Written Exam at the Department of Economics
Winter 2017–18

Advanced Microeconometrics

Final Exam

— Suggested Answers —

Problem 1

To investigate the determinants of female labor force participation, you are provided with an independent sample of $N = 2,477$ married women from the US National Longitudinal Survey of Youth (NLSY).

The dependent variable $y \equiv$ `WEEKS` measures the number of weeks worked by these women in 1990. Approximately 45% of the sample reports working full-time ($y = 52$) and 17% reports being out of the labor force ($y = 0$). Part-time workers report a number of hours that can be considered to be continuous.

The following explanatory variables are available:

- `AFQT`: A measure of cognitive ability (Armed Forces Qualifying Test, standardized test score with zero mean and unit variance).
- `EDUC`: Number of years of schooling completed.
- `HUSBINC`: Husband's income in the previous year (in thousand USD).
- `KIDS`: Binary indicator for having children (= 1 if at least one child, 0 otherwise).

These covariates are stored in a $N \times 5$ matrix X, where the first column is a vector of 1s for the intercept term of the model. The column vector x_i is used to denote the i^{th} row of X.

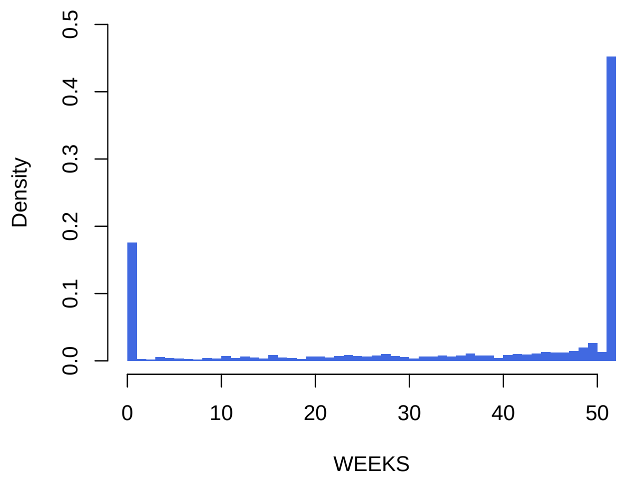

Figure 1.1: Histogram of the dependent variable `WEEKS`.

Question 1.1: Describe the features of the dependent variable `WEEKS`, and explain how the Tobit model can be extended to accommodate them. You should state your model analytically as precisely as possible.

[Note: If you do not manage to answer this question, use the Tobit model instead for this question and the remaining ones.]

Suggested answer

The dependent variable y displayed in Fig. 1.1 shows a bunching at 52 weeks (woman working full-time), and another one at 0 (women not working). The Tobit model studied in the course is an appropriate candidate for this type data, but it has to be extended to account for this double corner solution of the dependent variable.

The double corner solution happens because it is not possible to work a negative number of hours (left-censoring), and a year only has 52 weeks that can be worked (right-censoring).

A double-censored model (two-sided Tobit) can be constructed based on the following observational rule:

$$y_i = \begin{cases} 0 & \text{if } y_i^\star \leq 0, \\ y_i^\star & \text{if } 0 < y_i^\star < 52, \\ 52 & \text{if } y_i^\star \geq 52, \end{cases}$$

where the latent variable is expressed as

$$y_i^\star = x_i'\beta + \varepsilon_i, \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}\big(0, \sigma^2\big)\,. \tag{1}$$

The error terms ε_i are assumed to be independent and identically distributed (*iid*) across women, as we are working with an independent sample of individuals.

This model can be used to explain the following statistics, among other quantities of interest:

- Probability of not working: $\Pr(y_i = 0 \mid x_i)$.
- Probability of working part-time: $\Pr(0 < y_i < 52 \mid x_i)$.
- Probability of working full-time: $\Pr(y_i = 52 \mid x_i)$.

- Expected number of hours worked for those working part-time: $\mathrm{E}[y_i \mid 0 < y_i < 52, x_i]$.

Question 1.2: Derive the likelihood function $L_N(\theta; y, X)$ of the model specified in Question 1.1, where θ is the vector of model parameters.

Suggested answer

Let $\theta = (\beta', \sigma)'$. To derive the likelihood function, the three cases have to be considered:

(a) Probability of being out of the labor force:

$$\begin{aligned}
\Pr(y_i = 0 \mid x_i, \theta) &= \Pr(x_i'\beta + \varepsilon_i \leq 0 \mid x_i, \theta)\,, \\
&= \Pr\left(\frac{\varepsilon_i}{\sigma} \leq -\frac{x_i'\beta}{\sigma} \mid x_i, \theta\right), \\
&= \Phi\left(-\frac{x_i'\beta}{\sigma}\right) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right), \qquad (2)
\end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of the standard normal distribution, i.e., $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp\{-t^2/2\} dt$, and the last equality is obtained from the symmetry of the normal distribution.

(b) Probability of working full-time:

$$\begin{aligned}
\Pr(y_i = 52 \mid x_i, \theta) &= \Pr(x_i'\beta + \varepsilon_i \geq 52 \mid x_i, \theta)\,, \\
&= 1 - \Pr(x_i'\beta + \varepsilon_i < 52 \mid x_i, \theta)\,, \\
&= 1 - \Pr\left(\frac{\varepsilon_i}{\sigma} < \frac{52 - x_i'\beta}{\sigma} \mid x_i, \theta\right), \\
&= 1 - \Phi\left(\frac{52 - x_i'\beta}{\sigma}\right) = \Phi\left(\frac{x_i'\beta - 52}{\sigma}\right).
\end{aligned}$$

(c) Density function of y_i for women working part-time (the latent variable is observed in that case):

$$f(y_i \mid x_i, \theta) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right),$$

where $\phi(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$ is the probability density function

(PDF) of the standard normal distribution.

The probability of working part-time would therefore be
$\Pr(0 < y_i < 52 \mid x_i, \theta) = 1 - \Pr(y_i = 0 \mid x_i, \theta) - \Pr(y_i = 52 \mid x_i, \theta)$.

Putting the three pieces together provides the full likelihood function:

$$L_N(\theta; y, X) = \prod_{i=1}^{N} \left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]^{\mathbb{1}\{y_i=0\}} \left[\Phi\left(\frac{x_i'\beta - 52}{\sigma}\right)\right]^{\mathbb{1}\{y_i=52\}} \times \left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]^{\mathbb{1}\{0<y_i<52\}},$$

where the indicator function $\mathbb{1}\{\cdot\}$ is equal to 1 if the corresponding condition is fulfilled, to 0 otherwise, and allows to distinguish between the three different cases depending on the observed outcome y_i.

Question 1.3: Discuss the identification of the model, and especially, whether any parameter restrictions are required for identification.

Suggested answer

As in the standard Tobit model seen in the course, this model is identified, provided that some women are observed in the three different cases (not working/working part-time/working full-time), and that there is no multicollinearity problem with the explanatory variables. Then, no parameter restrictions are required.

Multicollinearity of the explanatory variables arises if at least one of them is a linear combination of some (or all) of the other ones. This results in a matrix X that is not full rank, and therefore creates identification problems (e.g., $X'X$ cannot be inverted in that case). To see this, simplify the model to have only two covariates x_{1i} and x_{2i}, and assume that $x_{2i} = cx_{1i}$, for a fixed constant c. Then in the likelihood function, $x_i'\beta$ enters as $x_{1i}\beta_1 + x_{2i}\beta_2 = x_{1i}(\beta_1 + c\beta_2) \equiv x_{1i}\delta$. Therefore, only $\delta \equiv \beta_1 + c\beta_2$ can be identified (and subsequently estimated to provide $\widehat{\delta} \equiv \widehat{\beta_1 + c\beta_2}$), but there is an infinite number of values for β_1 and β_2 that would provide the same likelihood.

The variance of the error term may be a source of identification problems in some latent variables (e.g., probit). This happens when the latent variable can be rescaled without affecting the likelihood. In this model, however, the outcome is observed for the women working part-time ($0 < y_i < 52$), which provides identification of σ.

This can be seen formally by observing that the term $-N_1 \log \sigma$ enters the log-likelihood function (where N_1 is the numer of women working part-time), so that any change in σ results in a change in $L_N(\theta; y, X)$. Thus, σ is uniquely identified. As a consequence, β can also be identified.

In the special case where there are no observations in the interval $]1, 52[$ (i.e., no women working part-time), σ is not identified. In this case, we are back to the framework of the probit model, where we can only explain the probabilities of not working or working full-time. This problem can also arise empirically if too few women working part-time are observed (empirical identification), creating estimation issues for the parameter σ (e.g., convergence problems).

Question 1.4: Derive analytically the marginal effect of a given (continuous) explanatory variable x_j on the probability of working (either part-time or full-time). You are not asked to compute this marginal effect.

Suggested answer

This marginal effect can be obtained using the probability derived in Eq. (2):

$$\begin{aligned}\frac{\partial}{\partial x_j} \Pr(y_i^\star > 0 \mid x_i, \theta) &= \frac{\partial}{\partial x_j} \left(1 - \Pr(y_i^\star \leq 0 \mid x_i, \theta)\right), \\ &= \frac{\partial}{\partial x_j} \Phi\left(\frac{x_i'\beta}{\sigma}\right), \\ &= \frac{\beta_j}{\sigma}\, \phi\left(\frac{x_i'\beta}{\sigma}\right),\end{aligned}$$

using the fact that $\partial \Phi(t)/\partial t = \phi(t)$.

Question 1.5: Discuss the maximum likelihood estimation results presented in Table 1.1, which shows the parameter estimates (Est.) and their standard errors (SE).

Suggested answer

These parameter estimates should be interpreted carefully, as they measure the impact of the explanatory variables on the *latent* variable, not on the *observed* outcome.

To get more insights, it would be necessary to derive and compute marginal effects, such as $\partial \mathrm{E}[y \mid 0 < y < 52, x, \theta] / \partial x_j$ for the impact of covariate x_j on the expected number of hours for women working part-time. This is beyond the current question asked.

It is possible, however, to interpret the significance and the signs of these coefficients. This can be done, for example, by using a Student t-test for each regression coefficient β_k, using the estimates reported in Table 1.1. Remember that the t-statistic is computed as follows and can be approximated by a standard normal distribution asymptotically:

$$\frac{\widehat{\beta}_k}{\widehat{\mathrm{SE}}(\widehat{\beta}_k)} \quad \to \quad \mathcal{N}(0,1)\,.$$

A quick calculation shows that all t-statistics are above the 5%-critical value (in absolute value) of the standard normal distribution, which is equal to 1.96 for a two-sided test. Therefore, all covariates have an impact significantly different from zero on the outcome.

All covariates have the signs we would expect: More educated women and with higher cognitive abilities tend to work more, while having children and husband's income reduce their number of hours and their probability or working. These results are consistent with a model of substitutability between the spouses' amount of work, and household production (including child care).

Table 1.1: Maximum likelihood estimation results.

	Est.	SE
CONST	49.97	7.85
AFQT	0.14	0.05
HUSBINC	-0.24	0.05
KIDS	-31.19	2.67
EDUC	1.56	0.62
σ	44.51	1.23

Problem 2

Consider an *iid* sample $y = (y_1, \ldots, y_N)'$ drawn from an exponential distribution with scale parameter $\theta > 0$ such that, for each $i = 1, \ldots, N$:

$$y_i \overset{iid}{\sim} \mathcal{E}xpon(\theta), \qquad f(y_i \mid \theta) = \frac{1}{\theta} \exp\left\{-\frac{y_i}{\theta}\right\}, \qquad \mathrm{E}[y_i] = \theta.$$

Question 2.1: Propose a *natural conjugate* prior distribution for θ. Derive the corresponding posterior distribution.

You may use one of the distributions of Table 2.1. Justify your choice.

Suggested answer

Remember that a prior is said to be *conjugate* if it leads to a posterior distribution that belongs to the same family of distribution.

The likelihood function is

$$L_N(\theta) = \prod_{i=1}^{N} \theta^{-1} \exp\left\{-\frac{y_i}{\theta}\right\} = \theta^{-N} \exp\left\{-\frac{\sum_{i=1}^{N} y_i}{\theta}\right\}. \tag{3}$$

It appears to match with the kernel of the inverse-gamma distribution provided in Table 2.1, which is $p(\theta) \propto \theta^{-a-1} \exp\{-b/\theta\}$. Therefore, the inverse-gamma distribution is a natural conjugate prior for θ.

Conjugacy can be confirmed by applying Bayes' theorem to obtain the posterior. Using $\theta \sim \mathcal{IG}(a_0, b_0)$ *a priori* and working directly with the kernels of the likelihood and of the prior provides:

$$\begin{aligned} p(\theta \mid y) &\propto p(y \mid \theta) p(\theta), \\ &\propto \theta^{-N} \exp\left\{-\frac{\sum_{i=1}^{N} y_i}{\theta}\right\} \times \theta^{-a_0-1} \exp\left\{-\frac{b_0}{\theta}\right\}, \\ &\propto \theta^{-a_0-N-1} \exp\left\{-\frac{b_0 + \sum_{i=1}^{N} y_i}{\theta}\right\}, \end{aligned}$$

which corresponds to the kernel of an inverse-gamma distribution (thus confirming the conjugacy of the inverse-gamma distribution):

$$\theta \mid y \sim \mathcal{IG}\left(a_0 + N, b_0 + \sum_{i=1}^{N} y_i\right).$$

Question 2.2: Show that Jeffreys' prior for this model is $p(\theta) \propto \theta^{-1}$. Derive the corresponding posterior distribution of θ.

Hint: Remember that Jeffreys' prior is proportional to the square root of the determinant of the information matrix, i.e., $p(\theta) \propto |\mathcal{I}(\theta)|^{1/2}$, which simplifies to $p(\theta) \propto \mathcal{I}(\theta)^{1/2}$ in the present scalar case.

Suggested answer

To derive Jeffreys' prior, let us first obtain the information matrix (which is just a scalar in this particular case).

Log-likelihood function of this model, obtained from Eq. (3):

$$\ln L_N(\theta) = -N \ln \theta - \frac{\sum_{i=1}^{N} y_i}{\theta}.$$

First derivative of the log-likelihood function:

$$\frac{\partial \ln L_N(\theta)}{\partial \theta} = -\frac{N}{\theta} + \frac{\sum_{i=1}^{N} y_i}{\theta^2}.$$

Second derivative of the log-likelihood function:

$$\frac{\partial^2 \ln L_N(\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{\partial \ln L_N(\theta)}{\partial \theta}\right) = \frac{N}{\theta^2} - \frac{2\sum_{i=1}^{N} y_i}{\theta^3}.$$

Information:

$$\begin{aligned}\mathcal{I}(\theta) &= \mathrm{E}\left[-\frac{\partial^2 \ln L_N(\theta)}{\partial \theta^2}\right] = \mathrm{E}\left[-\frac{N}{\theta^2} + \frac{2\sum_{i=1}^{N} y_i}{\theta^3}\right], \\ &= -\frac{N}{\theta^2} + \frac{2\sum_{i=1}^{N} \mathrm{E}[y_i]}{\theta^3} = -\frac{N}{\theta^2} + \frac{2N\theta}{\theta^3} = \frac{N}{\theta^2},\end{aligned}$$

because the expectation is taken with respect to the distribution of the random variable y_i, and $\mathrm{E}[y_i] = \theta$ for the exponential distribution.

Hence, Jeffreys' prior is $p(\theta) \propto |N\theta^{-2}|^{1/2} \propto \theta^{-1}$.

Corresponding posterior distribution:

$$p(\theta \mid y) \propto L_N(\theta)p(\theta) \propto \theta^{-N-1} \exp\left\{-\frac{\sum_{i=1}^N y_i}{\theta}\right\},$$

which is the kernel of the following inverse-gamma distribution:

$$\theta \mid y \sim \mathcal{IG}\left(N, \sum_{i=1}^{N} y_i\right).$$

Note that this posterior distribution appears to be the limit distribution of the one derived in Question 2.1 when $a_0 \to 0$ and $b_0 \to 0$.

Table 2.1: Some probability distributions.

Distribution	Density $f(\theta \mid a, b)$	Mean
Uniform	$\frac{1}{b-a}$	$\frac{a+b}{2}$
Beta	$\frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}$	$\frac{a}{a+b}$
Gamma	$\frac{1}{\Gamma(a)b^a}\theta^{a-1}\exp\left\{-\frac{\theta}{b}\right\}$	ab
Inverse-Gamma	$\frac{b^a}{\Gamma(a)}\theta^{-a-1}\exp\left\{-\frac{b}{\theta}\right\}$	$\frac{b}{a-1}$ (for $a > 1$)

Problem 3

Consider the following MATLAB code:

```
1  % Inputs:
2  %      y     vector  Nx1
3  %      x     vector  Nx1
4  %      x0    scalar  1x1
5  %      h     scalar  1x1
6  function [yhat] = regress(y,x,x0,h)
7      kern = @(z) normpdf(z);
8      w = kern((x-x0)./h);
9      w = w./sum(w);
10     yhat = sum(w.*y);
11 end
```

Question 3.1: Express in mathematical terms what this MATLAB function computes. You should only provide one or two equations to answer this question. Be explicit about the notation.

Suggested answer

Given the comments of this code (lines 1–5), this function takes as arguments two vectors $y = (y_1, \ldots, y_N)'$ and $x = (x_1, \ldots, x_N)'$, two scalars x_0 and h, and returns a scalar $\widehat{y}$ computed as

lines 7–9:

$$w_i = \frac{\phi\left(\dfrac{x_i - x_0}{h}\right)}{\sum_{j=1}^{N} \phi\left(\dfrac{x_i - x_0}{h}\right)},$$

line 10:

$$\widehat{y} = \sum_{i=1}^{N} w_i y_i,$$

where $\phi(\cdot)$ is the probability density function (PDF) of the standard normal distribution, implemented in MATLAB by the function `normpdf()`.

Question 3.2: Using only words *(no equations required)*, explain *briefly* the methodology implemented by this function. In particular, describe the role of the two scalars "`x0`" and "`h`", and explain the values they can take.

Suggested answer

This function implements kernel regression—more precisely, the Nadaraya-Watson estimator—using the Gaussian (normal) kernel. It provides, for a given value x_0 of the explanatory variable, a nonparametric estimate $\widehat{y}$ of the dependent variable, computed as the weighted average of all the observed values of y. The weights are computed as a function of the explanatory variable x and depend on how close to x_0 these values are. The PDF of the standard normal distribution is used as *kernel* to measure this distance.

The parameter h is called bandwidth and allows to adjust the weight each observation gets, depending on how far it is from the point x_0. The smaller the bandwidth, the less weight the observations get when they are far from x_0. Conversely, the larger, the more weight. As a consequence, h determines how smooth the estimator is, and is often referred to as *smoothing parameter*. Appropriate methods have been developed to find the optimal bandwidth. The bandwidth h should be strictly positive, while x_0 can take any real value on the support of the explanatory variable x.

Usually, this function will be called several times for different values of x_0, to then plot the corresponding nonparametric regression line.