Y^{(\lambda)}$ is the Box-Cox transformation of Y .

- (a) Is this a nonlinear regression model in the parameters $(\lambda, \beta_0, \beta_1)$? (Careful, this is tricky.)

No, the model is not a nonlinear regression model in parameters $(\lambda, \beta_0, \beta_1)$.

The model can be rewritten as:

$$\frac{Y^\lambda - 1}{\lambda} = \beta_0 + \beta_1 X + e \implies Y = [\lambda(\beta_0 + \beta_1 X + e) + 1]^{1/\lambda}$$

Thus, the conditional mean of Y is

$$E[Y|X = x] = \int [\lambda(\beta_0 + \beta_1 X + e) + 1]^{1/\lambda} f(e|x) de$$

Which is a function of the conditional density of e on x , which is not known.

Exercise 23.7

Suppose that $Y = m(X, \theta) + e$ with $E[e|X] = 0$, $\hat{\theta}$ is the NLLS estimator, and \hat{V} the estimator of $\text{var}(\hat{\theta})$. You are interested in the conditional mean function $E[Y|X = x] = m(x)$ at some x . Find an asymptotic 95% confidence interval for $m(x)$.

Under regularity assumptions, the 95% confidence interval for θ is $[\hat{\theta} - 1.96\sqrt{\hat{V}}, \hat{\theta} + 1.96\sqrt{\hat{V}}]$ because $\hat{\theta}$ is asymptotically normal. Thus, the asymptotic 95% confidence interval for $m(x)$ is $[m(X, \hat{\theta} - 1.96\sqrt{\hat{V}}), m(X, \hat{\theta} + 1.96\sqrt{\hat{V}})]$.

Exercise 23.8

The file PSS2017 contains a subset of the data from Papageorgiou, Saam, and Schulte (2017). For a robustness check they re-estimated their CES production function using approximated capital stocks rather than capacities as their input measures. Estimate the model (23.3) using this alternative measure. The variables for y_i , X_{1i} , and X_{2i} are EG_total, EC_c_alt, and EC_d_alt, respectively. Compare the estimates with those reported in Table 23.1.

```
PSS2017 <- read.xlsx("PSS2017.xlsx", sheetIndex = 1) %>%
  as_tibble() %>%
  mutate(y_ln = log(as.numeric(EG_total)),
         x_1 = as.numeric(EC_c),
         x_2 = as.numeric(EC_d),
         x_1_alt = as.numeric(EC_c_alt),
         x_2_alt = as.numeric(EC_d_alt)) %>%
  filter(!is.na(y_ln + x_1 + x_2 + x_1_alt + x_2_alt))

nls_main <- nls(formula = y_ln ~ beta +
                (nu/rho) * log(alpha * x_1^rho + (1-alpha) * x_2^rho),
                start = list(alpha = 0.39, beta = 1.66, nu = 1.05, rho = 0.36),
                data = PSS2017)

nls_robustness <- nls(formula = y_ln ~ beta +
                     (nu/rho) * log(alpha*x_1_alt ^ rho+(1-alpha) * x_2_alt^rho),
                     start = list(alpha = 0.39, beta = 1.66, nu = 1.05, rho = 0.36),
                     data = PSS2017)

summary(nls_main)

##
## Formula: y_ln ~ beta + (nu/rho) * log(alpha * x_1^rho + (1 - alpha) *
##      x_2^rho)
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## alpha 0.384314   0.011053  34.771 < 2e-16 ***
## beta  1.660470   0.080779  20.556 < 2e-16 ***
## nu    1.048639   0.008211 127.718 < 2e-16 ***
## rho   0.335022   0.060136   5.571 5.22e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2024 on 334 degrees of freedom
##
## Number of iterations to convergence: 4
## Achieved convergence tolerance: 3.675e-06

summary(nls_robustness)

##
## Formula: y_ln ~ beta + (nu/rho) * log(alpha * x_1_alt^rho + (1 - alpha) *
##      x_2_alt^rho)
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
```

```

## alpha    0.319427    0.011581   27.581   < 2e-16 ***
## beta   -12.582163    0.184574  -68.169   < 2e-16 ***
## nu      1.042580    0.007826  133.220   < 2e-16 ***
## rho     0.411489    0.058496    7.035  1.14e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1951 on 334 degrees of freedom
##
## Number of iterations to convergence: 5
## Achieved convergence tolerance: 8.251e-06

```

The estimates for α , ν , and ρ are about the same between the main results and the robustness check, but the estimates for β are very different. Since $\beta = \log A - e$, this difference is likely due to different scales for the alternate measures.

Exercise 24.3

Define $\psi(x) = \tau - \mathbb{1}\{x < 0\}$. Let θ satisfy $E[\psi(Y - \theta)] = 0$. Is θ a quantile of the distribution of Y ?
Yes,

$$E[\psi(Y - \theta)] = 0 \implies E[\tau - \mathbb{1}\{Y - \theta < 0\}] = 0 \implies E[\tau] = E[\mathbb{1}\{Y < \theta\}] \implies \tau = P(Y < \theta)$$

Exercise 24.4

Take the model $Y = X'\beta + e$ where the distribution of e given X is symmetric about zero.

(a) Find $E[Y|X]$ and $\text{med}[Y|X]$.

If the distribution of e given X is symmetric about zero, $E[Y|X] = E[X'\beta + e|X] = X'\beta$ and $\text{med}[Y|X] = \text{med}[X'\beta + e|X] = X'\beta$ because $P(e \leq 0|X) = 0.5 \implies \text{med}[e|X] = 0$

(b) Do OLS and LAD estimate the same coefficient β or different coefficients?

Different, although with symmetric errors the median and mean are equal, OLS and LAD estimated from a finite sample will differ. For example, the OLS estimate minimizes the square of the residual, so it will be more influenced by particularly large or small realizations of the error term.

(c) Under which circumstances would you prefer LAD over OLS? Under which circumstances would you prefer OLS over LAD? Explain.

LAD estimates the conditional median function and OLS estimates the conditional mean function. When the error term is symmetrically distributed, the mean and the median are equal, but generally you should use the estimate that is estimating what you want to be estimating. The median is less sensitive to influential observations and provides a better measure of the “typical” observation when a RV is asymmetrically distributed (i.e., the typical income level is measured by the median not the mean).

Exercise 24.5

You are interested in estimating the equation $Y = X'\beta + e$. You believe the regressors are exogenous, but you are uncertain about the properties of the error. You estimate the equation both by least absolute deviations (LAD) and OLS. A colleague suggests that you should prefer the OLS estimate, because it produces a higher R^2 than the LAD estimate. Is your colleague correct?

First, observe that $R_{LAD}^2 \leq R_{OLS}^2$. Since $\hat{\beta}_{OLS}$ minimizes the sum of square deviation by definition:

$$\begin{aligned} \sum_i (Y_i - X_i' \hat{\beta}_{LAD})^2 &\geq \sum_i (Y_i - X_i' \hat{\beta}_{OLS})^2 \\ \implies R_{LAD}^2 &= 1 - \frac{\sum_i (Y_i - X_i' \hat{\beta}_{LAD})^2}{\sum_i (Y_i - \bar{Y})^2} \leq 1 - \frac{\sum_i (Y_i - X_i' \hat{\beta}_{OLS})^2}{\sum_i (Y_i - \bar{Y})^2} = R_{OLS}^2 \end{aligned}$$

Second, a lower R^2 does not necessarily mean that an estimation approach is preferred. OLS estimates the conditional mean function and LAD estimates the conditional median function. If you want to conditional median function based on its advantages over conditional mean function (i.e., less sensitivity to influential observations), then LAD is more appropriate.

Exercise 24.14

Using the `cps09mar` dataset take the sample of Hispanic women with education 11 years or higher. Estimate linear quantile regression functions for log wages on education. Interpret.

```
cps09mar <- read_delim(file = "cps09mar.txt",
  delim = "\t",
  col_names = c("age", "female", "hisp", "education", "earnings", "hours",
    "week", "union", "uncov", "region", "race", "marital"),
  col_types = cols()) %>%
  mutate(l_wage = log(earnings/hours/week))

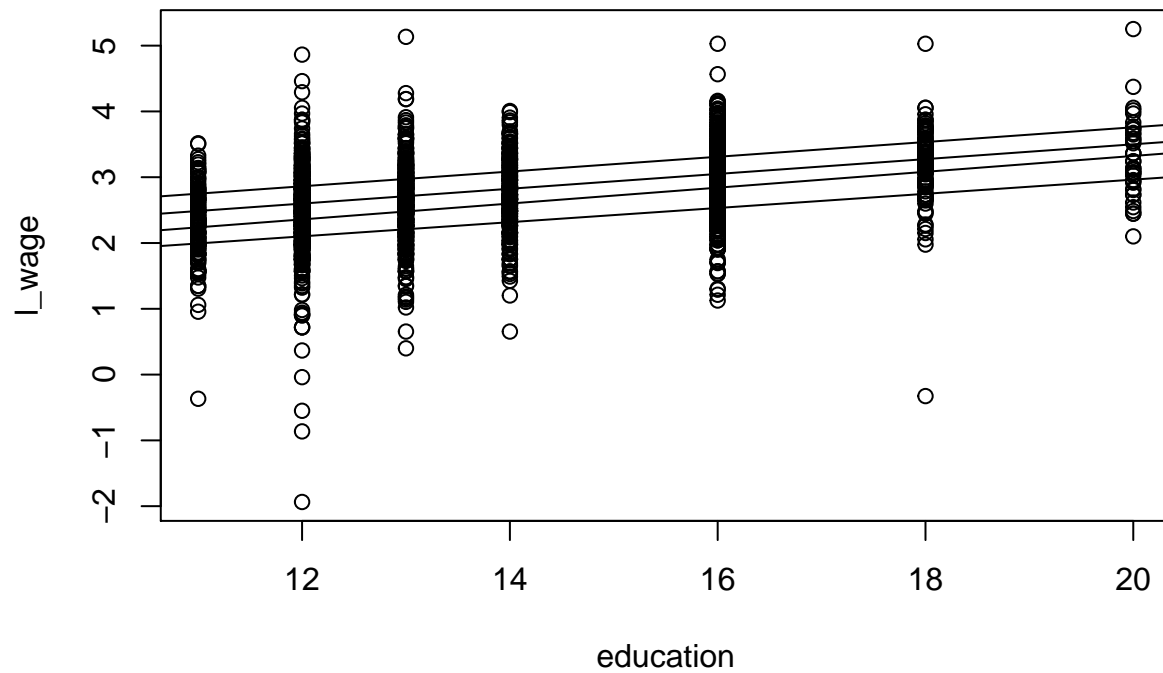
sample <- cps09mar %>%
  filter(hisp == 1, female == 1, education >= 11)

quant_reg <- rq(formula = l_wage ~ education, tau = 1:4/5, data = sample)

summary(quant_reg)
```

```
##
## Call: rq(formula = l_wage ~ education, tau = 1:4/5, data = sample)
##
## tau: [1] 0.2
##
## Coefficients:
##              Value      Std. Error t value Pr(>|t|)
## (Intercept)  0.80707   0.09748     8.27960  0.00000
## education    0.10781   0.00728    14.81870  0.00000
##
## Call: rq(formula = l_wage ~ education, tau = 1:4/5, data = sample)
##
## tau: [1] 0.4
##
## Coefficients:
##              Value      Std. Error t value Pr(>|t|)
## (Intercept)  0.91170   0.08327    10.94845  0.00000
## education    0.12058   0.00600    20.11279  0.00000
##
## Call: rq(formula = l_wage ~ education, tau = 1:4/5, data = sample)
##
## tau: [1] 0.6
##
## Coefficients:
##              Value      Std. Error t value Pr(>|t|)
## (Intercept)  1.24750   0.08154    15.29888  0.00000
## education    0.11264   0.00593    18.97964  0.00000
##
## Call: rq(formula = l_wage ~ education, tau = 1:4/5, data = sample)
##
## tau: [1] 0.8
##
## Coefficients:
##              Value      Std. Error t value Pr(>|t|)
## (Intercept)  1.51675   0.08832    17.17245  0.00000
## education    0.11212   0.00667    16.79944  0.00000
```

```
plot(l_wage ~ education, data = sample)
for (j in 1:ncol(quant_reg$coefficients)) abline(coef(quant_reg)[, j])
```



I estimate quantile regression function for the inner ventiles (i.e., 0.2, 0.4, 0.6, 0.8). Positive slopes across all ventiles suggest that higher education and higher wages are correlated across the wage distribution. If we assumed that someone would stay in the same quantile as they got more education, it would suggest a similar positive return to education across the wage distribution. Since the quantile functions are roughly parallel, it suggests homoskedastic errors.