Empirical Industrial Organization (ECO 310)
University of Toronto. Department of Economics
Fall 2018. Instructor: Victor Aguirregabiria

Solution to Midterm Exam
October 29th, 2018

MIDTERM EXAM. Monday, October 29th, 2018. From 12pm to 1pm (1 hour)

INSTRUCTIONS:
- **This is a closed-book exam.**
- **No study aids, including calculators, are allowed.**
- **Please, answer all the questions.**

TOTAL POINTS = 100

QUESTION 1 [50 points]. Consider a firm panel dataset with the following variables: id (firm id number); year (year); logy (logarithm of firm's output); logn (logarithm of firm's labor); logk (logarithm of firm's capital). Using this dataset and Stata software package, a researcher runs the following commands:

xtset id year

xtreg logy logn logk i.year, fe

test logn + logk = 1

And obtains the results that you can see in the next page.

Based on these results, answer the following questions.

Q1.1. [10 points] Write the regression equation that the command "xtreg logy logn logk i.year, fe" estimates.

ANSWER: The original model is:

$$y_{it} = \alpha_L \; \ell_{it} + \alpha_K \; k_{it} + \gamma_t + \eta_i + u_{it}$$

where y_{it}, ℓ_{it}, and k_{it} are the logarithm of output, the logarithm of labor, and the logarithm of capital, respectively. This command implements a Fixed Effects estimator of this model, where the fixed effects are the time-invariant unobserved firm effects η_i. The model also includes year dummies to capture the time effects γ_t.

Q1.2. [10 points] Explain the estimator that this command implements.

ANSWER: The Fixed Effects estimator first transforms the model in deviations with respect to the firm-specific means of the variables. The transformed model is:

$$y_{it} - \overline{y}_i = \alpha_L \ (\ell_{it} - \overline{\ell}_i) + \alpha_K \ (k_{it} - \overline{k}_i) + \gamma_t^* + (u_{it} - \overline{u}_i)$$

where $\overline{y}_i$, $\overline{\ell}_i$, and $\overline{k}_i$ are the sample means of log-output, log-labor, and log-capital, respectively, for firm i. Then, the Fixed Effects estimator is the OLS estimator applied to this transformed model. That is, an OLS estimator where the regression model where the dependent variable is $(y_{it} - \overline{y}_i)$ and the regressors are $(\ell_{it} - \overline{\ell}_i)$, $(k_{it} - \overline{k}_i)$, and time dummies.

Q1.3. [10 points] Under what assumptions (on the error term and the regressors) does this estimator provide unbiased (consistent) estimates of the true parameters in the production function?

ANSWER: This estimator is consistent (asymptotically unbiased) if and only if the regressors in the transformed equation, $(\ell_{it} - \overline{\ell}_i)$ and $(k_{it} - \overline{k}_i)$, are not correlated with the error term in the transformed equation, $(u_{it} - \overline{u}_i)$. This is the case if the transitory shock u at any period t is not correlated with labor and capital inputs at any period t', for any value of t and t'.

Q1.4. [10 points] Explain the F-test at the bottom of the output table from the xtreg command. Interpret its result.

ANSWER: This is a F-test for the null hypothesis that all the fixed effects are the same (and they are zero, because there is a common intercept) such that there is not time-invariant unobserved heterogeneity: $H_0 : \eta_1 = \eta_2 = ... = \eta_N = 0$. For these data, the p-value of the test is practically equal to zero. It is smaller than the significance levels of 1%, 5%, or 10% that we typically use to reject or not a null hypothesis. Therefore, we reject the null hypothesis and conclude that there is time-invariant unobserved firm heterogeneity in the production function.

Q1.5. [10 points] Explain the F-test from the command test logn + logk = 1. Interpret its result.

ANSWER: This is a F-test for the null hypothesis that the sum of the regression coefficients for log-labor and log-capital are equal to one: $H_0 : \alpha_L + \alpha_K = 1$. This is the null hypothesis of Constant Returns to Scale (CRS) in the production function. For these data, the p-value of the test is practically equal to zero. It is smaller than the significance levels of 1%, 5%, or 10% that we typically use to reject or not a null hypothesis. Therefore, we reject the null hypothesis and conclude that there is not CRS. There is evidence of decreasing returns to scale because $\widehat{\alpha}_L + \widehat{\alpha}_K < 1$.

```
. // ----------------------------------------------
. // 4. Question 2.2: Fixed Effects estimation
. // ----------------------------------------------
. xtreg logy logn logk i.year, fe

Fixed-effects (within) regression               Number of obs     =      4,072
Group variable: id                              Number of groups  =        509

R-sq:                                           Obs per group:
     within  = 0.7379                                         min =          8
     between = 0.9706                                         avg =        8.0
     overall = 0.9661                                         max =          8

                                                F(9,3554)         =    1111.47
corr(u_i, Xb)  = 0.5988                         Prob > F          =     0.0000

------------------------------------------------------------------------------
        logy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        logn |   .6544609   .0144048    45.43   0.000     .6262184    .6827034
        logk |   .2329072    .013637    17.08   0.000     .2061702    .2596443
             |
        year |
       1983  |  -.0376406   .0093042    -4.05   0.000    -.0558828   -.0193985
       1984  |  -.0076445   .0096071    -0.80   0.426    -.0264805    .0111914
       1985  |  -.0234513   .0100955    -2.32   0.020    -.0432449   -.0036578
       1986  |  -.0136103   .0105543    -1.29   0.197    -.0343034    .0070829
       1987  |   .0314121   .0108748     2.89   0.004     .0100907    .0527335
       1988  |   .0753576   .0111072     6.78   0.000     .0535805    .0971347
       1989  |   .0764164   .0118166     6.47   0.000     .0532485    .0995844
             |
       _cons |   3.863804   .0529288    73.00   0.000      3.76003    3.967578
-------------+----------------------------------------------------------------
     sigma_u |  .42922318
     sigma_e |  .14715329
         rho |  .89482518   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(508, 3554) = 38.90                  Prob > F = 0.0000

. test logn + logk = 1

 ( 1)  logn + logk = 1

       F(  1,  3554) =  121.32
            Prob > F =    0.0000
```

QUESTION 2 [50 points]. Consider the consumer utility function:

$$U = c \ (q_1 - \gamma_1)^{\alpha_1} \ (q_2 - \gamma_2)^{\alpha_2} ... \ (q_J - \gamma_J)^{\alpha_J}$$

where α's and γ's are positive parameters, and c is the composite or outside good. In this model, the marginal utility of product j has the form $U_j = \alpha_j \dfrac{U}{q_j - \gamma_j}$, and similarly, the marginal utility of the composite good is $U_c = \dfrac{U}{c}$. The consumer budget constraint is:

$$c + p_1 q_1 + p_2 q_2 + ... + p_J q_J \leq y$$

Q2.1. [5 points] Write the consumer's decision problem. Represent it as a Lagrange optimization problem.

ANSWER: The consumer decision problem is:

$$\begin{array}{ll} \max_{c, q_1, ..., q_J} & c \ (q_1 - \gamma_1)^{\alpha_1} ... \ (q_J - \gamma_J)^{\alpha_J} \\ \text{subject to:} & c + p_1 q_1 + ... + p_J q_J \leq y \end{array}$$

We can represent this decision problem using the Lagrange representation:

$$\max_{c, q_1, ..., q_J, \lambda} \quad c \ (q_1 - \gamma_1)^{\alpha_1} \ ... \ (q_J - \gamma_J)^{\alpha_J} + \lambda \left[y - c - p_1 q_1 - ... - p_J q_J \right]$$

where λ is the Lagrange multiplier of the budget constraint.

Q2.2. [5 points] In this optimization problem, obtain the marginal conditions of optimality: with respect to product j; with respect to the composite good c; and with respect to the Lagrange multiplier λ.

ANSWER: Taking into account that the marginal utilities are $U_j = \alpha_j \dfrac{U}{q_j - \gamma_j}$ and $U_c = \dfrac{U}{c}$, we have that the marginal conditions of optimality are:

$$\begin{aligned} \alpha_j \frac{U}{q_j - \gamma_j} - \lambda \ p_j &= 0 \\ \frac{U}{c} - \lambda &= 0 \\ y - c - p_1 q_1 - ... - p_J q_J &= 0 \end{aligned}$$

Q2.3. [10 points] Combine the marginal conditions with respect to q_j and with respect to c to obtain a linear equation that relates q_j with c/p_j.

ANSWER: The marginal condition with respect to j implies that:

$$\alpha_j \frac{U}{\lambda} = p_j \ (q_j - \gamma_j)$$

And the marginal condition with respect to c implies that:

$$\frac{U}{\lambda} = c$$

Plugging the last equation into the former, we have:

$$\alpha_j \ c = p_j \ (q_j - \gamma_j)$$

Solving for q_j, we get the following linear relationship between q_j and c/p_j.

$$q_j = \gamma_j + \alpha_j \ \frac{c}{p_j}$$

Q2.4. [10 points] Combine the J linear equations in Q2.3 and the budget constraint to obtain the optimal amount of the composite good, c, as a function of prices, income, and utility parameters. [Hint: use the following steps (i) multiply by the price p_j the linear equation for product j in Q2.3; (ii) sum this equation over all the J products; (iii) taking into account that the budget constraint implies that $\sum_{j=1}^{J} p_j q_j = y - c$, combine this budget constraint with the equation in (ii) to obtain an equation where the only endogenous variable is c; and (iv) using this equation, solve for c to get an expression for this variable in terms on prices, income, and utility parameters].

ANSWER: We follow the four steps in the Hint. First, we multiply by price p_j the linear equation for product j in Q2.3:

$$p_j \ q_j = p_j \ \gamma_j + \alpha_j \ c$$

Second, we sum this equation over all the J products:

$$\sum_{j=1}^{J} p_j \ q_j = \sum_{j=1}^{J} p_j \ \gamma_j + \sum_{j=1}^{J} \alpha_j \ c$$

Third, taking into account this equation and the budget constraint $\sum_{j=1}^{J} p_j q_j = y - c$, we can obtain the following equation where the only endogenous variable is c:

$$y - c = \sum_{j=1}^{J} p_j \ \gamma_j + \sum_{j=1}^{J} \alpha_j \ c$$

Finally, solving for c in this equation, we get:

$$c = \frac{y - \sum_{j=1}^{J} p_j \ \gamma_j}{1 + \sum_{j=1}^{J} \alpha_j}$$

Q2.5. [10 points] Combine the equations in Q2.3 and Q2.4 to obtain an expression for the demand of q_j in terms of prices, income, and utility parameters. Using this demand equation, obtain the derivative $\partial q_j/\partial p_i$ for $i \neq j$. Are products i and j complements or substitutes? Why?

ANSWER: Combining the equations in Q2.3 and Q2.4 we get:

$$q_j \;=\; \gamma_j + \alpha_j \, \frac{1}{p_j} \left[\frac{y - \sum_{i=1}^{J} p_i \, \gamma_i}{1 + \sum_{i=1}^{J} \alpha_i} \right]$$

Therefore, the derivative $\partial q_j/\partial p_i$ for $i \neq j$ is:

$$\frac{\partial q_j}{\partial p_i} \;=\; -\frac{\alpha_j \, \gamma_i}{p_j \left[1 + \sum_{i=1}^{J} \alpha_i \right]} < 0$$

Products i and j (and for that matter, any pair of products in this model) are complements because an increase in the price of one of the products implies a reduction in the quantity of the other product, i.e., $\partial q_j/\partial p_i < 0$ for $i \neq j$.

Q2.6. [10 points] Suppose that we have data of individual purchases and prices over T periods of time, $t = 1, 2, ..., T$: $\{c_t,\ q_{1t},\ q_{2t},\ ...,\ q_{Jt}\}$ and $\{p_{1t},\ p_{2t},\ ...,\ p_{Jt}\}$. Use the linear equation in Q2.3 to obtain a linear regression model. Explain how to estimate the utility parameters of the model using this regression model.

ANSWER: Given the equation in Q2.3, we have the regression model:

$$q_{jt} \;=\; \gamma_j + \alpha_j \; x_{jt} + \xi_{jt}$$

where $x_{jt} = c_t/p_{jt}$. The error term ξ_{jt} can be the result of measurement error in the dependent variable q_{jt}, or measurement error in the regressor x_{jt}, or variation over time in the parameters γ_j or/and α_j.

Suppose that the error term ξ_{jt} comes only from measurement error in the dependent variable q_{jt} and that this measurement error is not correlated with the regressor. Then, $cov(p_{jt}, \xi_{jt}) = 0$ and we can estimate consistently the parameters of the model using an OLS estimator. Note that each product j has its own regression equation with its own parameters. Therefore, we need to estimate J regression equations, one for each product.

END