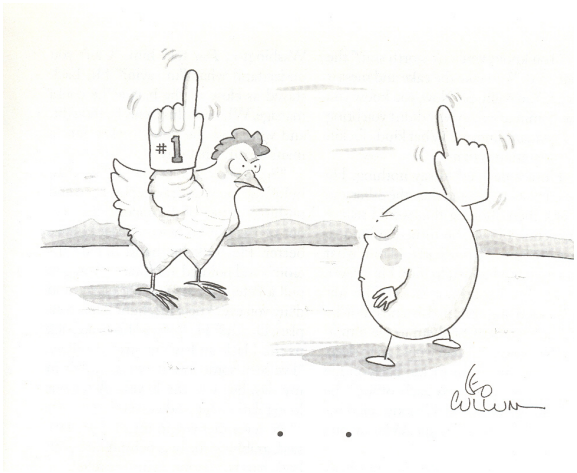
University of Illinois
Fall 2006

Department of Economics
Professor R. Koenker

ECON 508
Final Exam Review Questions



1. In the seminal paper of Thurman and Fisher (1988) evidence is presented to resolve the long-standing conundrum: Which came first, the chicken or the egg?

   (a) Explain briefly the conclusion of the paper and the nature of the empirical evidence used to support it.

   (b) Do the statistical methods used by Thurman and Fisher constitute a reasonable method for assessing causation? Why and/or why not?

2. Consider a fictional, balanced panel data model for household demand for gasoline,

$$y_{it} = \alpha_i + x_{it}\beta + z_i'\gamma + u_{it}$$

   where $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and

$$\begin{aligned} y_{it} &= \text{log of household demand for gasoline in gallons/month} \\ x_{it} &= \text{log of average price of gasoline paid by household } i \text{ in month } t \\ z_i &= \text{a vector of time invariant household characteristics including} \\ &\quad \text{income, family structure, etc.} \end{aligned}$$

   A serious potential problem with the model is that the price variable, x_{it}, may be correlated with the individual specific effect α_i. This problem arises because the price paid by household i must

be inferred by dividing recorded expenditure on gasoline by the number of gallons purchased – since some households may search more intensively than others for a lower price, some of the observed price variation may be due to this endogenous "shopping-effect" rather than purely exogenous market variability. In particular one might expect that this endogeneity would be correlated with the household specific demand effect α_i.

(a) Explain briefly why OLS is an unsatisfactory method of estimating this model.

(b) Since the primary objective in estimating this model is to recover an accurate estimate of the price elasticity, β, suggest a way to accomplish this which avoids the problems alluded to above regarding the endogeneity of the price variable.

(c) Suppose you now estimate the model by the technique recommended in part (b.) and obtain $\hat{\beta} = -.70$ with a standard error of .08. Then, ignoring the endogeneity effect, you also estimate the model using the random effects estimator, i.e., treating the α_i's as a random sample with mean α_0 and variance σ_α^2. From this you obtain $\hat{\beta} = -.40$ with a standard error of .06. Someone suggests using these results to test for bias due to the endogeneity. Explain the test briefly and carry it out.

(d) How do the conclusions drawn in part (c.) affect your ability to estimate the parameter vector γ. Explain briefly how the availability of a new time varying covariate, say, household income, would affect your estimation strategy.

3. You have estimated the logit model

$$\text{logit}(p_i) = -4.5 + 1.7x_i - .25x_i^2$$

where p_i is the probability that a paper submitted to the *Phuzics Review* is accepted for publication and x_i is the natural logarithm of the length of the paper in pages.

(a) If you would like to maximize the probability of acceptance, how long should your paper be?

(b) By how much do you change the probability of acceptance when you cut the length of a 50 page paper to 40 pages, assuming the content is undamaged?

4. Consider the linear model

$$y = X_1\beta_1 + X_2\beta_2 + u$$

where u is assumed to be iid $\mathcal{N}(0, \sigma^2)$. An investigator suggests using the test statistic

$$T = (\hat{\beta}_1 - \tilde{\beta}_1)^1 V^{-1} (\hat{\beta}_1 - \tilde{\beta}_1)$$

where

$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 y$$

and

$$\tilde{\beta}_1 = (X_1' X_1)^{-1} X_1' y$$

(a) Explain *carefully*, what is the null hypothesis for which this statistic would provide an appropriate test?

(b) What is the alternative hypothesis?

(c) What is the matrix V? Be as specific as possible here.

(d) What is the matrix M_2?

5. Consider the Cox proportional hazard survival model with hazard function

$$\lambda(t|x) = \lambda_0(t)e^{x'\beta}$$

where $\lambda_0(t)$ is the " baseline hazard", i.e. the hazard function for $x'\beta = 0$.

(a) Show that the probability of survival until time t given covariates, x is given by

$$S(t|x) = S_0^{\gamma(x)}(t)$$

where $\gamma(x) = exp(x'\beta)$ and $S_0(t) = exp(-\int_0^t \lambda_0(s)ds)$

(b) Suppose you have estimated such a model of unemployment duration and that for *otherwise identical* men and women the estimated probability of an unemployment spell greater than one year is $\frac{1}{2}$ for men and $\frac{1}{3}$ for women. What is the coefficient on the gender dummy variable assuming men 1 and women 0 in the estimated Cox model?

6. Consider a model for a consumer durable good in which the willingness to pay for one unit of the good for individual i is given by

$$w_i = x_i'\beta + u_i$$

where x_i denote a vector of individual observable characteristics. Suppose that individuals buy the good if their willingness to pay, w_i, exceeds the price that they encounter, v_i.

(a) Given a sample of individuals facing widely varying prices, v_i, i.e. given data on $(y_i, x_i, v_i : i = 1, \ldots, n)$ where $y_i = 1$ indicates purchase, and $y_i = 0$ nonpurchase. Assuming (as usual) that the u_i's are iid normal, suggest a method for estimating the parameter β.

(b) Explain the following statement: "For persons with characteristics x we can interpret the survival function $S(v|x) = P(w > v|x)$ as a demand function for consumers of type x."

7. Consider a life cycle productivity model like the one estimated for in PS4 phuzicists in which you decide to adopt a more flexible model for the "experience" effect. You estimate the model

$$y_{it} = \alpha_1 y_{it-1} + \alpha_2 y_{it-2} + \sum_{k=0}^{p} \varphi_k(x_{it})\beta_k + z_i^\top \gamma + \alpha_i + u_{it}$$

where y_{it} is productivity of individual i in year t, x_{it} is experience of i in year t, the vector z_i consists of time invariant characteristics of individual i, and the α_i are individual specific effects.

(a) Suppose you adopt the linear spline model with

$$\varphi_k(x) = \max\{0, x - \tau_k\}$$

where $p = 3, \tau_0 = 0, \tau_1 = 6, \tau_2 = 12, \tau_3 = 24$. Given an estimated version of the model explain how you would test that productivity declines after 24 years of experience.

(b) There are good reasons to suppose that the α_i's are correlated with lagged productivity and some of the z_i variables, so you decide to adopt a fixed effect approach to estimation. Discuss the advantages and disadvantages of this decision.