Discussion Papers No. 264, January 2000
Statistics Norway, Research Department

John K. Dagsvik

Multinomial Choice and Selectivity

Abstract:
In this paper we discuss two types of selection problems. The first problem is motivated by labor market analyses such as the estimation of sector-specific wage equations where the sector for which the wages are observed are chosen by the agents. In contrast to previous formulations which usually are based on a probit framework, we assume here that the discrete choice is generated by a multinomial logit model with random coefficients (mixed multinomial logit model). The advantage compared to the multinomial probit setting is that choice sets with many alternatives become almost as easy to handle as the binary case.

The second problem we analyze is motivated by studies where the interest is to estimate the effect of for example labor market training programs on the labor market opportunities. Previous works have, to the best of my knowledge, focused solely on the effect of labor market programs on earnings. As in the first case we allow for arbitrarily large choice sets of feasible first stage choices (programs) as well as the second stage choices (labor market status).

Keywords: Selection bias, discrete/continuous choice.

JEL classification: C13, C35

Address: John K. Dagsvik, Statistics Norway, Research Department. E-mail: john.dagsvik@ssb.no

Discussion Papers comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1. Introduction

Ever since the seminal papers by Heckman (1973, 1979) the econometric methodology for dealing with different types of selection problems has developed rapidly. Two types of selection problems that arise in labor market research have been particularly popular. The first one is related to the estimation of regression equations conditional on endogenous discrete choices, similarly to Roy (1951), Willis and Rosen (1979), Heckman and Sedlacek (1985, 1995), Lee (1983), Hanemann (1984), and Dubin and McFadden (1984). In this paper we demonstrate that when the choice model is assumed to be a random coefficient multinomial logit model (mixed multinomial logit model) the analysis becomes (under additional assumptions) simple even with a large number of sectors, in contrast to the case where the discrete choices are generated by a multinomial Probit model. More generally, this approach should be of interest in other discrete/continuous choice settings.

The motivation for the second type of selection problems we consider stems from the literature on the effect on earnings and labor market opportunities as a result of participation in educational- and training programs. While this literature has focused on the effect on earnings, we are in this paper interested in the possible selection problem that arises when one wishes to analyze the effect on labor market opportunities as a result of training. Abstracting from this particular applicaiton, a general description of the second type of selection problem we analyze is as follows: The agent faces a two stage choice problem: In stage one he chooses from a choice set of mutually exclusive alternatives (training programs). Conditional on the choice in the first stage he chooses from a choice set of second stage alternatives. The question of interest is whether or not the choice in the first stage has increased the second stage choice set. Since some of the choice opportunities are latent and possibly correlated with the choice made in the first stage, a selection problem arises. As in the first case one could also apply a multinomial Probit setting in this case. This would, however, not lead to the same interprtation as in this paper and it would also increase the complexity of the modeling framework.

The paper is organized as follows. In Section 2 we discuss the selection bias problem in a regression model, where the dependent variable depends on a choice variable that is generated by a mixed multinomial logit model. In Section 3 we similarly discuss the selection bias problem when the (first stage) dependent variable is a multinomial logit choice variable and the corresponding choice set is partially latent and correlated with a first stage multinomial choice variable.

2. The selection bias in regression models conditioned on a choice variable generated by a mixed multinomial Logit model

We consider a setting where the worker faces the choice to work in one out of m feasible sectors—or regions. For simplicity we suppress the indexation of the individual agent in the notation. The reduced form utility function of worker i of working in sector j has the form

(2.1) $$U_j = Z_j \bar{\beta} + Z_j \beta + \varepsilon_j ,$$

where Z_j is a vector of sector and individual-specific attributes. For example, the components of Z_j may be pure alternative specific attributes and products of alternative-specific attributes and individual characteristics. $\bar{\beta}$ is a K-dimensional vector of coefficients and β is a K-dimensional vector of individual specific coefficients which are assumed to be random and independent of the attribute vectors and with zero mean. Some of the components of β may be alternative-specific, but we shall not include this in the formalism. It will, however, be rather obvious how alternative-specific coefficients can be accounted for in the analysis below. The terms ε_j, $j = 1,2,\ldots,m$, are assumed to be i.i.d. with cumulative distribution function

(2.2) $$P\left(\varepsilon_j \le y \mid \beta, \{Z_k\}\right) = \exp\left(-e^{-y}\right), \ y \in R .$$

The regression equation is given as

(2.3) $$\log W_j = X\gamma_j + \eta_j$$

where W_j is the dependent variable which depends on j, X is a vector of individual- and possibly sector specific variables, γ is a vector of coefficients and η_j are random variables which distribution is independent of the X-variables. The following Lemma is well known and easily demonstrated.

Lemma 1

Assume that $U_j = v_j + \varepsilon_j$, *where* ε_j, $j = 1,2,\ldots,m$, *are i.i.d. with distribution function*

$$P\left(\varepsilon_j \le y \mid \{v_k\}\right) = exp\left(-e^{-y}\right),$$

$y \in R$, *and* v_j, $j = 1,2,...,m$, *are deterministic terms. Then*

(2.4) $$P\left(U_j \le y \mid U_j = max_k U_k\right) = P\left(max_k U_k \le y \mid U_j = max_k U_k\right) = P\left(max_k U_k \le y\right).$$

Loosely speaking, Lemma 1 states that the distribution of the indirect utility (maximum utility) is independent of which alternative maximum utility is achieved. For the readers' convenience, a proof is given in the appendix.

We shall next allow η_j to be correlated with ε_j and the components of β. Assume as a first order approximation that

(2.5) $$\eta_j = \rho_1\left(\varepsilon_j - 0.5772\right) + \sum_r \rho_{2jr} \beta_r + e_j$$

where e_j is independent of ε_j and β and β_r is the r-th component of β. It follows from (2.5) that

$$\rho_1 = \frac{6\,\mathrm{Cov}\left(\eta_j, \varepsilon_j\right)}{\pi^2} \tag{2.6}$$

and

$$\rho_{2r} = \frac{\mathrm{Cov}\left(\eta_j, \beta_r\right)}{\mathrm{Var}\,\beta_r}. \tag{2.7}$$

At this point we may note that in a Probit type of framework the decomposition (2.5) is guaranteed due to the properties of the multivariate normal distribution. In the general case it is of course not so.

As in Dubin and McFadden (1984), it may in some applications be of interest to allow η_j also to depend on $\{\varepsilon_k, k \neq j\}$. This is, however, ruled out in the present paper. The results obtained here can easily be extended to include this case as well.

Let

$$P_j(\beta) \equiv \frac{\exp\left(Z_j\beta + Z_j\bar{\beta}\right)}{\sum_{k=1}^{m} \exp\left(Z_k\beta + Z_k\bar{\beta}\right)}. \tag{2.8}$$

We can now prove the following result.

Proposition 1

Suppose that ε_j, $j = 1,2,\ldots,m$, are i.i.d. and that (2.2) and (2.5) hold. Then

$$E\left(\beta_r \,\middle|\, U_j = \max_k U_k\right) = \frac{E\left(\beta_r\, P_j(\beta)\right)}{E\, P_j(\beta)} \tag{2.9}$$

and

$$E\left(\varepsilon_j - 0.5772 \,\middle|\, U_j = \max_k U_k\right) = -E\log P_j(\beta) - \sum_r Z_{jr} E\left(\beta_r \,\middle|\, U_j = \max_k U_k\right). \tag{2.10}$$

A proof of Proposition 1 is given in the appendix.

From (2.5) and Theorem 1 we get

(2.11) $$E\left(\eta_j \,\middle|\, U_j = \max_k U_k\right) = -\rho_1 \, E \log P_j(\beta) + \frac{\sum_r \left(\rho_{2r} - Z_{jr}\right) E\left(\beta_r \, P_j(\beta)\right)}{E\, P_j(\beta)}.$$

But this means that η_j^*, defined by

(2.12) $$\eta_j^* = \log W_j - X\gamma_j + \rho_1 \, E \log P_j(\beta) - \frac{\sum_r \left(\rho_{2r} - Z_{jr}\right) E\left(\beta_r \, P_j(\beta)\right)}{E P_j(\beta)}$$

has the property

$$E\left(\eta_j^* \,\middle|\, U_j = \max_k U_k\right) = 0.$$

Thus, if $E \log P_j(\beta)$, $E\beta P_j(\beta)$ and $E\, P_j(\beta)$ were known one could estimate γ, ρ_1 and $\{\rho_{2r}\}$ consistently by means of the regression equation

(2.13) $$\log W_j + \frac{\sum_r Z_{jr} \, E\left(\beta_r P_j(\beta)\right)}{E P_j(\beta)} = X\gamma_j - \rho_1 \, E \log P_j(\beta) + \frac{\sum_r \rho_{2r} \, E\left(\beta_r P_j(\beta)\right)}{E P_j(\beta)} + \eta_j^*.$$

Estimation

The choice probabilities $\bar{P}_j$ are given by

(2.14) $$\bar{P}_j \equiv P\left(U_j = \max_k U_k\right) = E\, P\left(U_j = \max_k U_k \,\middle|\, \beta\right) = E\, P_j(\beta).$$

The simplest version of the model is obtained when $\beta_r = \sigma_r \theta_r$, where $\sigma_r > 0$ and θ_r, $r = 1,2,\ldots,\ldots,K$ are i.i.d. with c.d.f. that is independent on $\{\sigma_r\}$. Then a simple way to compute $E\, P_j(\beta)$ (conditional on $\{\sigma_r\}$) is to draw S independent θ_r^s, $s = 1,2,\ldots,S$, and then for each given $\bar{\beta}$ and $\{\sigma_r\}$ simulate $E\, P_j(\beta)$ by

(2.15) $$E\, P_j(\beta) \approx \frac{1}{S} \sum_{s=1}^{S} P_j\left(\beta^s\right)$$

where $\beta_r^s = \sigma_r \theta_r^s$, $r = 1,2,\ldots,K$. For a detailed discussion of the estimation of mixed multinomial logit models, see McFadden and Train (1998).

In the same way one can simulate $E\left(\beta_r \, P_j(\beta)\right)$ and $E \log P_j(\beta)$. Estimation can now be carried out in two stages:

Stage one:

Estimate $\{\sigma_r\}$ and $\bar{\beta}$ on the basis of $\bar{P}_j$.

Stage two:

In addition to $\bar{P}_j$, compute $E\left(\beta_r\, P_j(\beta)\right)$ and $E \log P_j(\beta)$. Then estimate γ, ρ_1 and $\{\rho_{2r}\}$ on the basis of (2.13) with

$$\log W_j + \frac{\sum_r Z_{jr}\, E\left(\beta_r\, P_j(\beta)\right)}{E\, P_j(\beta)}$$

as dependent variable and X_j,

$$E \log P_j(\beta) \quad \text{and} \quad \frac{E\left(\beta_r\, P_j(\beta)\right)}{E\, P_j(\beta)}$$

as independent variables.

Note that when $\beta_r = \sigma_r \theta_r$ we get

$$\sum_r Z_{jr}\, E\left(\beta_r\, P_j(\beta)\right) = E\left(\theta_1\, P_j(\beta)\right) \sum_r Z_{jr}\, \sigma_r \tag{2.16}$$

and

$$\sum_r \rho_{2r}\, E\left(\beta_r\, P_j(\beta)\right) = E\left(\theta_1\, P_j(\beta)\right) \sum_r \rho_{2r}\, \sigma_r\,. \tag{2.17}$$

We realize that one cannot in this case identify $\{\rho_{2r}\}$. But this is not of great interest anyway. Accordingly, it is in this case sufficient to specify two independent variables in addition to X_j, namely $E \log P_j(\beta)$ and

$$\frac{E\left(\theta_1\, P_j(\beta)\right)}{E\, P_j(\beta)}$$

with γ, ρ_1 and $\rho_3 \equiv \sum_r \rho_{2r}\sigma_r$ as unknown parameters. In this case the regression equation (2.13) reduces to

$$\log W_j + \frac{E\left(\theta_1 P_j(\beta)\right)}{EP_j(\beta)} \sum_r Z_{jr}\, \sigma_r = X\gamma_j - \rho_1 E \log P_j(\beta) + \frac{\rho_3 E\left(\theta_1 P_j(\beta)\right)}{EP_j(\beta)} + \eta_j^*. \tag{2.18}$$

A special case

Consider now the special case when $\mathrm{Var}\,\beta_r = 0$, i.e., there are no variations in the parameters in the utility function across individuals. Then $\rho_{2r} = 0$ for all r and (2.18) reduces to

$$\log W_j = X\gamma_j - \rho_1 \log P_j(0) + \eta_j^*. \tag{2.19}$$

3. The selection bias in a two-stage multinomial discrete choice setting where the choice in the second stage is correlated with the choice in the first stage

In this section we discuss a fairly simple framework that can be applied for structural analysis of the following general setting: We consider an agent that faces a two stage choice problem. In the first stage the agent makes a choice from a set of "schooling" (training) alternatives. In the second stage, and after the chosen schooling alternative has been made, the agent chooses between a set of different types of opportunities (jobs). We can interpret the choice in the first stage as one in which the agent maximizes expected utility with respect to "investment" in human capital. Investment in human capital in stage one is assumed by the agent to increase the available opportunities in stage two.

A typical example of the above setting is found in the so-called job-training literature. In this case the agent chooses a specific job-training program (or is allocated to a program by a central manager) in stage one. In stage two he chooses from a set of feasible jobs.

Evidently, one can apply a two period multinomial probit framework to formulate choice models of this type. The multinomial multiperiod model is, however, still complicated to estimate for the average researcher. More fundamentally, the approach proposed here has, in addition of being computationally simple, the advantage compared to the probit framework in that it enables us to accommodate the notion of endogenous and latent second stage choice set in an explicit manner.

Let us next introduce some theoretical concepts and formalism. For notational simplicity we drop the indexation of the agent in the notation. Let U_j be the utility of the agent of alternative j, where j belongs to the individuals choice set B. The choice set may be individual specific, but this is suppressed in the notation here. We assumed that $\{U_j\}$ has the structure

$$U_j = u_j + \varepsilon_j \tag{3.1}$$

where u_j is a systematic term that depends on individual and possible alternative-specific variables. The terms $\varepsilon_j, j = 1, 2, ..., N,$ are independent random variables with

$$P\left(\varepsilon_j \le y \,\middle|\, \{u_j\}\right) = \exp\left(-e^{-y+0.5772}\right) \tag{3.2}$$

for $y \in R$. Assumption (3.2) implies that the corresponding first stage choice probabilities fulfill the assumption, "Independence from Irrelevant Alternatives".

Consider next stage two. In stage two there are K observable types of choice opportunities which each consists of a set of latent opportunities. Let J denote the chosen alternative in stage one, and let C_{Jk} denote the agent-specific index set of opportunities of type k, and let V_{ks} be the agent's utility of opportunity s in C_{Jk} in stage two. A typical example of this setup is that C_{Jk} is the set of jobs of type k and V_{ks} is the utility of job s of type k. Let m_{Jk} be the number of opportunities in C_{Jk}.

Assume moreover that

(3.3) $$V_{ks} = v_k + \eta_{ks}$$

for $s \in C_{Jk}, k = 1, 2, ..., K$, where v_k is a systematic term that may depend on individual characteristics as well as observable attributes that characterize opportunities of type k. The term η_{ks} is random and account for unobservables that affect the utility of opportunity s in C_{Jk}. Similarly to stage one the random variables $\{\eta_{ks}\}$ are assumed i.i.d. with

(3.4) $$P\left(\eta_{ks} \le y \middle| \{v_k\}\right) = \exp\left(-e^{-y}\right)$$

for $y \in R$.

As mentioned above, the observing analyst does, however, not observe the choice sets $\{C_{Jk}\}$ nor does he observe which opportunity (job) the agent chooses. He only observes the chosen opportunity. Let

(3.5) $$\tilde{V}_{Jk} = \max_{s \in C_{Jk}} V_{ks}.$$

Note that $\tilde{V}_{Jk}$ is the highest utility the agent can attain conditional on the choice set C_{Jk}, and $\{\tilde{V}_{Jk}\}$ are therefore the utilities that correspond to the actual *observed* choices. It follows readily from (3.4) that $\tilde{V}_{Jk}$ can be expressed as

(3.6) $$\tilde{V}_{Jk} = v_k + \log m_{Jk} + \tilde{\eta}_{Jk}$$

where $\{\tilde{\eta}_{Jk}\}$ are i.i.d. with

(3.7) $$P\left(\tilde{\eta}_{Jk} \le y \middle| \{m_{Jk}\}, \{v_k\}\right) = \exp\left(-e^{-y}\right)$$

for $y \in R$.

The variables $\{m_{Jk}\}$ are unobservable. They are presumably endogeneous in the sense that they depend upon the "investment" that was made in the first stage. It follows from (3.2) that the indirect utility U_J of stage one can be written as

$$U_J = \log\left(\sum_{r\le N} e^{u_r}\right) + \varepsilon_J^* \tag{3.8}$$

where ε_J^* is a random variable with conditional distribution

$$P\left(\varepsilon_J^* \le y \,\middle|\, \{u_r\}\right) = \exp\left(-e^{-y}\right) \tag{3.9}$$

for $y \in R$. Let alternative one be a reference alternative in the first stage. Then $U_J - U_1$ represents the highest predicted value (as of stage one) of human capital attainable (to the agent). In the second stage $U_J - U_1$ can thus be viewed as a proxy for the information about the human capital acquired through the behavior in stage one.

Thus, we are lead to postulate the following relationship:

$$\log m_{Jk} = \theta_{Jk}\left(U_J - U_1\right) + b_k \tag{3.10}$$

with $\theta_{J1} = \theta_{1k} = 0$. The choice set C_1 is assumed not to be affected by the choice in the first stage. The parameter θ_{Jk} determines to which extent the increase in indirect utility, $U_J - U_1$, is relevant for the opportunity set C_k. The motivation for the structure (3.10) is that since $U_J - U_1$ can be interpreted as the highest value of the investment in human capital, $U_J - U_1$ is a sufficient statistic for the effect of the first stage choice.

Now from (3.1) and (3.2) follows that

$$P_j \equiv P\left(U_j = \max_{r\le N} U_r\right) = \frac{e^{u_j}}{\sum_{r\le N} e^{u_r}} . \tag{3.11}$$

Proposition 2:

Under Assumptions (3.2), (3.4) and (3.10) we have

$$P\left(\tilde{V}_{jk} = \max_{r\le K} \tilde{V}_{jr} \,\middle|\, U_j = \max_{r\le N} U_r\right) = E\left(\frac{exp\left(v_k + b_k + \theta_{jk}\xi_j\right)}{\sum_{r\le K} exp\left(v_r + b_r + \theta_{jr}\xi_j\right)}\right) \tag{3.12}$$

where ξ_j *is a random variable with distribution function*

(3.13) $$P\left(\xi_j \le y \middle| \left\{\mathrm{v}_k + b_k\right\},\ \left\{u_k\right\}\right) \equiv P\left(U_J - U_1 \le y \middle| J = j\right) = 1 - \frac{1}{1 - P_1 + P_1 e^{-y}}$$

for $y \ge R$.

A proof of Proposition 2 is given in the appendix.

Corollary 1

The distribution of ξ_j *can be expressed as*

$$P\left(\xi_j \le y\right) = \frac{1}{1 - P_1} P\left(max\left(\delta_j + log\left(\frac{1 - P_1}{P_1}\right), 0\right) \le y\right)$$

where δ_j *is a random variable with c.d.f.*

$$P\left(\delta_j \le y\right) = \frac{1}{1 + e^{-y}}$$

for $y \in R$.

The proof of this result is straight forward.

The interest of Corollary 1 is that it is useful for computing the second stage choice probabilities by stochastic simulation, due to the fact that the distribution of δ_j does not depend on parameters nor variables in the model.

Corollary 2

Under the assumptions of Proposition 2 a first order approximation of the second stage conditional choice probabilities is given by

$$P\left(\tilde{V}_{jk} = \max_{r \le K} \tilde{V}_{jr} \middle| U_j = \max_{r \le N} U_r\right) \approx \frac{exp\left(\mathrm{v}_k + b_k + \theta_{jk} Z\right)}{\sum_{r \le K} exp\left(\mathrm{v}_r + b_r + \theta_{jr} Z\right)}$$

where

$$Z = -\frac{log\, P_1}{1 - P_1}.$$

A proof of Corollary 2 is given in the appendix.

A weaker assumption than (3.10)

In the discussion above we postulated that the choice set in stage two depends on $U_J - U_1$. We shall now relax this assumption. Specifically, assume now that it makes sense to postulate

$$\text{(3.14)} \qquad \log m_k = \theta_{Jk}(\varepsilon_J - \varepsilon_1) + b_k + \gamma_{Jk}$$

where $\{\theta_{Jk}\}$ are unknown parameters with $\theta_{J1} = 0$. The parameter θ_{Jk} determines to which extent the unobservables associated with the first stage investment in human capital is relevant for the opportunity set C_k. The parameters $\{\gamma_{jk}\}$ account for the systematic effect from the first stage investment. Specifically, γ_{jk} captures the systematic (average) effect of first stage decision on the opportunity set C_k, given that $J = j$.

We have the following result

Proposition 3

If (3.1) to (3.4), and (3.14) hold, we have that

$$\text{(3.15)} \qquad P\left(\tilde{V}_{jk} = \max_{r \le K} \tilde{V}_{jr} \,\middle|\, U_j = \max_{r \le N} U_r\right) = E\left(\frac{exp\left(v_k + b_k + \gamma_{jk} + \theta_{jk}\xi_j^*\right)}{\sum_{r \le K} exp\left(v_r + b_r + \gamma_{jr} + \theta_{jr}\xi_j^*\right)}\right)$$

where ξ_j^* is a random variable with distribution function

$$P\left(\xi_j^* \le y\right) = 1 - \frac{1}{1 - P_1 + P_1 e^{y + u_j - u_1}}.$$

A proof of Proposition 3 is given in the Appendix.

Corollary 3

Under the assumptions of Proposition 3 a first order approximation of the second stage conditional choice probabilities is given by

$$P\left(\tilde{V}_{jk} = \max_{r \le K} \tilde{V}_{jr} \,\middle|\, U_j = \max_{r \le N} U_r\right) \approx \frac{exp\left(v_k + b_k + \gamma_{jk} + \theta_{jk} Z_j^*\right)}{\sum_{r \le K} exp\left(v_r + b_r + \gamma_{jr} + \theta_{jr} Z_j^*\right)}$$

where

$$Z_j^* = -\frac{1}{1-P_l} log\left(\frac{P_j}{1-P_l+P_j}\right).$$

Proof:

Similarly to the proof of Corollary 2, the result follows from (A.15) in the appendix.

Q.E.D.

Eq. (3.12) implies that we can compute Q_{jk} by simulation as follows: Draw M independent realizations, $\delta_j^1, \delta_j^2, ..., \delta_j^M$, from the logistic distribution and compute the corresponding values $\xi_j^1, \xi_j^2, ..., \xi_j^M$, by Corollary 1. Then

(3.16)
$$Q_{jk} \approx \tilde{Q}_{jk} \equiv \frac{1}{M}\sum_{s=1}^{M} \frac{1}{1-P_l} \frac{\exp\left(v_k + b_k + \theta_{jk}\xi_j^s\right)}{\sum_{r \le K} \exp\left(v_r + b_r + \theta_{jr}\xi_j^s\right)}, \quad \theta_{1k} = \theta_{j1} = 0.$$

Estimation

Here we need to introduce the indexation of the agents. Let $Y_{jk}^i = 1$ if individual i has chosen alternative j in stage one and alternative k in stage two. Then

(3.17)
$$P\left(Y_{jk}^i = 1\right) = Q_{jk}^i \cdot P_j^i \approx \widetilde{Q}_{jk}^i \cdot P_j^i .$$

Consequently, the corresponding loglikelihood equals

(3.18)
$$\ell = \sum_i \sum_k \sum_j Y_{jk}^i \left(\log Q_{jk}^i + \log P_j^i\right) \approx \sum_i \sum_k \sum_j Y_{jk}^i \left(\log \widetilde{Q}_{jk}^i + \log P_j^i\right).$$

Estimation in two stages

Stage one: Estimate P_j^i by maximum likelihood.

Stage two: Estimate the remaining parameters in Q_{jk} by maximum likelihood. The second stage likelihood equals

(3.19)
$$\ell^* = \sum_i \sum_k \sum_j Y_{jk}^i \log Q_{jk}^i \approx \sum_i \sum_k \sum_j Y_{jk}^i \log \widetilde{Q}_{jk}^i .$$

4. Conclusion

In this paper we have analyzed two typical selectivity bias problems. The first one is related to the estimation of regression equations conditional on endogenous multinomial discrete choice variables. Specifically, we consider the case where the discrete choices are generated by a mixed multinomial logit model. In the second part we analyze a two stage choice problem where the dependent variable in each stage is discrete. In both cases we demonstrate that the respective selectivity bias problem can be taken into account in a simple way.

References

Dubin, J. and D. MacFadden (1984): An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica*, **52**, 345-362.

Hanemann, M.W. (1984): Discrete/Continuous Models for Consumer Demand. *Econometrica*, **52**, 541-561.

Heckman, J.J. and G. Sedlacek (1985): Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market. *Journal of Political Economy*, **93**, 1077-1125.

Heckman, J.J. and G. Sedlacek (1995): Self-Selection and the Distribution of Hourly Wages. *Journal of Labor Economics*, **8**, S329-S363.

Lee, L.F. (1983): Generalized Econometric Models with Selectivity. *Econometrica*, **51**, 507-512.

McFadden, D. and K. Train (1998): Mixed MNL Models for Discrete Responses. Mimeo, Department of Economics, University of California, Berkeley.

Roy, A.D. (1951): Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, **3**, 135-146.

Willis, R.J. and S. Rosen (1979): Education and Self-Selection. *Journal of Political Economy*, **87**, S7-S36.

Appendix

Proof of Lemma 1:

We have

$$P\left(U_j > \max_{k\neq j} U_k, \max_k U_k \in (y, y+\Delta y) \right)$$

(A.1)
$$= P\left(U_j \in (y, y+\Delta y), \max_{k\neq j} U_k < y \right) + o(\Delta y)$$
$$= P\left(U_j \in (y, y+\Delta y)\right) \prod_{k\neq j} P\left(U_k < y\right) + o(\Delta y)$$
$$= \exp\left(-e^{v_j - y}\right) \cdot e^{v_j - y} \Delta y \exp\left(-e^{-y} \sum_{k\neq j} e^{v_k} \right) + o(\Delta y)$$
$$= e^{v_j - y} \exp\left(-e^{-y} \sum_k e^{v_k}\right) + o(\Delta y).$$

Since

(A.2)
$$P\left(\max_k U_k \leq y\right) = \exp\left(-e^{-y} \sum_k e^{v_k}\right)$$

it follows from (A.1) that

$$P\left(U_j > \max_{k\neq j} U_k, \max_k U_k \in (y, y+dy) \right)$$

(A.3)
$$= \frac{e^{v_j}}{\sum_k e^{v_k}} \cdot e^{-y} \sum_k e^{v_k} \exp\left(-e^{-y} \sum_k e^{v_k}\right) + o(\Delta y)$$
$$= P_j \cdot P\left(\max_k U_k \in (y, y+\Delta y)\right) + o(\Delta y).$$

But this implies that

$$P\left(U_j > \max_{k\neq j} U_k \mid \max_k U_k = y \right) = P_j .$$

Q.E.D.

Proof of Proposition 1:

Let $f(\beta)$ be the density of β. We have for $x \in R^K$:

(A.4)
$$P\left(\beta \in (x, x+dx), U_j = \max_k U_k\right) = P\left(U_j = \max_k U_k \mid \beta = x\right) f(x)\,dx = P_j(x)\, f(x)\,dx .$$

Consequently,

(A.5)
$$P\left(\beta \in (x, x+dx) \,\middle|\, U_j = \max_k U_k\right) = \frac{P_j(x)\, f(x)\, dx}{E\, P_j(\beta)}.$$

From (A.5) we immediately obtain that

(A.6)
$$E\left(\beta_r \,\middle|\, U_j = \max_k U_k\right) = \frac{E\left(\beta_r\, P_j(\beta)\right)}{E\, P_j(\beta)}.$$

Now by Lemma 1 and (A.6) we obtain

(A.7)
$$\begin{aligned}
E\left(\varepsilon_j \,\middle|\, U_j = \max_k U_k\right) &= E\left(U_j \,\middle|\, U_j = \max_k U_k\right) - E\left(Z_j \beta \,\middle|\, U_j = \max_k U_k\right) - Z_j \bar{\beta} \\
&= E\, E\left(U_j \,\middle|\, U_j = \max_k U_k, \beta\right) - \sum_r Z_{jr}\, E\left(\beta_r \,\middle|\, U_j = \max_k U_k\right) - Z_j \bar{\beta} \\
&= E\, E\left(\max_k U_k \mid \beta\right) - \frac{\sum_r Z_{jr}\, E\left(\beta_r\, P_j(\beta)\right)}{E\, P_j(\beta)} - Z_j \bar{\beta} \\
&= E \log\left(\sum_k \exp\left(Z_k \beta + Z_k \bar{\beta}\right)\right) + 0.5772 - Z_j \bar{\beta} - \frac{\sum_r Z_{jr}\, E\left(\beta_r\, P_j(\beta)\right)}{E\, P_j(\beta)}.
\end{aligned}$$

But since

(A.8)
$$E \log\left(\sum_k \exp\left(Z_k \beta + Z_k \bar{\beta}\right)\right) = -E \log P_j(\beta) + Z_j \bar{\beta}$$

(A.7) implies that we can write

(A.9)
$$E\left(\varepsilon_j - 0.5772 \,\middle|\, U_j = \max_k U_k\right) = -E \log P_j(\beta) - \frac{\sum_r Z_{jr}\, E\left(\beta_r\, P_j(\beta)\right)}{E P_j(\beta)}$$

which was to be proved

Q.E.D.

Proof of Proposition 2:

Consider

$$P\left(U_J - U_1 > y \,\middle|\, U_j = U_J\right)$$

for $j \in \{2,3,...,N\}$ and $y \geq 0$. Let J^* denote the preferred alternative within $\{2,3,...,N\}$. Then we can write

(A.10)
$$P\left(U_J > y + U_1, U_j = U_J\right) = P\left(\max\left(U_{J^*}, U_1\right) > y + U_1, U_j = U_{J^*}, U_{J^*} > U_1\right)$$
$$= P\left(U_{J^*} > y + U_1, J^* = j, U_{J^*} > U_1\right) = P\left(U_{J^*} > y + U_1, J^* = j\right).$$

From Lemma 1 it follows that J^* and U_{J^*} are independent. Furthermore, U_1 and J^* are independent due to the assumption that $\varepsilon_1, \varepsilon_2, \ldots$, are independent, and J^* does not depend on ε_1. Hence, we obtain

(A.11)
$$P\left(U_{J^*} > U_1 + y, J^* = j\right) = P\left(U_{J^*} > U_1 + y\right) P\left(J^* = j\right).$$

Moreover,

(A.12)
$$P\left(U_{J^*} > U_1 + y\right)\; P\left(J^* = j\right) = \frac{\sum_{r \le N} e^{u_r} - e^{u_1}}{\sum_{r \le N} e^{u_r} - e^{u_1} + e^{u_1 + y}} \cdot \frac{e^{u_j}}{\sum_{r \le N} e^{u_r} - e^{u_1}}$$
$$= \frac{1 - P_1}{1 - P_1 + P_1 \; e^y} \cdot \frac{P_j}{1 - P_1} = \frac{P_j}{1 - P_1 + P_1 \; e^y} \quad .$$

Hence we obtain that

(A.13)
$$P\left(U_J > U_1 + y \,\middle|\, J = j\right) = \frac{P\left(U_J > U_1 + y, U_j = U_J\right)}{P\left(U_j = U_J\right)} = \frac{1}{1 - P_1 + P_1 e^y}$$

for $y > 0$. Let δ_j be a random variable with c.d.f.

(A.14)
$$P\left(\delta_j \le y\right) = \frac{1}{1 + e^{-y}}$$

for $y \in R$. Then the result of Proposition 2 follows.

This completes the proof.

Q.E.D.

Proof of Corollary 2:

From (A.13) in the appendix we obtain that

$$E\left(U_J - U_1 \,\middle|\, J = j\right) = \int_0^\infty \frac{dy}{1 - P_1 + P_1 e^y} = -\Big|_0^\infty \frac{1}{1 - P_1} \log\left(P_1 + \left(1 - P_1\right) e^{-y}\right) = -\frac{\log P_1}{1 - P_1}.$$

Q.E.D.

Proof of Proposition 3:

Similarly to the proof of Proposition 2 we have that

(A.15)

$$\begin{aligned} & P(\varepsilon_J - \varepsilon_1 > y \mid J = j) = P(\varepsilon_j - \varepsilon_1 > y \mid J = j) \\ & = P(U_j - U_1 > y + u_j - u_1 \mid J = j) = P(U_J - U_1 > y + u_j - u_1 \mid J = j) \\ & = \frac{P(U_{J^*} > U_1 + y + u_j - u_1)}{P(U_{J^*} > U_1)} = \frac{1}{1 - P_1 + P_1 e^{y + u_j - u_1}} \end{aligned}$$

for $y > u_1 - u_j$. The results of Proposition 3 now follows.

Thus, the proof is complete.

Q.E.D.