Written Exam at the Department of Economics
Winter 2018–19

Advanced Microeconometrics

Final Exam

— Suggested Answers —

Problem 1 (50%)

Consider the following nonlinear regression for the binary outcome $y_i \in \{0, 1\}$:

$$y_i = g(x_i, \beta) + u_i, \tag{1}$$

$$\mathrm{E}[y_i \mid x_i] = g(x_i, \beta) = \frac{\exp\{x_i'\beta\}}{1 + \exp\{x_i'\beta\}}, \tag{2}$$

for $i = 1, \ldots, N$, where x_i is a vector of observed characteristics, and β is the corresponding vector of regression coefficients.

Question 1.1: Within what range does this model restrict $\mathrm{E}[y_i \mid x_i]$ to lie? Justify your answer.

Suggested answer

The conditional mean $\mathrm{E}[y_i \mid x_i]$ of the dependent variable lies within $(0, 1)$ (both bounds excluded): For any given real vector x_i, the product $x_i'\beta$ takes values on $(-\infty, +\infty)$ since β is not restricted. This implies that $\exp\{x_i'\beta\}$ lies between $(0, +\infty)$. Therefore, the ratio in Eq. (2) tends to 0 when $x_i'\beta \to -\infty$, and to 1 when $x_i'\beta \to +\infty$.

Question 1.2: Express the optimization problem that can be used to obtain the nonlinear least squares (NLS) estimator of β.

Suggested answer

The optimization problem can be expressed as:

$$\widehat{\beta}_{\mathrm{NLS}} = \arg\min_{\beta} Q_N(\beta), \qquad \text{with } Q_N(\beta) = \frac{1}{2}\sum_{i=1}^{N}(y_i - g(x_i, \beta))^2.$$

(Note that the factor $1/2$ is only used to simplify the subsequent derivations.)

Question 1.3: Show that the first-order conditions for the NLS estimator of

β can be expressed as

$$\sum_{i=1}^{N} w_i(y_i - g(x_i, \beta))x_i = 0, \qquad \text{with } w_i = \frac{\exp\{x_i'\beta\}}{(1+\exp\{x_i'\beta\})^2}. \tag{3}$$

Suggested answer

First-order conditions:

$$\begin{aligned} \frac{\partial Q_N(\beta)}{\partial \beta} = 0 \qquad &\Rightarrow \qquad \frac{\partial}{\partial \beta}\left(\frac{1}{2}\sum_{i=1}^{N}(y_i - g(x_i, \beta))^2\right) = 0, \\ &\Rightarrow \qquad \sum_{i=1}^{N} \frac{\partial g(x_i, \beta)}{\partial \beta}(y_i - g(x_i, \beta)) = 0. \end{aligned}$$

Given the expression of $g(x_i, \beta)$ in Eq. (2), the corresponding first derivative is:

$$\begin{aligned} \frac{\partial g(x_i, \beta)}{\partial \beta} &= \frac{\exp\{x_i'\beta\}x_i(1+\exp\{x_i'\beta\}) - \exp\{2x_i'\beta\}x_i}{(1+\exp\{x_i'\beta\})^2}, \\ &= \frac{\exp\{x_i'\beta\}}{(1+\exp\{x_i'\beta\})^2}x_i, \end{aligned}$$

which provides the weight w_i in Eq. (3).

Question 1.4: Which condition is required for consistency of the NSL estimator of β? Is this condition fulfilled in this model? Justify your answer analytically.

Suggested answer

Using the first-order conditions derived above, it comes that consistency will be achieved if

$$\mathrm{E}[w_i(y_i - g(x_i, \beta_0))x_i] = 0,$$

where β_0 denotes the true value of β.

This holds if $\mathrm{E}[u_i \mid x_i] = 0$. Indeed, since $u_i = y_i - g(x_i, \beta_0)$, we get

$$\begin{aligned} \mathrm{E}[w_i(y_i - g(x_i, \beta_0))x_i] = \mathrm{E}[w_i u_i x_i] &= \mathrm{E}[\mathrm{E}[w_i u_i x_i \mid x_i]]\,, \\ &= \mathrm{E}[w_i x_i \mathrm{E}[u_i \mid x_i]]\,, \\ &= 0, \end{aligned}$$

using the law of iterated expectations. Therefore, the conditional mean of the error term needs to be equal to 0 for consistency. This holds if the conditional mean of the dependent variable is correctly specified. In this model:

$$\mathrm{E}[u_i \mid x_i] = \mathrm{E}[y_i - g(x_i, \beta) \mid x_i]\,, = \mathrm{E}[y_i \mid x_i] - g(x_i, \beta) = 0$$

because of the assumption in Eq. (2).

Alternatively, the zero conditional mean of the error term can be shown by realizing that each error term only has two possible values in this model:

$$u_i = \begin{cases} 1 - g(x_i, \beta) & \text{if } y_i = 1, \\ -g(x_i, \beta) & \text{if } y_i = 0. \end{cases}$$

Therefore,

$$\begin{aligned} \mathrm{E}[u_i \mid x_i] &= (1 - g(x_i, \beta)) \Pr(y_i = 1 \mid x_i) - g(x_i, \beta)) \Pr(y_i = 0 \mid x_i)\,, \\ &= (1 - g(x_i, \beta))g(x_i, \beta) - g(x_i, \beta))(1 - g(x_i, \beta)), \\ &= 0, \end{aligned}$$

because $\Pr(y_i = 1 \mid x_i) = \mathrm{E}[y_i \mid x_i] = g(x_i, \beta)$.

Question 1.5: Show that $\mathrm{V}[y_i \mid x_i] = g(x_i, \beta)(1 - g(x_i, \beta))$. How can you use this result to improve on the NLS estimation of β? Describe briefly the alternative approach you suggest.

Suggested answer

Conditional variance:

$$\begin{aligned}
\mathrm{V}[y_i \mid x_i] &= \mathrm{E}\big[y_i^2 \mid x_i\big] - \mathrm{E}[y_i \mid x_i]^2\,, \\
&= \mathrm{E}[y_i \mid x_i] - \mathrm{E}[y_i \mid x_i]^2\,, \\
&= g(x_i, \beta) - g(x_i, \beta)^2, \\
&= g(x_i, \beta)(1 - g(x_i, \beta)),
\end{aligned}$$

using the fact that $\mathrm{E}[y_i^2 \mid x_i] = \mathrm{E}[y_i \mid x_i]$ because y_i is binary. This conditional variance is the variance of the corresponding Bernoulli distribution, as we are dealing with a binary dependent variable.

The previous expression of the conditional variance of the dependent variable implies that the error terms are heteroskedastic, which affects the efficiency of the basic NLS estimator. To improve efficiency, the Generalized Nonlinear Least Squares (GNLS) estimator controls for this heteroskedasticity:

$$\widehat{\beta}_{\text{GNLS}} = \arg\min_{\beta} \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_{0i}^2} (y_i - g(x_i, \beta))^2.$$

where $\sigma_{0i}^2 \equiv \mathrm{V}[u_i \mid x_i]|_{\beta_0} = g(x_i, \beta_0)(1 - g(x_i, \beta_0))$, where β_0 is the true value of the parameter β.

This estimator, however, cannot be obtained directly, as it depends on the unknown true parameter β_0. A feasible version is the Feasible Generalized Nonlinear Least Squares (FGNLS) estimator:

$$\widehat{\beta}_{\text{FGNLS}} = \arg\min_{\beta} \frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - g(x_i, \beta))^2}{g(x_i, \widehat{\beta})(1 - g(x_i, \widehat{\beta}))},$$

where $\widehat{\beta}$ is a consistent estimator of β that allows to consistently estimate the variance of the error term.

The FGNLS estimator can be obtained in a multi-step procedure: First run NLS to obtain $\widehat{\beta}_{\text{NLS}}$, a consistent estimator of β. Then, use this estimator in a second step to compute the corresponding variance of the errors. Finally, in a third step use this estimator of the variance of the error terms to solve the optimization problem above providing $\widehat{\beta}_{\text{FGNLS}}$.

Question 1.6: Which alternative econometric model, assuming the same conditional expectation of the outcome as in Eq. (2), could you use to estimate β with maximum likelihood estimation (MLE)? Discuss briefly the differences between this MLE approach and the NLS approach.

Suggested answer

The expression of the conditional mean of the dependent variable corresponds to the logit model:

$$\Pr(y_i = 1 \mid x_i) = \mathrm{E}[y_i \mid x_i] = \frac{\exp\{x_i'\beta\}}{1 + \exp\{x_i'\beta\}},$$

with likelihood function, asuming independence of the observation $i = 1, \ldots, N$:

$$\begin{aligned} L_N(\beta) &= \prod_{i=1}^{N} \Pr(y_i = 0 \mid x_i)^{1-y_i} \Pr(y_i = 1 \mid x_i)^{y_i}, \\ &= \prod_{i=1}^{N} \left(\frac{1}{1 + \exp\{x_i'\beta\}} \right)^{1-y_i} \left(\frac{\exp\{x_i'\beta\}}{1 + \exp\{x_i'\beta\}} \right)^{y_i} \end{aligned}$$

This alternative model can be estimated with maximum likelihood estimation:

$$\widehat{\beta}_{\mathrm{ML}} = \arg\max_{\beta} \ln L_N(\beta).$$

Both MLE and NLS provide consistent estimators, but not with the same efficiency. The ML estimator will be more efficient than NLS, but it requires a correct specification of the data generating process. This may be a strong requirement in practice. NLS is more flexible because it relies on weaker distributional assumption. Only the conditional mean of the dependent variable needs to be well specified to apply NLS. The feasible generalized version of NLS is more efficient than NLS, but requires to known the specification of the conditional variance of the error term (if only a guess can be made on the expression of the variance, then weighted nonlinear least squares (WNLS) can be used instead). Therefore, in practice there is a trade-off between the distribution assumptions the analyst is willing to make and the efficiency of the resulting estimator.

Problem 2 (30%)

Consider a stock with price p_t at each time period $t = 0, \ldots, T$. You are interested in modeling the price fluctuations of this stock using the binary indicator $y_t = \mathbb{1}\{p_t - p_{t-1} > 0\}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function that is equal to 1 if the corresponding condition is fulfilled, to 0 otherwise.

For simplicity, assume that y_t is independent across time periods.

The goal of the analysis is to make inference on the parameter $\theta \equiv \Pr(y_t = 1)$, for $t = 1, \ldots, T$.

Question 2.1: Propose a model for this analysis and derive the corresponding likelihood function. Specify a prior distribution on θ that is a natural conjugate prior and derive the corresponding posterior distribution.

[Hint: You may use a distribution from Table 2.1, or a different one.]

Suggested answer

Each random variable y_t can be assumed to follow a Bernoulli distribution with parameter $0 \leq \theta \leq 1$. Given the independence assumed across time periods, the corresponding likelihood function can be expressed as:

$$p(y \mid \theta) \equiv L_T(\theta) = \prod_{t=1}^{T} \theta^{y_t}(1-\theta)^{(1-y_t)} = \theta^S(1-\theta)^{T-S}.$$

with $y = (y_1, \ldots, y_T)'$ and $S = \sum_{t=1}^{T} y_t$. This likelihood function has the same form as the kernel of the Beta distribution for θ:

$$\theta \sim \mathcal{B}\text{eta}(a_0, b_0)\,, \qquad\qquad p(\theta) \propto \theta^{a_0-1}(1-\theta)^{b_0-1}.$$

Therefore, the Beta distribution is a natural conjugate prior in this model, due to the fact that the corresponding posterior distribution is also a Beta distribution. This can be verified by applying Bayes' theorem:

$$
\begin{aligned}
p(\theta \mid y) &\propto p(y \mid \theta) p(\theta), \\
&\propto \theta^{S}(1-\theta)^{T-S} \theta^{a_0-1}(1-\theta)^{b_0-1}, \\
&\propto \theta^{a_0+S-1}(1-\theta)^{b_0+T-S-1},
\end{aligned}
$$

which corresponds to the kernel of the following Beta distribution:

$$
\theta \mid y \sim \mathcal{B}\text{eta}(a_0+S, b_0+T-S)\,.
$$

Question 2.2: How would you choose the values of the prior parameters to obtain a flat prior? Show that with such a flat prior, the mean of the posterior distribution is asymptotically identical to the value of the maximum likelihood estimator.

Suggested answer

A flat prior puts the same weight on all possible values of the parameter, i.e., $p(\theta) \propto 1$. Since in this model $0 \leq \theta \leq 1$, a flat prior corresponds to a uniform prior on $[0,1]$, which can be obtained by specifying $a_0 = 1$ and $b_0 = 1$:

$$
p(\theta \mid a_0 = 1, b_0 = 1) \propto \theta^{(1-1)}(1-\theta)^{(1-1)} \propto 1.
$$

Using this flat prior and the expression of the posterior distribution derived in the previous question, the posterior mean of θ is (see Table 2.1):

$$
\mathrm{E}[\theta \mid y] = \frac{a_0+S}{a_0+b_0+T} = \frac{S+1}{T+2} \qquad \overset{T\to\infty}{\longrightarrow} \qquad \frac{S}{T},
$$

where S/T is the sample average of the binary variable y_t.

Given that the maximum likelihood estimator of θ is:

$$\widehat{\theta}_{\mathrm{ML}} = \arg\max_{\theta} \ln L_T(\theta) = \arg\max_{\theta} \left(S \ln(\theta) + (T-S)\ln(1-\theta) \right),$$

$$\begin{aligned}
\text{FOC:} \quad & \frac{\partial \ln L_T(\theta)}{\partial \theta} = 0, \\
\Rightarrow \quad & \frac{S}{\theta} - \frac{T-S}{1-\theta} = 0, \\
\Rightarrow \quad & \widehat{\theta}_{\mathrm{ML}} = \frac{S}{T}.
\end{aligned}$$

we can conclude that the posterior mean of θ coincides with the maximum likelihood estimator asymptotically.

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 (20%)

You would like to compute the cumulative distribution function (CDF) of the standard normal distribution using an alternative to the `normcdf()` function provided in `MATLAB`.

One of your colleagues gives you the following piece of code, where you need to select one option for the two ingredients A and B, respectively:

```matlab
function [cdf] = cdf_normal(x,M)

    rng(1);
    pdf_normal = @(t) exp(-(t.^2)/2)/sqrt(2*pi);

    % A: first ingredient
    z = rand(M,1);              % A1
    z = randn(M,1);             % A2
    z = x + randn(M,1);         % A3

    % B: second ingredient
    w = z.*x;                   % B1
    w = z.*pdf_normal(x);       % B2
    w = z < x;                  % B3
    w = pdf_normal(z) < x;      % B4

    % result
    cdf = mean(w);

end
```

Question 3.1: State your choice for the two ingredients A and B (for example, "A1 and B1"), and provide the corresponding mathematical expression of the formula used in this function.

Suggested answer

The ingredients A2 and B3 allow to approximate the CDF of the standard normal distribution (denoted $\Phi(\cdot)$) based on the following expression:

$$\Phi(x) = \int_{-\infty}^{x} \phi(z)dz = \int_{-\infty}^{+\infty} \mathbb{1}\{z \leq x\}\, \phi(z)dz \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\{z^{(m)} \leq x\}$$

where $z^{(m)} \sim \mathcal{N}(0,1)$ for $m = 1, \ldots, M$, and where $\phi(\cdot)$ denotes the probability distribution function (PDF) of the standard normal distribution.

Question 3.2: Describe precisely the theory behind this function. In particular, explain the role of the constant `M` and how it should be specified.

Suggested answer

This function relies on Monte Carlo integration to approximate the CDF of the standard normal distribution. It uses the fact that $\Phi(x) = \mathrm{E}[\mathbb{1}\{X \leq x\}]$, where $X \sim \mathcal{N}(0,1)$. Using the law of large numbers, this expectation can be approximated by the following sample average

$$\frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\{z^{(m)} \leq x\} \qquad \xrightarrow{p} \qquad \mathrm{E}[\mathbb{1}\{X \leq x\}]\,,$$

which converges in probability to the expectation on the right when $M \to +\infty$, where $z^{(1)}, \ldots, z^{(M)}$ is a sequence of independent random draws from the standard normal distribution.

The number of random draws M determines the precision of the approximation. The larger the number of draws, the better the approximation. In practice, however, only a finite number of draws can be used. It is therefore important to investigate the impact of M on the results. It should be large enough to provide a stable approximation, but not too large to prevent computational burden.

[Hint: see next page...]

[Hint: For any random variable Z with probability distribution function $f(\cdot)$ and cumulative distribution function $F(\cdot)$, remember that

$$E[Z] = \int z f(z) dz \qquad \textit{and} \qquad F(z) = \int_{-\infty}^{z} f(t) dt = E[\mathbb{1}\{Z \leq z\}],$$

where $\mathbb{1}\{\cdot\}$ is the indicator that is equal to 1 if the corresponding condition is fulfilled, to 0 otherwise.]