Fall 2012 R. Koenker

Economics 508
Midterm Exam Review

This is a merged and revised version of several recent midterm exams. The typical instructions are reproduced below. You can expect about 4 questions on the real midterm. You should bring a calculator in case there are questions that require some computation. Since the midterm will be somewhat earlier than usual this year, some of these questions may refer to material that we will cover after the midterm and therefore would not be appropriate for the exam.

Please answer all questions. Even if you are unsure about some aspect of the questions, try to write something sensible – partial credit *will* be given. The questions will be weighted equally. The exam is closed book, closed notes and will last 2 hours.

(1) You have just been hired by Starbuck's to forecast world coffee prices. Your predecessor has left a short paper describing his methods, which boil down to the estimated least squares equation

$$\log p_t = \underset{(0.23)}{-3.65} - \underset{(0.12)}{1.22} \log q_t + \underset{(0.42)}{2.86} \log x_t$$

where q_t, p_t and x_t are quantity, price and per capita income in billions of tons, dollars per pound, and 1000's of dollars per year, respectively. (Standard errors in parentheses.)

You feel that this is too simplistic, so you reestimate a more sophisticated model and obtain,

$$\log p_t = \underset{(0.18)}{-1.54} + \underset{(0.12)}{0.50} \log p_{t-1} + \underset{(0.07)}{0.28} \log p_{t-2} - \underset{(0.04)}{0.12} \log q_t - \underset{(0.02)}{0.05} (\log q_t)^2 + \underset{(0.31)}{1.04} \log x_t$$

It is presumed that world coffee production is determined primarily by weather conditions in growing areas, so it is reasonable to treat supply as exogonous and view the foregoing models as proper demand equations, but this year there has been talk about the possibility that the Organization of Coffee Exporting Nations will conspire to reduce world output by 5% from its current level of 2.718 trillion tons.

(a) Evaluate the stability of the dynamics of your estimated model.
(b) Compare the price predictions from the two models (from a long-run equilibrium viewpoint) based on the conjectured 5% reduction of supply.

(2) In his canonical work on cointegration in multivariate (VAR) settings Soren Johansen focused on the error correction model

$$\Delta y_t = \Pi y_{t-1} + B_1 \Delta y_{t-1} + \cdots + B_p \Delta y_{t-p} + \varepsilon_t \tag{1}$$

where y_t is an m-vector.

(a) Suppose, for a moment, that $m = 1$, so that Π and the B's are all scalar, interpret the hypothesis $\Pi = 0$ and explain *briefly* how to test it.
(b) Now in the general case $m \geq 2$ explain why Johansen's test relies on the estimated auxiliary VAR equations:

$$\Delta y_t = \hat{\delta}_0 + \hat{\delta}_1 \Delta y_{t-1} + \cdots + \hat{\delta}_p \Delta y_{t-p} + \hat{u}_t \tag{2a}$$

$$y_{t-1} = \hat{\theta}_0 + \hat{\theta}_1 \Delta y_{t-1} + \cdots + \hat{\theta}_p \Delta y_{t-p} + \hat{v}_t \tag{2b}$$

In particular how does estimating these equations relate to the estimation of (1)?

(c) Give a brief heuristic explanation of what it means that the coordinates of y_t are *cointegrated* and how this relates to the rank of the matrix Π. (Avoid discussion of any gory details of the Johansen test statistic.)

(3) Suppose that we have a structural equation

$$y = Y\gamma + X\beta + u$$

and we are willing to assume that $X \perp u$ but *not* $Y \perp u$. But we have instrumental variables $Z \perp u$ that might save the day. We proceed in the following somewhat unorthodox way:

Step 1 We multiply by $M_X = I - X(X'X)^{-1}X'$ to obtain

$$M_X y = M_X Y\gamma + M_X X\beta + M_X u$$

which we write as

$$\tilde{y} = \tilde{Y}\gamma + \tilde{u}$$

Step 2 Now, even though we know that Z doesn't belong in our structural equation and that Y does belong, but is endogenous, we consider estimating the linear model by least squares

$$\tilde{y} = \tilde{Y}\gamma + Z\delta + v$$

in particular, we consider a family of estimaters of δ as a function of γ,

$$\hat{\delta}(\gamma) = (Z'Z)^{-1}Z'(\tilde{y} - \tilde{Y}\gamma)$$

Step 3 Now since we are really interested (mainly) in γ we decide to estimate γ by minimizing the norm of $\hat{\delta}(\gamma)$ that is by solving

$$\hat{\gamma}_A = \arg\min_{\gamma} \parallel \hat{\delta}(\gamma) \parallel_A^2$$

where $\parallel X \parallel_A = x'Ax$.

(a) Explain Step 1. What happened to β in the model for $\tilde{y}$?

(b) Given that you are free to choose A in any way you would like, try to provide a justification for Steps 2 and 3. Be sure to be explicit about any reservations you have about this procedure, and suggest alternatives that would perform better if you can.

(4) Suppose that

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

with $Euu' = \sigma^2 I$, and let

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

and

$$\hat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2y$$

(a) Explain *in words* the difference between these two estimators.

(b) Define the matrix M_2.

(c) It is claimed that

$$\hat{\beta}_1 = \tilde{\beta}_1 - \Gamma\hat{\beta}_2.$$

Derive this expression providing in the process definitions of $\hat{\beta}_2$ and Γ.

(d) It is claimed that

$$V(\hat{\beta}_1) = V(\tilde{\beta}_1) + \sigma^2\Gamma(X_2'M_1X_2)^{-1}\Gamma'.$$

Interpret this expression and explain briefly "what happened to the covariance term?"

(e) The foregoing results establish a fundamental tradeoff in the consideration of omitted variables between bias and variance effects. Explain briefly the nature of this tradeoff and contrast it briefly with the model selection rules proposed by Akaike and Schwarz in the case that β_2 is a scalar.

(5) In Figures 2a-f we illustrate six scatterplots corresponding to possible models represented algebraically below. Try to match the figures and the equations; if there isn't a match try to explain why you don't think so.

$$\log y_i = \underset{(0.04)}{-2.00} + \underset{(0.026)}{2.01} \log x_i \tag{2.1}$$

$$y_i = \underset{(.056)}{5.03} - \underset{(.0.10)}{2.03} / \sqrt{x_i} \tag{2.2}$$

$$y_i = \underset{(1.03)}{49.92} - \underset{(0.13)}{1.91} x_i \tag{2.3}$$

$$1/y_i = \underset{(.030)}{1.047} - \underset{(.004)}{0.051} x_i \tag{2.4}$$

$$y_i = \underset{(.51)}{49.5} + \underset{(.065)}{1.95} x_i \tag{2.5}$$

$$y_i = \underset{(.013)}{2.05} - \underset{(.004)}{.060} x_i + \underset{(.0002)}{.004} x_i^2 \tag{2.6}$$

(6) Suppose you have estimated a cost function for a sample of international steel manufacturers

$$\log c_i = \alpha + \beta \log x_i + \gamma (\log x_i)^2 + u_i$$

where c_i denotes total annual cost of firm i and x_i denotes annual output in metric tons.

(a) Explain why optimal scale (minimum average cost), if it exists, occurs at the output level where the cost elasticity is unity, i.e., $\eta = \partial \log c / \partial \log x = 1$,

(b) Having estimated this model explain how you would make a point estimate of optimal scale. State clearly any necessary caveats.

(7) The Schwarz information criterion for model selection chooses the model that maximizes

$$\text{SIC}_j = \ell_j(\hat{\theta}) - \frac{1}{2} p_j \log n$$

(a) Explain briefly what each of the pieces mean: $\ell_j(\hat{\theta}), p_j, n$.

(b) Explain why such a criterion might be preferred to finding the model that maximized the likelihood function as a device for model selection.

(c) Explain how the SIC criterion is related to the classical theory of hypothesis testing. (Recall that under the null hypothesis that model j is correct, then for any model k within which model j is nested, i.e., $\Theta_k \supseteq \Theta_j$, we have

$$2(\ell_k(\hat{\theta}) - \ell_j(\hat{\theta})) \sim \chi^2_q$$

where $q = p_k - p_j$. In particular, relate SIC to the conventional t-test when $q = 1$, so the model dimensions only differ by one.

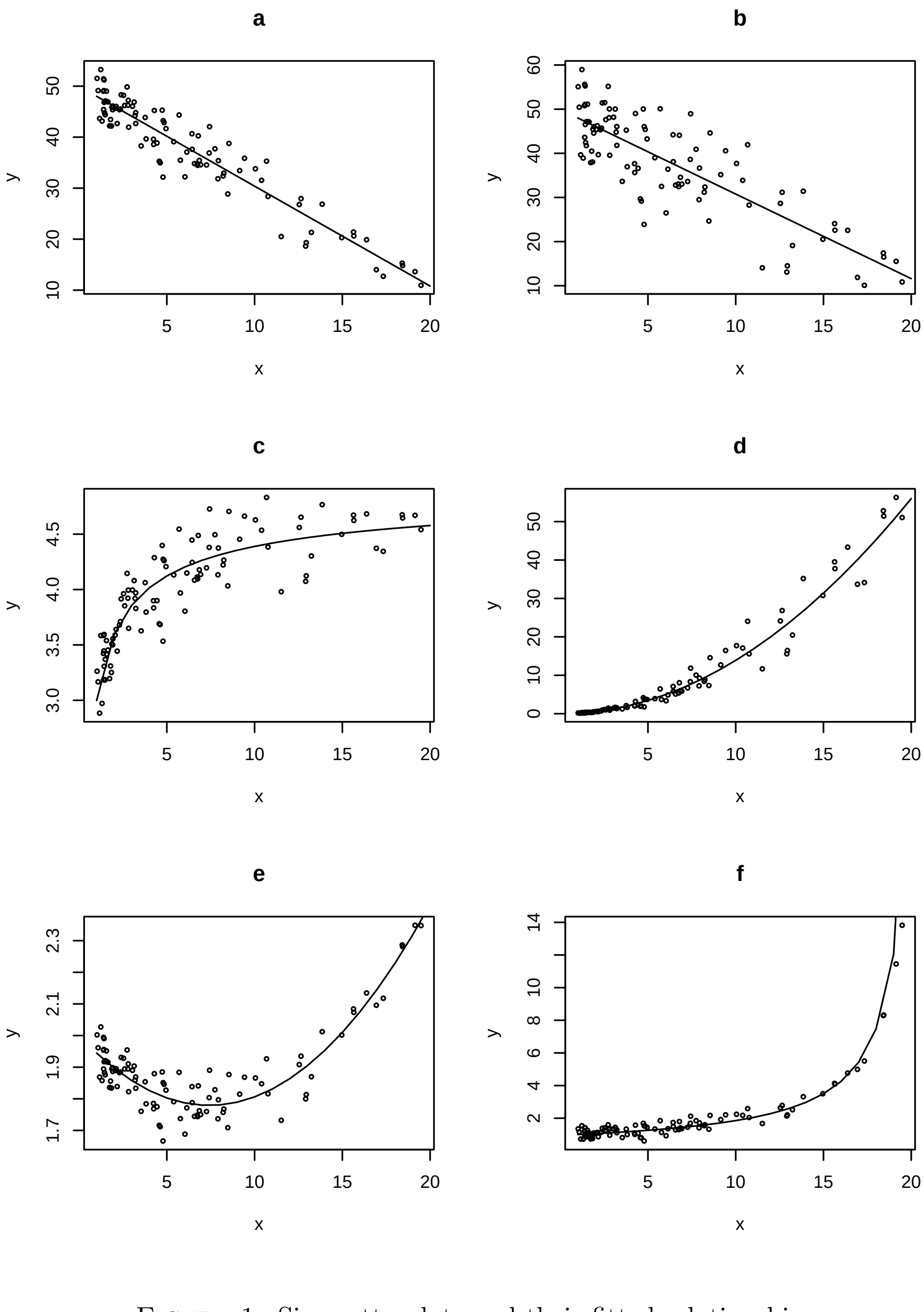

FIGURE 1. Six scatterplots and their fitted relationship.

(8) Given the simple cobweb model of supply and demand

$$\begin{aligned} &\text{(Supply)} \qquad & Q_t = \alpha_1 + \alpha_2 p_{t-1} + \alpha_3 z_t + u_t \\ &\text{(Demand)} \qquad & p_t = \beta_1 + \beta_2 Q_1 + \beta_3 w_t + v_t \end{aligned}$$

Suppose that z_t and w_t are exogenous in the sense that $E(z_t, w_t)'(u_t, v_t) = 0$.

(a) Supposing all the variables are in logs, find the long-run supply elasticities with respect to changes in z_t and w_t.

(b) Explain briefly how to use the δ method to compute a standard error for the elasticities found in part a.)

(9) A standard formula for the least squares estimator of scalar parameter β in the simple bivariate regression model

$$y_i = \alpha + \beta x_i + u_i$$

is

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Explain this formula using the language and algebra of the partial residual plot.

(10) The Vichy bottled water company has estimated the following log quadratic demand curve

$$\log q_t = \hat{\alpha} + \hat{\beta} \log p_t + \hat{\gamma} (\log p_t)^2 + \hat{\rho} \log q_{t-1}$$

(a) Since costs are negligible, the firm wants to maximize *revenue*. Explain what you would do to find the revenue maximizing price, explaining also any caveats along the way.

(b) Not satisfied with a point estimate for this revenue maximizing price, p^*, suppose the Board of Directors wants a 95% confidence interval for p^*. Describe how you would compute the confidence interval, using the δ-method?

(11) In the simplest treatment effect model we have a binary treatment variable, say x_i, that takes the value one if the subject is treated, and takes the value zero if the subject is a control. Often the response variable $y_i \equiv \Delta Y_i$ measures the change in something: wages, health status, etc. over the study period.

(a) Show that the least squares estimator of the treatment effect in the model

$$y_i = \alpha + \beta x_i + u_i \tag{1.1}$$

is simply $\hat{\beta} = \bar{y}_1 - \bar{y}_0$ the difference in mean differences for the two samples. You may find it more convenient to work with the reparameterized model

$$y_i = \gamma x_i + \delta(1 - x_i) + u_i \tag{1.2}$$

If so, justify the link back to the original parameterization in (1.1).

(b) Generalize this diff-in-diff interpretation of $\hat{\beta}$ to the case that the model includes other covariates,

$$y_i = \alpha + \beta x_i + z_i^\top \xi + u_i \tag{1.3}$$

in particular, use your interpretation to explain what would happen if treatment assignment, x_i, were exactly predictable given the other covariates, z_i for all of the observations in your sample, i.e. $x_i = z_i^\top \eta$ for some vector η.

(c) Explain briefly the critical assumption in justifying the least squares estimator in (1.1) that broke down in the Lanarckshire milk experiment, and why randomized treatment assignment would have been preferable from the point of view of accurately estimating the treatment effect.

(12) Suppose you are interested in estimating the causal effect of military experience on subsequent earnings. You have a large sample, but very limited information on each individual, leading you to consider the following model,

$$y_i = \beta D_i + \sum_{j=1}^{J} F_{ij}\gamma_j + u_i,$$

where y_i is log earnings, $D_i = 1$ if individual i has served in the military and $D_i = 0$ otherwise, and the F_{ij}'s are occupational dummies, so $F_{ij} = 1$ if individual i is in occupation j and $F_{ij} = 0$ otherwise. Each individual in the sample is in one, and only one, of the J occupations.

(a) Let $P_F = F(F^\top F)^{-1}F^\top$ be the projection onto the space spanned by the occupational dummies, and $M_F = I - P_F$. Explain the difference between the two estimators of β:

$$\hat{\beta} = (D^\top M_F D)^{-1} D^\top M_F y$$

and

$$\tilde{\beta} = (D^\top P_F D)^{-1} D^\top P_F y$$

Discuss briefly circumstances under which each of these estimators would be a desirable, that is to say consistent, estimator of the causal effect.

(b) Explain what the matrix M_F does to the vectors y and D. Be as explicit as possible. Suppose that we write, $\tilde{y} = M_F y$ and $\tilde{D} = M_F D$ and we now consider "running" the regression,

$$\tilde{y}_i = \alpha_0 + \alpha_1 \tilde{D}_i + v_i$$

What can you say about the resulting least squares estimators of α_0 and α_1. Try to provide, at least, some heuristic description of what values they will take.

(c) Suppose we now plot $\hat{\alpha} = (F^\top F)^{-1}F^\top y$ versus $\hat{\delta} = (F^\top F)^{-1}F^\top D$, interpret this scatter plot and suggest a way to compute the coefficients of a line through the scatter plot that would recover $\tilde{\beta}$ from part (a.).

(13) A recent study of portfolio managers finds that "high-flying" managers who performed in the top 20 percent over the period 2000-2002 had much more modest performance in the period 2008-2010, averaging only slightly above the overall mean performance of all managers for the latter period.

(a) Explain briefly the above phenomenon employing the simple model,

$$y_{it} = \alpha_i + \beta y_{i,t-1} + u_{it}$$

(b) For a manager with $\hat{\alpha}_i = 2$ and $\hat{\beta} = 0.6$, with $\mathbb{E}u_{it}^2$ estimated to be $\hat{\sigma}^2 = 2$, what would be his estimated long-run mean performance according to the model? Construct a 95 percent confidence interval for this expected performance.

(c) Suppose that the α_i were found to be approximately normally distributed with mean two and variance one, what is the long run mean and variance for the performance of all managers.

(d) If you suspected that the u_{it} were autocorrelated of order one, how would this influence your estimation strategy for the model in part (a.)?