James R. Walker
6462 Social Science

Economic Statistics III

Problem Set 1

This problem set gives you an opportunity to estimate censor regression models using "control function methodology." Data for this exercise come from Tom Mroz's 1987 *Econometrica* paper ("The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," Volume 55, [July], pages 765-799). With some modification, the questions are from Ernst Berndt, *The Practice of Econometrics: Classic and Contemporary*. New York: Addison-Wesley. 1991. Chapter 11.

The data needed for this exercise is available from my webpage. Download two ascii files: mroz.dat and mroz.doc. mroz.doc documents the contents (variable names) of mroz.dat.

1. To check whether the data are the same as those employed by Mroz (Table III, page 769), compute and print out the arithmetic means and standard deviations of each of the 19 variables in mroz.dat. Also, compute minimum and maximum values for each variable; this is a particularly useful practice since by doing this, one can often spot coding errors. Are any of the min-max values "suspicious"? Why or why not? It is also good practice to explore your data through graphs.

2. Compare the sample of working women (the first 428 observations in mroz.dat) with those not working for pay in 1975 (the final 325 observations). Compute arithmetic means and standard deviations for each of the 19 variables in these two subsamples. What differences appear and what might these differences imply concerning the reservation wages of women in the two samples? To the extent that previous labor market experience of women (variable AX, see mroz.doc) reflects their preferences or tastes for market work, one might expect means of AX to differ between the working and nonworking samples. Is this the case? Interpret the difference. Are there any other differences in means of variables in the two subsamples that might affect labor force participation or hours worked? If so, please briefly explain.

3. In the model of labor supply estimated by Mroz, it is assumed that in making her labor supply decisions the wife takes as given the household's entire nonlabor income plus her husband's labor income. Mroz calls this sum the wife's property income and computes it as total family income minus the labor income earned by the wife. For the entire sample of 753 observations, compute this property income variable (hereafter called PRIN) as PRIN = FAMINC - (WHRS · WW). Save PRIN, as we'll use it in subsequent exercises. The mean and standard deviation of PRIN are 20129 and 11635.

4. The wage rate is typically not observed for women who are not working. Restricting your sample to workers, take the natural logarithm of the wife's wage variable WW and call the variable LWW. Then, for the entire sample of 753 observations calculate the square of wife's experience (say, AX2) and the square of the wife's age (WA2). Using the sample of working women, regress LWW on a constant term, WA, WE, CIT, AX, and AX2. Does this equation make sense? Why or why not? Use the estimated equation to predict the log wage for the 325 nonworking women in the sample. Call this variable FLWW. Compute the arithmetic mean and standard deviation for FLWW for non

workers and compare them to those for LWW from the working subsample. Is the difference in means substantial? How do you interpret this result? Define the variable LWW1 (for subsequent use) as equal to LWW (for working women) and equal to FLWW for nonworking women. To ensure that you have constructed the LWW1 variable correctly, its mean and standard deviation should equal 1.10432 and 0.58268. We'll use LWW1 below.

5. Use the entire data set and estimate via OLS a regression of WHRS on a constant term, and the variables KL6, KL618, WA, WE, LWW1, and PRIN. Interpret the estimated coefficients. Do the signs of the estimated coefficients agree with your intuition? Why or why not? What is the value of R^2? Why might this value be so low?

6. Using the OLS parameter estimates from part (5) compute the elasticity of hours of worked with respect to wages and to property income. Evaluate the elasticities, using Mroz's point of reference (see Table I, page 766). Is the wage elasticity a compensated or an uncompensated elasticity? Why? Compute the implied response of hours worked to $1 change in the wage rate, evaluated at the same point. How does this estimate compare with those presented in Table I, page 766? Compute the implied response of hours work to a $1000 increase in property income. Again, how does this estimated effect compare to those reported in Table I, page 766?

7. The estimation procedure in step 5 is simple to implement. Are the estimated coefficients consistent? Efficient? Please explain.

8. Using the 428 observations of working women regress WHRS on a constant term, KL6, KL618, WA, WE, PRIN, and LWW to obtain a conditional OLS model of labor supply. Compare your results to those obtained by Mroz (row 1 of Table IV, page 770). Your standard error estimates might differ from those reported by Mroz, since his estimates are adjusted for heteroskedasticity using the White [1980] robust estimation procedure. If your software program permits, compute the White robust standard errors; your results should be close to those reported by Mroz.

9. Are the estimated effects obtained in part (8) consistent? Please explain. Are the estimates in (8) preferred to those obtained in (5)? Why or why not?

10. Using the specification in step 8, estimate a Tobit model of labor supply. What can you say regarding the statistical significance of these Tobit parameters?

11. The total effect of the change in a regressor on expected hours in the Tobit model can be decomposed into two parts: the change in hours worked for those already working weighted by the probability of working plus the change in the probability of working weighted by the expected value of hours worked for those who work. Do this decomposition evaluated at sample means. Of the total change in hours due to a $1 change in the wage rate, what amount results from changes in hours worked from those who are already working? What amount comes from new entrants into the labor force?

12. What are the limitations of the Tobit specification? Is there evidence in the Mroz sample against the Tobit specification?

13. Generate a number of polynomial transformations of the wife's age, education, and experience to be used as explanatory variables in a labor force participation equation. (In the discussion below, the

polynomials will be referenced by numerical suffices. For example, the cube of the woman's education [WE] is denoted as WE3.) Estimate a probit of labor force participation (variable LFP) using as regressors a constant term, KL6, K618, WA, WE, WA2, WE2, WAWE (an interaction between WA and WE), WA3, WE3, WA2WE, WAWE2, WFED, WMED, UN, CIT, and PRIN. Using the estimated coefficients, calculate the inverse Mill's ratio for each observation, save this variable and call it INVR1. (Some computer software programs offer this calculation as an optional command; for others it must be computed by brute force.) Re-estimate the probit equation this time adding the experience variables AX and AX2 and calling the corresponding inverse Mill's ratio values INVR2. Comment on the statistical significance of parameter estimated in these two probit equations. Note that the LWW1 variable is excluded as an exploratory variable in this probit model. Since economic theory suggests that LFP is affected by the wage rate, why is the LWW1 excluded? In what sense, however, might it be included indirectly?

14. Next, restrict your sample to those who work for pay, (following Mroz) estimate by OLS a wage equation with LWW as a linear function of a constant term, KL6, K618, WA, WE, WE2, WAWE, WA3, WE3, WA2WE, WAWE2, WMED, WFED, UN, CIT, and PRIN. Call this set of explanatory variables "Set A." (If your software permits, calculate White corrected robust standard errors.) Comment on the signs and statistical significance of the estimated parameters. Re-estimate this equation (again by OLS), adding to Set A the experience variables AX and AX2. Interpret any changes in the results. Then with the same 428 observations (of working women), estimate by OLS two wage equations that allow for sample selectivity: first, an LWW wage equation with the Set A variables and the INVR1 inverse Mills ratio variable from part (13) included as regressors and second, an LWW wage equation with Set A variables, the AX and AX2 experience measures and the corresponding INVR2 variable. Comment on the sensitivity of the estimated parameters to the inclusion of the experience variables and to the sample selectivity adjustment. Is sample selectivity statistically significant? Is correcting for sample selectivity (in this application) substantively important?

15. Restricting your sample to those who work for pay, use the fitted values from the wage determination equation in (14) as instruments in a sample selectivity adjusted two-stage least squares estimation of hours worked equation. Specifically, following Mroz, for a base comparison, first employ instrumental variables (IV) estimation (with White robust standard errors, if possible) of an hours worked equation in which WHRS is the dependent variable and the explanatory variables include a constant term, KL6, K618, WA, WE, LWW, and PRIN; call this set of explanatory variables "Set B." In this IV or 2SLS estimation, treat LWW as an endogenous variable, and use the Set A variables defined in (14) to form instruments. How do your results compare with those reported by Mroz, in specification 2 of Table VI**?** (Note, Mroz reports White's corrected robust standard errors.) Next allow for sample selectivity, but exclude the experience variables. That is, using the same Set A variables plus the INVR1 inverse Mills ratio variable to form the instrument for LWW, estimate by IV or 2SLS an hours worked equation with the Set B variables as regressors but with INVR1 added as a regressor. How do your results compare with those reported by Mroz in specification 3 of Table IX? Is sample selectivity significant? Why or why not? Now, estimate a model by 2SLS in which the experience variables AX and AX2 are included along with the Set A variables in the first stage wage determination equation, but in which sample selectivity is not taken into account and only the Set B variables are regressors. Compare your results with those reported by Mroz (specification 4, Table IV). Finally, include the experience variables AX and AX2, the Set A variables, and the

INVR2 inverse Mills ratio as variables in forming the instrument for LWW, and estimate by 2SLS the hours worked equation included as regressors the Set B variables and INVR2. Your results should conform closely with those reported by Mroz (specification 3, Table X). Is sample selectivity important when the experience variables are treated as exogenous? Why or why not?

16. Interpret your findings in part (15), commenting on the importance of allowing for sample selectivity and how this sensitivity is affected by the assumption of exogeneity of the experience variables. Mroz concluded that with his preferred specifications the uncompensated wage effect for this sample is small, as is the estimated income effect. Do you agree? Why or why not?

17. For the entire sample of 753 observations create and save several tax-related variables: a virtual income variable defined as VPRIN = OneMmtr*PRIN, where OneMmtr is one minus the marginal tax rate variable on the dataset and PRIN is defined in part (3); LTAX = ln(OneMmtr), and the logarithm of an after-tax variable as LTWW = LTAX + LWW1.

18. Estimate a model of labor supply with taxes. Use as exogenous variables a constant term, KL6, K618, WA, WE, WA2, WE2, WAWE, WA3, WE3, WA2WE, WAWE2, WMED, WFED, UN, CIT, PRIN, and the inverse Mills ratio variable, INVR1, from part (14) to form an instrument for the log of the after-tax wage rate variable LTWW. With these instruments for LTWW, estimate by 2SLS a model in which WHRS is a linear function of a constant term, LTWW, LTAX, KL6, K618, WA, WE, PRIN, VPRIN, INVR1, and a random disturbance term. Formulate and test the hypothesis that women optimally take income taxes into account when making labor supply decisions. Then formulate and test the hypothesis that women entirely ignore income taxes when making labor supply decisions. Interpret your results, do your conclusions agree with those reported by Mroz?

19. Experiment with any two other plausible specifications for the probit, wage determination, and/or hours worked equations with taxes included. Explain why these specifications are plausible and comment on the sensitivity of your results to changes in the specification. Do Mroz's results on the unimportance of income taxes appear to be robust?