Working Paper Series

Economics

No. 1802

http://shss.nu.edu.kz/shss/academics/departments/economics

November 2018

EDGEWORTH EXPANSION BASED CORRECTION OF SELECTIVITY BIAS IN MODELS OF DOUBLE SELECTION

Insan Tunali[1] and Berk Yavuzoglu[2]

[1] Department of Economics, Koc University, Istanbul, Turkey
[2] Department of Economics, Nazarbayev University, Astana, Kazakhstan
(c) copyright 2018, Insan Tunali & Berk Yavuzoglu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

EDGEWORTH EXPANSION BASED CORRECTION OF SELECTIVITY BIAS IN MODELS OF DOUBLE SELECTION*

Insan Tunali[†], Berk Yavuzoglu[‡]

November 2, 2018

Abstract

Edgeworth expansions are known to be useful for approximating probability distributions and moments. In our case, we exploit the expansion in the context of models of double selection embedded in a trivariate normal structure. We assume bivariate normality among the random disturbance terms in the two selection equations but allow the distribution of the disturbance term in the outcome equation to be free. This sets the stage for a control function approach to correction of selectivity bias that affords tests for the more common trivariate normality specification. Other recently proposed methods for handling multiple outcomes are Multinomial Logit based selection correction models. An empirical example is presented to document the differences among the results obtained from our selectivity correction approach, trivariate normality specification and Multinomial Logit based selection correction models.

Keywords: double selection models; Edgeworth expansion; female labor supply; Multinomial Logit based selection correction models; selectivity bias.

JEL codes: C34, C35, C63.

---

*We are grateful to Elvan Ceyhan for his detailed input to the paper. Feedback form Francis Vella and John Kennan are gratefully acknowledged.

[†]Department of Economics, Koc University; e-mail: itunali@ku.edu.tr

[‡]Corresponding author, Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, USA. Tel.: +1 608 263 2327; Fax: +1 608 608 262 2033; e-mail: yavuzoglu@wisc.edu

1 Introduction

Since the early work on selectivity pioneered by Heckman (1976, 1979) and Lee (1976, 1979), the control function approach has remained as the most popular and versatile method in parametric as well as semi-parametric models of selection. Unlike parametric models that can be reconciled with optimizing behavior, semi-parametric models do not invoke a behavioral interpretation of the control function. Their appeal rests in their ability to circumvent the misspecification problems associated with the strong parametric assumptions. When multiple sources of selection are present, semi-parametric approaches require additional assumptions. These in turn impose restrictions on the behavioral model, and usher in a new set of tradeoffs in gauging the merits of parametric vs. semi-parametric methods.

In this paper we return to the double selection version of the Heckman-Lee parametric framework which is studied by Tunali (1986). To solve the problem, Tunali (1986) uses a common choice structure where he assumes trivariate normality among the random disturbances of the two selection equations and the regression (partially observed outcome) equation. The present paper uses the same common structure but relaxes the trivariate normality assumption following the Edgeworth expansion approach of Lee (1982). In obtaining our correction, we do not impose any condition on the form of the distribution of the random disturbance in the regression equation, but conveniently assume bivariate normality between the random disturbances of the two selection equations. Although quasi-maximum likelihood methodology offers some limited justification for defending this distributional assumption, we do not pursue the more appealing alternatives developed in Ruud (1983) and Stoker (1986).

As documented in Miller (2011), Edgeworth expansion was introduced by Edgeworth (1905) and dubbed "Edgeworth's series" by Elderton (1906). H. Cramer established the theory of "Edgeworth's series" in the 1920s and summarized it later in Cramer (1937). The presently popular term Edgeworth expansion was used for the first time by David et al. (1951). The topic has received considerable attention in statistics and econometrics. A sampling of the contributions may be found in Bhattacharya and Rao (1976), Bhattacharya and Ghosh (1978), Rothenberg (1983) and Kolassa and McCullagh (1990).

A double selection problem may arise due to interdependent choices of distinct decision makers or due to a simultaneous or sequential decision undertaken by a single party. Although the problem can be tackled by employing other methodologies, we opt for Edgeworth expansion because it can be situated within a behavioral model and yields a richer functional form for selectivity correction. In particular skewed and kurtic distributions can be handled with ease. This form nests the version obtained under trivariate normality and provides a test of that restriction. We illustrate our Edgeworth expansion based correction for double selection in the empirical context of Tunali and Baslevent (2006), where the goal is to estimate a wage equation for married women subject to participation and mode of employment choices. Since less than 14 percent of the married women participate and an even smaller subset works for wages, selectivity is likely to be a serious problem. We also briefly revisit the empirical context in Tunali (1986) to address computational issues that arise during estimation.

Multinomial Logit based selection correction models have also been developed for the purpose of handling two or more selection equations [see Lee (1983), Dubin and McFadden (1984), Dahl (2002)]. These models imply severe restrictions on behavior because the random disturbances of the selection equations are assumed to be independent. Our model can only handle two selection equations but it has the advantage of allowing interdependence. Another advantage of our model is the ease of accommodation of additional structure in the behavioral model. In the example we pursue in some detail there are four outcomes: nonparticipation, self-employement, wage employment and unemployment. In effect unemployed individuals have opted for participation, but their employment mode is not known. This structure can easily be handled in our model, but not in a Multinomial Logit based alternative. Nevertheless we present the results obtained from some well-known Multinomial Logit based selection models to gauge the appeal of the various models empirically.

Das et al. (2003) propose the use of series expansions along the lines of Ahn and Powell (1993) under multiple selection. These models exploit power series and smooth piecewise polynomials in the propensity scores (conditional choice probabilities) which can be nonparametrically estimated. To our knowledge Dahl (2002) is the only paper that has implemented this idea. We discuss his methodology in some detail and apply a version of it in our empirical investigation.

In Section 2, we develop the theoretical framework. We start with a generic definition of the double selection problem, then discuss the our parametric solution strategy and its merits. We then introduce the Trivariate Edgeworth expansion, outline the derivation of the selectivity correction terms and describe the estimation procedure. The formal derivations are collected in the Appendix. In Section 3, we provide an empirical example which illustrates our selectivity correction methodology. In Section 4, we revisit the example using alternative approaches relying on Multinomial Logit distribution. We conclude in Section 5 with a brief summary and some closing remarks.

2 Theoretical Framework

2.1 Double Selection Problem

We take as our point of departure the following generic representation of the double selection problem:

$$\begin{aligned} y_{1i}^* &= \beta_1' X_{1i} + \sigma_1 u_{1i} \qquad \text{(first selection rule)}, && (1) \\ y_{2i}^* &= \beta_2' X_{2i} + \sigma_2 u_{2i} \qquad \text{(second selection rule)}, && (2) \\ y_{3i} &= \beta_3' X_{3i} + \sigma_3 u_{3i} \qquad \text{(regression equation)}. && (3) \end{aligned}$$

For $k = 1, 2, 3$ and individual i, X_{ki}'s are vectors of explanatory variables, β_k's are the corresponding vectors of unknown coefficients, u_{ki}'s are the random disturbances, and σ_k's are unknown scale parameters. The variables y_{1i}^* and y_{2i}^* are unobserved continuous random variables, but without loss of generality, functions of them classify individuals to different categories according to a *sample selection regime* denoted by Λ. From now on, we drop the observation subscript i to avoid notational clutter. Let $X = X_1 \cup X_2 \cup X_3$.[1] The feature of interest is

$$E(y_3 \mid X, \Lambda) = \beta_3' X_3 + \sigma_3 E(u_3 \mid X, \Lambda). \tag{4}$$

Selectivity is manifested via $E(u_3 \mid X, \Lambda) \neq 0$, which renders conventional linear regression inconsistent. The goal is to consistently estimate β_3 conditional on the sample selection regime

[1] Our notation accommodates the use of different sets of explanatory variables in the two selection equations and the outcome equation. When alternative specific variables are present, it is natural for X_1 and X_2 to be different. Justification of exclusions that would distinguish X_3 from X_1 and X_2 would rest on context specific arguments.

Λ. Parametric approaches exploit additional assumptions which yield a functional form for $E(u_3 \mid X, \Lambda)$, so that adjustment for selectivity can be implemented. It is customary to assume that the random disturbance vector (u_1, u_2, u_3) is distributed independently across individuals and of X. Once the selectivity adjustment component is consistently estimated, linear regression becomes a viable choice.

In the cases covered in Tunali (1986), the sample selection regime can be expressed as $\Lambda = \{D_1, D_2\}$ where D_1 and D_2 are two dichotomous variables indicating the outcomes of two selection rules:

$$D_1 = \left\{ \begin{array}{c} 1 \text{ if } y_1^* > 0 \\ 0 \text{ if } y_1^* \leqslant 0 \end{array} \right\} \text{ and } D_2 = \left\{ \begin{array}{c} 1 \text{ if } y_2^* > 0 \\ 0 \text{ if } y_2^* \leqslant 0 \end{array} \right\}. \tag{5}$$

Note that the support of $(y_1^*,\ y_2^*)$ is broken down into four mutually exclusive regions which results in the four-way classification given in the 2×2 table below, where S_j is the set of individuals in the j^{th} subsample for $j = 1,\ 2,\ 3,\ 4$.

Figure 1: Possible Double Selection Outcomes

	$D_2 = 0$	$D_2 = 1$
$D_1 = 0$	S_1	S_2
$D_1 = 1$	S_3	S_4

The case in which we observe all the cells of 2×2 table yields a complete classification of the original sample. However, incomplete classification cases, in which only three or two distinct cells of 2×2 table are observed, are also possible. Various sample selection regimes which fit this set-up are discussed in Tunali (1986). Without loss of generality, he assumes that y_3 is observed for the subsample S_4, so that the conditional expectation function of interest given in Equation (4) specializes to:

$$\begin{aligned} E(y_3 \mid X, D_1 = 1, D_2 = 1) &= \beta_3' X_3 + \sigma_3 E(u_3 \mid X, D_1 = 1, D_2 = 1) \\ &= \beta_3 X_3 + \sigma_3 E(u_3 \mid \sigma_1 u_1 > -\beta_1' X_1, \sigma_2 u_2 > -\beta_2' X_2). \end{aligned} \tag{6}$$

The conditioning on X is implicit throughout, but we drop it for clarity. Tunali (1986) assumes that u_1, u_2 and u_3 have a trivariate normal distribution and sets $\sigma_1 = \sigma_2 = 1$. The variance normalization is an innocuous assumption for his set-up because of the nature of of Λ: truncation with respect to u_1 does not involve u_2 (and vice versa). However, for other choices of Λ, this is no longer the case. In obtaining our solution, we relax both restrictions. In particular, we let the form of the distribution of the random disturbance in the regression equation be free. Under our weaker distributional assumption, we denote the disturbance term of the regression equation by $\tilde{u}_3$ instead of u_3. We assume that $\tilde{u}_3$ has zero expectation and unit variance conditional on X. However, we maintain that u_1 and u_2 have a standard bivariate normal distribution.

In Section 2.3, we set $\sigma_1 = \sigma_2 = 1$ for convenience and give the functional form of $E(\tilde{u}_3 \mid u_1 > -\beta_1' X_1, u_2 > -\beta_2 X_2)$ obtained via an Edgeworth expansion. That is, we use the exclusive selection rule set-up of Tunali (1986) and provide a strategy for consistent estimation of β_3 on subsample S_4. The methodology can easily be adopted to the other cases covered in Tunali (1986). Presently, we situate our approach within a broader context and justify our choice. We relax the variance restriction later, when we confront the empirical problem in Tunali and Baslevent (2006).

2.2 Solution Strategy

Given our distributional assumptions, numerous approaches can be pursued. Kolassa (1997) offers a broad list of series approximations and discusses the computational implications. One commonly used approximation is the Normal Inverse Gaussian, examined in detail in Eriksson et al. (2004). This method involves a partial specification: the conditional expectation is written without specifying the joint distribution fully. Even though this method works well for the univariate case, its behavior is not known in the multivariate case. Edgeworth expansion also involves a partial specification, but is easier to apply.

For our purposes, the most appealing characteristic of Edgeworth expansion is the fact that it nests the conventional trivariate normality assumption. Under trivariate normality, the conditional expectation function $E(u_3 \mid u_1, u_2)$ becomes a linear function of u_1 and u_2. Consequently trivariate normality allows us to express $E(u_3 \mid \Lambda)$ using only the first and second central moments. Edgeworth expansion provides an improvement because higher order moments are introduced in obtaining $E(\tilde{u}_3 \mid \Lambda)$. In the derivation of our Edgeworth expansion, we also use the third and fourth central moments. These allow us to relax (and test) the conventional restrictions imposed on the skewness and kurtosis of the distribution.

2.3 Trivariate Edgeworth Expansion

Let $f(u_1, u_2, \tilde{u}_3)$ be the joint density of u_1, u_2 and $\tilde{u}_3$. It is helpful to keep in mind that u_1, u_2 and $\tilde{u}_3$ are the random disturbances of Equations (1) - (3). The triplet $(u_{1i}, u_{2i}, \tilde{u}_{3i})$ is assumed to be independently and identically distributed across individuals with zero mean vector and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix},$$

and is independent of X. For brevity, we suppress the conditioning on covariate matrix X and the individual subscript i throughout this subsection.

We denote the standard trivariate normal density either by $STVN(\rho_{12}, \rho_{13}, \rho_{23})$ or by $g(u_1, u_2, u_3)$ and the standard bivariate normal density either by $SBVN(\rho_{12})$ or by $g(u_1, u_2)$. Under some general conditions given by Chambers (1967) and conditional on existence of all the moments of u_1, u_2, and $\tilde{u}_3$, the trivariate density $f(u_1, u_2, \tilde{u}_3)$ can be expanded in terms of a series of derivatives of $g(u_1, u_2, u_3)$:

$$f(u_1, u_2, \tilde{u}_3) = g(u_1, u_2, u_3) + \sum_{r+s+p \geq 3} \frac{(-1)^{r+s+p}}{r!s!p!} A_{rsp} D_{u_1}^r D_{u_2}^s D_{u_3}^p g(u_1, u_2, u_3), \tag{7}$$

where A_{rsp}'s are functions of the moments of $f(u_1, u_2, \tilde{u}_3)$, and

$$D_{u_1}^r D_{u_2}^s D_{u_3}^p g(u_1, u_2, u_3) = \frac{\partial^{r+s+p} g(u_1, u_2, u_3)}{\partial u_1^r \partial u_2^s \partial u_3^p}. \tag{8}$$

The iterative relation between the derivatives in Equation (8) can be captured with the help of Hermite polynomials. In our case, a bivariate version is sufficient. The bivariate Hermite polynomial of $(r, s)^{th}$ order, denoted by $H_{rs}(u_1, u_2)$, satisfies

$$D_{u_1}^r D_{u_2}^s g(u_1, u_2) = (-1)^{r+s} H_{rs}(u_1, u_2) g(u_1, u_2). \tag{9}$$

In other words, $H_{rs}(u_1, u_2)$ is a function of the ratio of $(r, s)^{th}$ to $(0, 0)^{th}$ order derivatives of $g(u_1, u_2)$. With the help of Equation (9), we can deduce from Equation (7) that the joint density of u_1 and u_2 can be written as:

$$h(u_1, u_2) = \left[1 + \sum_{r+s\geq 3} \frac{1}{r!s!} A_{rs0} H_{rs}(u_1, u_2)\right] g(u_1, u_2). \tag{10}$$

Employing the usual practice in the related literature (see Lee (1982); Lee (1984); Lahiri and Song (1999)), we consider terms up to 4^{th} order ($r + s + p = 4$ in Equation (7)). This amounts to truncating the series approximation. The truncated series is known as a Gram Charlier Expansion for the univariate case. For the bivariate case, it is called as a Type AaAa surface in Pretorius (1930) and a Type AA surface in Mardia (1970). For the trivariate case, we may use shorter designation provided in Mardia (1970) as our guide and term the truncated expansion as a Type AAA surface.

The truncation eases the computational burden in two ways. First, it reduces the number of terms we have to consider. Second, we know that for the Type AAA surface ($r + s + p \leq 4$), $A_{rsp} = K_{rsp}$ where K_{rsp}'s denote trivariate cumulants (Lahiri and Song (1999)). Lee (1984) exploits the truncated version of Equation (10) for deriving a test for bivariate normality while Lahiri and Song (1999) use Equation (7) to obtain a test for trivariate normality. Our approach is different in the sense that we focus on correction for selectivity.

Returning to the generic representation of the double selection problem, recall that we assume $(u_1, u_2) \sim SBVN(\rho_{12})$ in Equations (1) and (2), but allow the distribution of $\tilde{u}_3$ to be free with $E(\tilde{u}_3) = 0$ and $V(\tilde{u}_3) = 1$ in Equation (3) and set $\sigma_1 = \sigma_2 = 1$. Extending the steps given in Mardia (1970), we get the following operationally useful formula for the Type AAA surface:

$$\begin{aligned} E(\tilde{u}_3 \mid u_1, u_2) &= K_{101}H_{10}(u_1, u_2) + K_{011}H_{01}(u_1, u_2) + \frac{1}{2}K_{201}H_{20}(u_1, u_2) \\ &+ \frac{1}{2}K_{021}H_{02}(u_1, u_2) + K_{111}H_{11}(u_1, u_2) + \frac{1}{6}K_{301}H_{30}(u_1, u_2) \\ &+ \frac{1}{6}K_{031}H_{03}(u_1, u_2) + \frac{1}{2}K_{211}H_{21}(u_1, u_2) + \frac{1}{2}K_{121}H_{12}(u_1, u_2) \end{aligned} \tag{11}$$

Equivalently,

$$
\begin{aligned}
E(\tilde{u}_3 \mid u_1, u_2) &= \frac{1}{1-\rho_{12}^2}[(\rho_{13}-\rho_{12}\rho_{23})u_1 + (\rho_{23}-\rho_{12}\rho_{13})u_2] + \frac{1}{2}K_{201}H_{20}(u_1,u_2) \\
&+ \frac{1}{2}K_{021}H_{02}(u_1,u_2) + K_{111}H_{11}(u_1,u_2) + \frac{1}{6}K_{301}H_{30}(u_1,u_2) \\
&+ \frac{1}{6}K_{031}H_{03}(u_1,u_2) + \frac{1}{2}K_{211}H_{21}(u_1,u_2) + \frac{1}{2}K_{121}H_{12}(u_1,u_2) \qquad (12)
\end{aligned}
$$

The derivation is provided in Appendix 6.1. Appendix 6.2 gives explicit expressions for the bivariate Hermite polynomials in Equation (11), and Appendix 6.3 presents derivations of moments of the truncated *SBVN* distribution up to the third order.[2]

Incorporating the explicit formulas for the Hermite polynomials into Equation (12) and using the moment formulas of the truncated *SBVN* distribution, we obtain:

$$
\begin{aligned}
E(\tilde{u}_3 \mid u_1 > -\beta_1' X_1, u_2 > -\beta_2 X_2) &= \rho_{13}\lambda_1 + \rho_{23}\lambda_2 + \frac{1}{2}K_{201}\lambda_3 + K_{111}\lambda_4 + \frac{1}{2}K_{021}\lambda_5 \\
&+ \frac{1}{6}K_{301}\lambda_6 + \frac{1}{6}K_{031}\lambda_7 + \frac{1}{2}K_{211}\lambda_8 + \frac{1}{2}K_{121}\lambda_9. \qquad (13)
\end{aligned}
$$

Explicit formulas of λ's are provided in Appendix 6.4. Note that λ_1 and λ_2 are the terms given in Tunali (1986) and omission of the remaining terms in Equation (13) corresponds to the standard trivariate normality specification.

In theory, we can exploit Equation (10) and use an approximation for the bivariate component, $h(u_1, u_2)$ to the same order ($r + s = 4$), instead of assuming standard bivariate normality. However, this will add 9 terms to the denominator of our conditional expectation. Since each of these terms includes a different cumulant, the problem quickly becomes computationally intractable. We therefore rely on the *SBVN* distribution. Two strands in the econometrics literature provide theoretical justification for our apparently cavalier approach to the specification of the distribution of the selection equation disturbances. First, as Ruud (1983) and Stoker (1986) have shown, slope coefficients of index-function models can be consistently estimated up to a factor of proportionality using any commonly used technique like probit or logit. Second is the work on quasi maximum

[2]Our moment derivations of the truncated SBVN distribution up to the second order corroborate with the formulas in Henning and Henningsen (2007), but differ from Rosenbaum (1961). In our view, Rosenbaum (1961) made a sign error while calculating integrals. Besides, Henning and Henningsen (2007) cross-check their formulas using numerical integration and Monte Carlo simulation.

likelihood methodology: If the correct specification of the distribution is in the linear exponential family and we incorrectly specify the model by using another distribution from that family, we still get consistent estimates.

With Equation (13) in hand, the regression equation may be expressed as

$$y_3 = \beta_3 X_3 + \gamma_1\lambda_1 + \gamma_2\lambda_2 + \gamma_3\lambda_3 + \gamma_4\lambda_4 + \gamma_5\lambda_5 + \gamma_6\lambda_6 + \gamma_7\lambda_7 + \gamma_8\lambda_8 + \gamma_9\lambda_9 + \nu_3 \qquad (14)$$

where the new random disturbance ν_3 has the desired property $E(\nu_3 \mid u_1 > -\beta_1' X_1, u_2 > -\beta_2 X_2) = 0$ in view of the fact that $\nu_3 = \tilde{u}_3 - E(\tilde{u}_3 \mid \Lambda)$. As seen from the formulas of λ's in Appendix 6.4, the new regression equation is highly non-linear. Therefore, adopting the practice in the literature, we rely on two-step estimation procedure provided in Heckman (1979) to estimate it. In the first step, we target the parameters of the selection equations (1) - (2) and maximize the relevant likelihood function (which depends on Λ, the sample selection regime).[3] This allows us to form consistent estimates of the λ's. We use the estimated $\hat{\lambda}$'s as additional regressors and fit Equation (14) using linear regression on the subsample of individuals in S_4 to estimate β_3 and γ's. Since generated regressors are involved, we take heteroscedasticity into account and report Hubert-White standard errors. NOTE: We need to generate the consistent variance-covariance matrix here!!

We may refer to the $\hat{\lambda}$'s as selectivity correction terms. Then, we can test for the presence of selectivity bias by testing joint significance of all selectivity correction terms, $\hat{\lambda}_1$, $\hat{\lambda}_2$, ..., $\hat{\lambda}_9$. Furthermore, by setting $\gamma_3 = \gamma_4 = ... = \gamma_9 = 0$ we get a test of the conventional trivariate normality specification. Both tests are conducted under the maintained assumption that the random disturbances in the two selection equations are bivariate normally distributed. The latter test can also be thought as a test for linearity of the conditional expectation of $\tilde{u}_3$ given u_1 and u_2.

3 Application

3.1 Problem

We illustrate our methodology in the context of the empirical work undertaken in Tunali and Baslevent (2006). The data come from October 1988 Household Labor Force Survey in Turkey. The

[3] The formation of the likelihood function according to different sample selection regimes is explained in detail by Tunali (1986).

working sample consists of $8,962$ married women between $20 - 54$ years old who reside in urban areas (population $\geqslant 20,000$), are not in school and identified as the only wife of the household head. There are four labor force participation categories in the data set: $7,770$ non-participants, 745 wage workers (or regular/casual wage and salary earners), 181 self-employed which include employers, own-account work and unpaid family workers, and 266 unemployed. The ultimate goal of the paper is to estimate labor supply elasticities for wage workers. This calls for estimation of a wage equation on a subsample of wage workers which constitute less than 10 percent of the full sample. Hence there is good reason to take selection into consideration.

Tunali and Baslevent (2006) assume that home-work (or non-participation), self-employment and wage work utilities can be expressed as follows.

$$\text{Home-work utility} \quad : \quad U_0^* = \theta_0' z + \upsilon_0, \tag{15}$$

$$\text{Self-employment utility} \quad : \quad U_1^* = \theta_1' z + \upsilon_1, \tag{16}$$

$$\text{Wage work utility} \quad : \quad U_2^* = \theta_2' z + \upsilon_2, \tag{17}$$

where z is a vector of observed variables, θ_j's are the corresponding vectors of unknown coefficients and υ_j's are the random disturbances. Assuming that individuals choose the state with highest utility, their decisions can be captured using the utility differences:

$$y_1^* = U_1^* - U_0^* = (\theta_1' - \theta_0')z + (\upsilon_1 - \upsilon_0) = \beta_1' z + \sigma_1 u_1, \tag{18}$$

$$y_2^* = U_2^* - U_1^* = (\theta_2' - \theta_1')z + (\upsilon_2 - \upsilon_1) = \beta_2' z + \sigma_2 u_2. \tag{19}$$

This pair of selection equations has the same form as the ones in Equations (1) - (2) with $X_1 = X_2 = z$. Note that y_1^* can be expressed as the propensity to be self-employed rather than being a non-participant and y_2^* as the incremental propensity to engage in wage work rather than self-employment. Then, $y_1^* + y_2^*$ gives the propensity to engage in wage work over home-work. The preferences of unemployed women over employment options is not known, so Tunali and Baslevent (2006) follow Magnac (1991) and define the unemployed as people obtaining higher utility either from self-employment or wage work relative to home-work. Under this assumption, the four way classification observed in the sample arises as follows:

$$LFP = \left\{ \begin{array}{c} 0 = \text{home-work, if } y_1^* < 0 \text{ and } y_1^* + y_2^* < 0, \\ 1 = \text{self-employment, if } y_1^* > 0 \text{ and } y_2^* < 0, \\ 2 = \text{wage labor, if } y_2^* > 0 \text{ and } y_1^* + y_2^* > 0, \\ 3 = \text{unemployed, if } y_1^* > 0 \text{ or } y_1^* + y_2^* > 0. \end{array} \right\}. \tag{20}$$

The sample selection regime is given by $\Lambda = \{LFP\}$. Note that in this case the support of $(y_1^*,\ y_2^*)$ is broken down into three mutually exclusive regions, which respectively correspond to $LFP = 0$, 1, and 2. The region for $LFP = 3$ is the union of those for $LFP = 1$ and 2. We see that the classification in our sample is obtained via a pair from the triplet $\{y_1^*,\ y_2^*,\ y_1^* + y_2^*\}$. Suppose we were to normalize the variances of y_1^* and $y_1^* + y_2^*$ to 1, but this has an implication for the variance of y_2^* $(\sigma_2^2 = -2\rho_{12})$. Thus the usual variance normalization is no longer innocuous. We may apply the normalization to one of $\sigma_{11} = \sigma_1^2$ and $\sigma_{22} = \sigma_2^2$, but must leave the other variance free to take on any positive value. In our analysis, we take $\sigma_{11} = 1$ and let σ_{22} be free. In the first step, we rely on maximum likelihood estimation and obtain consistent estimates of β_1, β_2, ρ_{12} and σ_2 subject to $\sigma_1 = 1$. The likelihood function is given by

$$L = \prod_{LFP=0} P_0 \prod_{LFP=1} P_1 \prod_{LFP=2} P_2 \prod_{LFP=3} P_3, \tag{21}$$

where $P_j = Pr(LFP = j)$ for $j = 0,\ 1,\ 2,\ 3$. There are two restrictions on the $P_j's$: (i) $P_0 + P_1 + P_2 = 1$, and (ii) $P_3 = 1 - P_0$. The explicit definitions of $P_j's$ are provided in Appendix 6.5 where the methodology given in Section 2.3 is modified to handle the current case.

The regression equation for this problem is a Mincer-type wage equation obtained by setting $y_3 = \log(wage)$ in Equation (3) where X_3 includes human capital variables and labor market characteristics:

$$\log(wage) = \beta_3^{'} X_3 + \sigma_3 u_3. \tag{22}$$

The aim is to estimate β_3 for wage workers, on the subsample with $LFP = 2$. After forming the estimates of selectivity correction terms via first step estimates, we run a linear regression equation with 9 selectivity correction terms in the second step, as shown in Equation (14).

Table 1: Sample Means (Standard Deviations) by Labor Force Participation Status

Variable	Full Sample	Non-participant	Self-employed	Wage Worker	Un-employed
Own wage	-	-	-	1.39 (3.17)	-
Age	33.29	33.41	34.28	32.82	30.65
Age Squared/100	11.75	11.86	12.37	11.18	9.84
Experience	20.29	20.82	21.31	15.76	16.93
Experience Squared/100	4.87	5.08	5.26	3.11	3.39
Illiterate (Reference)	0.25	0.27	0.21	0.055	0.13
Literate without a Diploma	0.090	0.095	0.12	0.034	0.071
Elementary School	0.49	0.52	0.47	0.22	0.48
Middle School	0.058	0.055	0.10	0.060	0.094
High School	0.087	0.059	0.072	0.35	0.20
University	0.030	0.0069	0.022	0.28	0.023
Husband Self-employed	0.33	0.35	0.39	0.16	0.18
Children Aged $0-2$	0.26	0.26	0.18	0.22	0.22
Children Aged $3-5$	0.34	0.35	0.28	0.27	0.34
Female Children aged $6-14$	0.41	0.42	0.49	0.32	0.36
Male Children aged $6-14$	0.44	0.45	0.49	0.34	0.43
Extended Household	0.12	0.13	0.14	0.11	0.079
Ext. HH $\times$ Children Aged $0-2$	0.029	0.029	0.028	0.031	0.019
Ext. HH $\times$ Children Aged $3-5$	0.036	0.036	0.033	0.039	0.019
Ext. HH $\times$ Female Ch. Aged $6-14$	0.045	0.046	0.044	0.032	0.030
Ext. HH $\times$ Male Ch. Aged $6-14$	0.047	0.049	0.050	0.039	0.030
Share of Textiles	0.31	0.31	0.39	0.30	0.31
Share of Agriculture	0.17	0.18	0.20	0.15	0.19
Share of Finance	0.060	0.059	0.053	0.064	0.056
Migration Rate	0.024	0.023	0.021	0.035	0.018
Marmara (Reference)	0.35	0.35	0.32	0.37	0.28
Aegean	0.11	0.11	0.10	0.14	0.21
South	0.12	0.12	0.26	0.099	0.13
Central	0.19	0.18	0.10	0.23	0.17
North West	0.040	0.037	0.028	0.055	0.075
East	0.078	0.081	0.088	0.044	0.086
South East	0.081	0.087	0.088	0.039	0.015
North East	0.033	0.034	0.0055	0.027	0.041
Population $200,000$ or more	0.67	0.66	0.69	0.74	0.60
Population 1 million or more	0.51	0.50	0.43	0.61	0.46
Share of Welfare Party	0.073	0.074	0.059	0.062	0.058
Share of Left of Center	0.35	0.35	0.33	0.36	0.36
Sample size	8,962	7,770	181	745	266

3.2 Results

The list of variables and descriptive statistics broken down by subsample are given in Table 1. In the first step of estimation, we imposed a single variance normalization ($\sigma_{11} = 1$) and restricted the correlations and variances to their theoretical ranges using non-linear transformations. Since sampling frame relies on stratification, the probability weights coming from the data were used. Robust standard errors are reported throughout the section.

Table 2 provides the results of the first step. The estimate of the correlation coefficient between the random disturbances of the two selection equations, $\hat{\rho}_{12}$, is very close to -1, which suggests that computational fragility is an issue. Evidently this is a weakness that our double selection model shares with the broader class of multinomial probit models as discussed in Keane (1992). Since our interests lie elsewhere, we do not dwell on the problem here.[4] Detailed discussion of the results and a sensitivity analysis which explores other variance restrictions can be found in Tunali and Baslevent (2006). The negative sign of $\hat{\rho}_{12}$ is justified by the argument that the unobserved characteristics that make a woman more likely to choose self-employment rather than home-work, also make her less likely to choose wage work over self-employment in Tunali and Baslevent (2006). We touch on some of their substantive findings in Section 4.3 in order to compare and contrast them with the first stage results of Multinomial Logit based selection correction models.

We proceed with our results from the second step which are collected in Table 3. Following Tunali and Baslevent (2006), (i) we excluded 15 variables that appeared in the selection equations, (ii) included a second degree polynomial in experience instead of the polynomial in age, and (iii) dropped 10 observations due to missing wage information. We report the least squares estimates of the wage equation in three columns, respectively without selectivity correction, with conventional correction (corresponds to trivariate normality specification), and with Edgeworth expansion based correction. A Wald test for exclusion of all the selectivity correction terms yields strong evidence in favor of non-random selection ($p - value = 0.0006$). Moreover, a Wald test of the restriction $\gamma_3 = \gamma_4 = ... = \gamma_9 = 0$ favors our Edgeworth expansion approach rather than conventional correction ($p - value = 0.0424$).

[4] Keane (1992) investigates the issue of computational fragility in multinomial probit models and proposes using alternative specific variables as well as exclusions in each selection equation. Since alternative specific variables are not available, this venue is not a viable option in the present case.

Table 2: Maximum Likelihood Bivariate Probit Estimates of Participation Equations (Normalized Version)

Variable	First Selection		Second Selection	
	Coefficient	Std. Error	Coefficient	Std. Error
Age	0.077	0.066	−0.010	0.029
Age Squared/100	−0.124	0.096	0.024	0.041
Literate without a Diploma	0.256***	0.099	−0.185	0.122
Elementary School	0.124	0.085	−0.048	0.075
Middle School	0.471***	0.138	−0.194	0.187
High School	0.534	0.580	0.090	0.116
University	0.931	1.050	0.149	0.154
Husband Self-employed	−0.081	0.242	−0.152*	0.092
Children Aged $0-2$	−0.178**	0.075	0.077	0.089
Children Aged $3-5$	−0.078	0.077	0.012	0.053
Female Children aged $6-14$	0.055	0.103	−0.075	0.094
Male Children aged $6-14$	0.031	0.089	−0.072	0.065
Extended Household	0.189	0.133	−0.167	0.130
Ext. HH × Children Aged $0-2$	0.096	0.181	−0.047	0.182
Ext. HH × Children Aged $3-5$	−0.058	0.178	0.091	0.149
Ext. HH × Female Ch. Aged $6-14$	−0.252	0.165	0.221	0.161
Ext. HH × Male Ch. Aged $6-14$	−0.083	0.161	0.103	0.137
Share of Textiles	0.058	0.242	0.172	0.153
Share of Agriculture	−0.759*	0.432	0.952***	0.356
Share of Finance	−3.686*	2.121	2.918	2.637
Migration Rate	−3.642**	1.613	4.214***	1.515
Aegean	−0.229	0.228	0.326*	0.180
South	−0.038	0.101	−0.015	0.093
Central	−0.508*	0.263	0.547*	0.292
North West	−0.392	0.306	0.526**	0.262
East	−0.381	0.233	0.424*	0.233
South East	−0.291*	0.151	0.149	0.201
North East	−0.850	0.546	0.970*	0.528
Population 200,000 or more	−5.969***	2.063	5.628**	2.857
Population 1 million or more	−0.697	0.695	1.172**	0.508
Share of Welfare Party	0.114	0.088	−0.029	0.123
Share of Left of Center	−0.160*	0.092	0.139	0.099
Constant	−1.688	1.340	−0.242	0.643
σ_{11}	1 [normalized]			
σ_{22}	0.358 (0.514)			
ρ_{12}	−0.971*** (0.046)			
Number of Observations	8,962			
Log-Likelihood Without Covariates	−3,247.25			
Log-Likelihood With Covariates	−2,443.05			

Notes: Robust standard errors are reported. * is significant at 10%; ** is significant at 5%; *** is significant at 1%.

Table 3: Least Squares Estimates of the Wage Equation

Variable	With Robust Correction		With Conventional Correction		Without Selectivity Correction	
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Experience	0.032***	0.012	0.034***	0.013	0.031**	0.013
Experience Squared/100	−0.064**	0.033	−0.072**	0.037	−0.064*	0.037
Literate without a Diploma	−0.119	0.174	−0.080	0.181	−0.112	0.186
Elementary School	−0.055	0.116	−0.028	0.120	−0.011	0.121
Middle School	0.215	0.149	0.256*	0.144	0.231*	0.125
High School	0.594***	0.191	0.437**	0.184	0.465***	0.119
University	1.254***	0.272	1.083***	0.250	1.038***	0.128
Share of Textiles	−0.308**	0.150	−0.282**	0.143	−0.407***	0.122
Share of Agriculture	0.098	0.237	0.085	0.238	0.150	0.238
Share of Finance	3.437***	1.319	2.557**	1.225	4.241***	1.158
Aegean	−0.135*	0.072	−0.169**	0.073	−0.131*	0.069
South	0.162**	0.081	0.172**	0.080	0.159*	0.082
Central	0.035	0.072	0.007	0.073	0.029	0.073
North West	−0.151	0.100	−0.170*	0.099	−0.166*	0.096
East	0.109	0.114	0.096	0.115	0.142	0.113
South East	0.130	0.118	0.086	0.114	0.191*	0.100
North East	−0.137	0.117	−0.210*	0.113	−0.129	0.115
$\hat{\lambda}_1$	−0.044	0.687	0.171	0.146	—	
$\hat{\lambda}_2$	2.441	3.132	−0.458**	0.182	—	
$\hat{\lambda}_3$	0.135	0.412	—		—	
$\hat{\lambda}_4$	2.033	2.784	—		—	
$\hat{\lambda}_5$	2.076	2.897	—		—	
$\hat{\lambda}_6$	−0.060	0.215	—		—	
$\hat{\lambda}_7$	0.625	1.006	—		—	
$\hat{\lambda}_8$	0.595	1.291	—		—	
$\hat{\lambda}_9$	−1.149	2.080	—		—	
Constant	−1.087**	0.519	−0.927***	0.345		
Number of Observations	735		735		735	
R^2	0.387		0.378		0.367	

Notes: Robust standard errors are reported. * is significant at 10%; ** is significant at 5%; *** is significant at 1%.

In light of the first stage results, participation probability increases with education; furthermore, wage work and self-employment orientations of participants differ systematically. One would think that all this sorting has implications for the estimates of the wage equation. Indeed, comparison of the substantive results from the three models underscores the importance of proper adjustment for selection. According to our specification, while the coefficient of middle school becomes insignificant (implying people having middle school education or less earn the same wage), high school and university coefficients increase. Evidently, the minority which opts for wage work is highly selective in terms of unobservables and failure to account for this yields underestimation of the returns to higher levels of education. With illiterates as the reference category, high school results in 81 percent higher returns on average according to our Edgeworth expansion based correction whereas it is 55 percent on average under conventional correction. In the case of university graduates the contrast is equally dramatic: 250 percent under our method, and 195 percent in the trivariate normality specification. According to our trusted estimates, average incremental returns are 21.9 percent per additional year of high school and 28 percent per additional year of university education.

The other statistically significant variables in the Edgeworth expansion based correction have coefficient estimates with magnitudes between those under random selection and under conventional correction for non-random selection. Based on the our robust correction, a one percent increase in the share of textile jobs is predicted decrease wages by 0.31 percent while a one percent increase in share of finance jobs increases wages by 3.4 percent. These effects are underestimated in the trivariate normality specification.

3.3 Multicollinearity

Another issue that arises in the context of our robust correction approach is the large standard errors of individual selectivity correction terms (see Table 3). Letting R_j^2 for $j = 1, 2, ..9$ denote the coefficient of determination in the regression of $\hat{\lambda}_j$ on the remaining selectivity correction terms, we obtained values in the range $0.995 < R_j^2 < 1$. This implies extremely high multicollinearity. In the context of the original Heckman-Lee formulation, Little (1985) and Leung and Yu (1996) have pointed out that the correction term (so-called inverse Mills-ratio) is linear over much of the range of variation of the underlying index function. Evidently the same observation applies to our formulation which exploits nine correction terms. However, one might think that the problem is exacerbated by the computational fragility of the double selection model. We pursued this line of thinking by engaging in a secondary empirical investigation. First, to side step the computational fragility issue, we re-estimated our double selection model by imposing the second variance normalization: $\sigma_{22} = 1$. This version yielded $\hat{\rho}_{12} = -0.746$, a correlation estimate which is comfortably away from the boundary. However multicollinearity remained extremely high, with $0.964 < R_j^2 < 0.9997$ for $j = 1, 2, ..9$.

Next, to assure that multicollinearity among the correction terms was not specific to a particular data set, we also generated $\hat{\lambda}$'s for the migration-remigration problem of Tunali (1986), which is another example of double selection at work. The empirical context of the problem is as follows. There are three earnings profiles: y_s^* under the stay option, y_o^* under the one-time move option, y_f^* under the frequent move option. Let $\delta_1^* = y_o^* - y_s^*$ denote anticipated earnings gain from the one-time move to the best potential destination, and τ_1^* be the cost of that move. Let $\delta_2^* = y_f^* - y_o^*$ denote anticipated incremental earnings gain from the frequent move involving the best potential combination of intermediate and final destinations, and τ_2^* be the cost of that move. Tunali (1986) assumes that individuals are rational, in that sense anticipations are unbiased. Then, the latent dependent variables of Equations (1) - (2) in the generic formulation in Section 2.1 are obtained as $y_1^* = \delta_1^* - \tau_1^*$ and $y_2^* = \delta_2^* - \tau_2^*$. Upon introducing the variable $r = sup\{0,\ y_1^*,\ y_1^* + y_2^*\}$, the decision rule becomes

$$\text{Stay if } r = 0, \tag{23}$$

$$\text{Move once if } r = y_1^*, \tag{24}$$

$$\text{Move more than once if } r = y_1^* + y_2^*. \tag{25}$$

However, modified decision rule in Tunali (1986) treats $r = y_1^* + y_2^* > 0 > y_1^*$ (move more than once will be optimal in this case) as a stay decision on the grounds that the first move will not be feasible when $0 > y_1^*$. Under this feasibility constraint, the modified decision rule may be expressed using the two dichotomous variables given in Equation (5):

$$\text{Stay if } D_1 = 0, \tag{26}$$

$$\text{Move once if } D_1 = 1 \text{ and } D_2 = 0, \tag{27}$$

$$\text{Move more than once if } D_1 = 1 \text{ and } D_2 = 1. \tag{28}$$

Thus, D_2 is observed if and only if $D_1 = 1$. The sample selection regime for this problem is a 3-way classification with $\Lambda = \{D_1, D_2 \text{ if and only if } D_1 = 1\}$, whereby the cells of the first row of

Figure 1 cannot be distinguished. A detailed empirical investigation of this incomplete classification problem may be found in Tunali (1986). Returning to our examination of the multicollinearity issue, the first step yields $\hat{\rho}_{12} = 0.389$ but high multicollinearity is still present in the second step: $0.910 < R_j^2 < 0.9999$ for $j = 1, 2, ..9$. It appears that extremely high multicollinearity is part and parcel of our approach.[5]

4 Multinomial Logit Based Selection Correction Models

Multinomial Logit based selection correction models are used as alternatives to Multivariate Normal based models in the literature. In what follows we offer a brief exposition to adapt the relevant models to our empirical context and underscore the differences among them. Estimation proceeds in two steps. The Multinomial Logit based selection correction models differ in the second step. Conveniently, the empirical results for all the models reviewed here can be obtained using the STATA module *selmlog* developed by Bourguignon et al. (2007).

4.1 Selection Step

The starting point is a selection equation for each possible state. In the context of Tunali and Baslevent (2006), upon suppression of the subscripts for the individuals these may be expressed as

$$y_j^* = \theta_j' z + \eta_j \text{ for } j = 0,\ 1,\ 2,\ 3, \tag{30}$$

where y_j^*'s denote the unobserved utility obtained from the choice of labor force participation status j, z is the vector of explanatory variables, θ_j's are the corresponding vectors of unknown coefficients and η_j's are the random disturbances. Let $r = \max\ (y_0^*, y_1^*, y_2^*, y_3^*)$. The four way classification is

[5]To understand the implications for standard errors, we may partition the full covariate matrix $\tilde{X}_3$ as $(X_{31} \mid X_{32})$, where X_{31} includes all explanatory variables of the model without selectivity correction terms ($X_{31} = X_3$ in terms of the notation given in the generic statement of the double selection model), and X_{32} includes all the selectivity correction terms. We can write:

$$(\tilde{X}_3'\tilde{X}_3)^{-1} = \begin{pmatrix} X_{31}'X_{31} & X_{31}'X_{32} \\ X_{32}'X_{31} & X_{32}'X_{32} \end{pmatrix}^{-1}. \tag{29}$$

We use the lower right block of the matrix $(\tilde{X}_3'\tilde{X}_3)^{-1}$ in calculating the standard errors of the coefficients on the $\hat{\lambda}_j$'s. This block may be computed as $[(X_{32}'X_{32}) - X_{32}'X_{31}(X_{31}'X_{31})^{-1}X_{31}'X_{32}]^{-1}$ via submatrix inverse theorem (see Goldberger (1991)). Extremely large R_j^2's imply that the matrix $(X_{32}'X_{32})$ is nearly singular. However, this matrix is not inverted. Instead, the difference between $(X_{32}'X_{32})$ and $X_{32}'X_{31}(X_{31}'X_{31})^{-1}X_{31}'X_{32}$ is inverted. Therefore, it is possible to compute the standard errors despite extremely high multicollinearity between the $\hat{\lambda}_j$'s.

obtained via the following categorical variable:

$$LFP' = \left\{ \begin{array}{l} 0 = \text{home-work, if } r = y_0^*, \\ 1 = \text{self-employment, if } r = y_1^*, \\ 2 = \text{wage labor, if } r = y_2^*, \\ 3 = \text{unemployed, if } r = y_3^*. \end{array} \right\}. \tag{31}$$

Hence, the sample selection regime is given by $\Lambda = \{LFP'\}$ and Equation (30) corresponds to the selection equations.

Under the assumption that the random disturbances (η_j's) are independently and identically Gumbel distributed, independently of the vector of explanatory variables, McFadden (1974) proves that the selection probabilities are given by the Multinomial Logit model (MLM):

$$\pi_j = \Pr(LFP' = j \mid z) = \frac{\exp(\theta_j' z)}{\sum_{\zeta=0}^{3} \exp(\theta_\zeta' z)}, \ j = 0, \ 1, \ 2, \ 3. \tag{32}$$

Since $\sum_{k=0}^{3} \pi_k = 1$, we choose non-participants as the reference group and set $\theta_0 = 0$. We may obtain consistent estimates of θ_j's by maximizing the likelihood function in Equation (33):

$$L = \prod_{LFP=0} \pi_0 \prod_{LFP=1} \pi_1 \prod_{LFP=2} \pi_2 \prod_{LFP=3} \pi_3. \tag{33}$$

Note that our selectivity correction approach allows correlation between the random disturbances of the two selection equations, but computational difficulties limit its generalization (unless simulation based estimation is used at the first stage). On the other hand, Multinomial Logit based selection correction models do not allow correlation between the random disturbances of the selection equations. Consequently they can handle a large number of selection equations. However, the independence assumption ushers in the so-called Independence of Irrelevant Alternatives (IIA) property, whereby the odds of choice between any two options is strictly a function of the systematic utility components of these two options. In our empirical context this implies that choice of self-employment over the home-work option is not affected by how attractive the wage work option is. In similar vein, the desirability of the self-employment option does not influence the choice of wage work over home-work. Thus strong (if not untenable) implicit behavioral assumptions may be present when the MLM set-up is adopted.

Note that in our model the unemployed prefer either self-employment or wage work over home-work, but we do not know their preferences between the two employment options. This restricts the choice probabilities in a particular way; namely, $\pi_3 = 1 - \pi_0$. Because of the IIA assumption, this restriction cannot be incorporated into the MLM. Instead, unemployment needs to be treated as a distinct state. Therefore, another potential limitation of these models is the failure to capture the choice structure appropriately. Despite this failure, we still discuss some well-known Multinomial Logit based selection correction models.

4.2 Regression Step

In our evaluation we considered the methods developed by Lee (1983), Dubin and McFadden (1984), Dahl (2002) and two variants of the Dubin-McFadden model proposed by Bourguignon et al. (2007). Regression equation given in Equation (22) remains intact. The aim as before is to estimate the parameter vector of the wage equation for the subsample of individuals with $LFP' = 2$. Returning to the generic statement of the selectivity problem in Equation (4), once again we are confronted with $E(u_3 \mid X, \Lambda) \neq 0$. Multinomial Logit based correction methods invoke different assumptions and end up with different parametric forms for $E(u_3 \mid X, \Lambda)$. As in the conventional Heckman-Lee formulation, the correction terms arise because of correlation between the random components of the selection equations and the random disturbance term of the regression equation. In what follows we denote the correlations between η_j and u_3 by ρ_j for $j = 0,\ 1,\ 2,\ 3$ and their differences by $\Upsilon_l = \rho_l - \rho_2$ for $l = 0,\ 1,\ 3$. Utilizing Equation (30), we may define the maximum utility difference relative to the wage labor state by

$$\epsilon_2 = \max(y_l^* - y_2^*) \text{ for } l \in \{0, 1, 3\}. \tag{34}$$

Then,

$$LFP' = 2 \Longleftrightarrow \epsilon_2 < 0. \tag{35}$$

4.2.1 Lee's (1983) Model (LEE)

Lee (1983) generalizes the two-step selection correction methodology provided by Heckman (1976, 1979) and Lee (1976, 1979) by extending it to polychotomous choice. He has two insights: (i) by treating the utility associated with the chosen state as the maximum of an order statistic, he is able to express the choice as a condition on a univariate random variable [see equations (34) and (35)]; (ii) by transforming the distribution of that random variable to normality, he is able to relocate the problem in the conventional Heckman-Lee framework.

Assume that the cumulative distribution function for ϵ_2 is given by $F_{\epsilon_2}(\epsilon_2 \mid \theta' z)$ where θ is a matrix of unknown coefficient vectors in each selection equation, i.e. $\theta = (\theta_0 \mid \theta_1 \mid \theta_2 \mid \theta_3)$. Lee (1983) relies on the transformation

$$J_{\epsilon_2}(\epsilon_2 \mid \theta' z) = \Phi^{-1}(F(\epsilon_2 \mid \theta' z)), \tag{36}$$

where Φ is the standard normal distribution function. He makes the following assumption:

$$\text{The joint distribution of } (u_3,\ J_{\epsilon_2}(\epsilon_2 \mid \theta' z)) \text{ does not depend on the index } \theta' z. \tag{37}$$

This assumption is contested in the literature since it has strong implications. First, the transformation implies that the marginal distribution of $J_{\epsilon_2}(\epsilon_2 \mid \theta' z)$ is independent of $\theta' z$, but it does not tell us anything about the joint distribution of u_3 and $J_{\epsilon_2}(\epsilon_2 \mid \theta' z)$. Second, Schmertmann (1994) shows that under index independence assumption of Lee (1983), $\Upsilon_l = \rho_l - \rho_2$ for $l = 0, 1, 3$ must have the same sign. Moreover, if we assume $\eta_l - \eta_2$'s are identically distributed, Υ_l's will be identical for $l = 0, 1, 3$ (Bourguignon et al. (2007)).

Lee (1983) also assumes that

$$E(u_3 \mid \epsilon_2,\ \theta' z) = c_2 J_{\epsilon_2}(\epsilon_2 \mid \theta' z), \tag{38}$$

where c_2 is the correlation between u_3 and $J_{\epsilon_2}(\epsilon_2 \mid \theta' z)$. Clearly Equation (38) would follow if u_3 and ϵ_2 are bivariate normally distributed, but this is a strong assumption. Another way to justify the functional form is to translate the arbitrary distribution of u_3 to standard normal via $J_{u_3}(u_3)$ say, before assuming that the joint distribution of $(J_{u_3}(u_3),\ J_{\epsilon_2}(\epsilon_2 \mid \theta' z))$ is standard bivariate normal, independently of $\theta' z$. When u_3 is not normally distributed, this equation may be viewed as a best linear approximation to the conditional expectation function.

This approach yields

$$E(u_3 \mid LFP' = 2, \theta' z) = -c_2 \lambda_2^{LEE}, \tag{39}$$

where $\lambda_2^{LEE} = \frac{\phi(J_{\epsilon_2}(0 \mid \theta' z))}{F_{\epsilon_2}(0 \mid \theta' z)}$. The regression equation becomes

$$\log(wage) = \beta_3' X_3 + \gamma^{LEE} \lambda_2^{LEE} + v_3 \tag{40}$$

where $\gamma^{LEE} = -\sigma_3 c_2$ with $\sigma_3 = V(u_3)$ and v_3 is a heteroscedastic random disturbance term with $E(v_3 \mid \Lambda) = 0$. Note that λ_2^{LEE} is the familiar term in conventional single selection correction. Maximum likelihood estimates of θ which come from the first step are used to obtain consistent estimates of λ_2^{LEE}, and the second step estimates of β_3 and γ^{LEE} are obtained via linear regression.

4.2.2 Dubin and McFadden's (1984) Model (DMF) and Its Two Variants

All three approaches share the assumption that the conditional expectation function of the regression equation random disturbance is a linear function of the selection equation random disturbances:[6]

$$E(u_3 \mid \eta_1, \eta_2, \eta_3, \eta_4) = \frac{\sqrt{6}}{\pi} \sum_{j=0}^{3} \rho_j (\eta_j - E(n_j)). \tag{41}$$

Since η_j's are assumed to have independent Gumbel distributions, it follows that

$$E(\eta_2 - E(n_2) \mid LFP' = 2,\ \theta' z) = -\log(\pi_2), \tag{42}$$

$$E(\eta_l - E(n_l) \mid LFP' = 2,\ \theta' z) = \frac{\pi_l \log(\pi_l)}{1 - \pi_l} \text{ for } l = 0,\ 1, 3. \tag{43}$$

where log denotes the natural logarithm. Note that $E(\eta_j) = 0.57721..$, the Euler-Mascheroni con-

[6]Linear conditional expectation function assumption is a convenient starting point going back to Olsen (1980). See Vella (1998) for a broad discussion.

stant, and $V(\eta_j) = \pi^2/6$. To assure that $E(u_3 \mid X_3) = 0$, Dubin and McFadden (1984) also assume

$$\sum_{j=0}^{3} \rho_j = 0. \tag{44}$$

Under the assumptions given by Equations (41) - (44), the regression equation becomes

$$\log(wage) = \beta_3' X_3 + \sum_{l=0,1,3} \gamma_l^{DMF} \lambda_l^{DMF} + v_3, \tag{45}$$

where $\lambda_l^{DMF} = \frac{\pi_l \log(\pi_l)}{1-\pi_l} + \log(\pi_2)$, $\gamma_l^{DMF} = \frac{\sigma_3 \sqrt{6}}{\pi} \rho_l$ for $l = 0,\ 1,\ 3$ and v_3 is a heteroscedastic random disturbance term with $E(v_3 \mid \Lambda) = 0$. Estimates of λ_l^{DMF} are formed using the selection step maximum likelihood estimates. The estimates of the parameters in Equation (45) are obtained in the second step via linear regression.[7]

Observe that if the linear functional forms in Equations (39) and (41) are not contested, DMF can be regarded as superior to LEE because the only restriction on the correlation structure is via Equation (44). However, Bourguignon et al. (2007) show via Monte Carlo experiments that the restriction in Equation (44) causes biased results when incorrectly imposed; in contrast, efficiency loss is small by not imposing it when the restriction holds. Hence, they drop the restriction in Equation (44) and refer to it as the first variant of Dubin and McFadden (1984) model (DMF1). For the DMF1 version, the regression equation is given by

$$\log(wage) = \beta_3' X_3 + \sum_{l=0,1,2,3} \gamma_l^{DMF1} \lambda_l^{DMF1} + v_3, \tag{46}$$

where $\lambda_2^{DMF1} = -\log(\pi_2)$, $\lambda_l^{DMF1} = \frac{\pi_l \log(\pi_l)}{1-\pi_l}$ for $l = 0,\ 1,\ 3$, $\gamma_l^{DMF1} = \frac{\sigma_3 \sqrt{6}}{\pi} \rho_l$ for $l = 0, 1, 2, 3$, and v_3 is a zero mean, heteroscedastic random disturbance term. Observe that if we set $\rho_2 = -(\rho_0 + \rho_1 + \rho_3)$, we get Equation (45). Hence DMF1 is more general compared to DMF. After forming the four selectivity correction terms using the first step maximum likelihood estimates, we run a least squares on Equation (46) to obtain the second step estimates of the DMF1 model.

Since the linearity assumption in Equation (41) restricts the distribution of u_3, Bourguignon

[7] Even though the second step does not provide us with an estimate for ρ_2 , it can be calculated using Equation (44).

et al. (2007) propose a second variant of Dubin and McFadden (1984) model (DMF2). Assuming that the cumulative distribution function for η_j is given by $G(\eta_j)$ they rely on the transformation idea of Lee (1983). Let

$$\eta_j^* = J(\eta_j) = \Phi^{-1}(G(\eta_j)) \tag{47}$$

and assume that conditional expectation function of interest is a linear function η_j^*'s (rather than the η_j's as in DMF and DMF1):

$$E(u_3 \mid \eta_1, \eta_2, \eta_3, \eta_4) = \sum_{j=1}^{4} \rho_j^* \eta_j^*, \tag{48}$$

where ρ_j^* is the correlation between u_3 and η_j^*. Note that the assumption in Equation (48) holds if u_3 and η_j^*'s are multivariate normally distributed. Assuming that $g(.)$ denotes the density function of η_j, let

$$M(\pi_j) = \int J(v - \log(\pi_j))g(v)dv \text{ for } j = 0,\ 1,\ 2,\ 3. \tag{49}$$

Then, Bourguignon et al. (2007) derive the following:

$$E(\eta_2^* \mid LFP' = 2, \theta' z) = M(\pi_2). \tag{50}$$

$$E(\eta_l^* \mid LFP' = 2, \theta' z) = M(\pi_l)\frac{\pi_l}{\pi_l - 1} \text{ for } l = 0, 1, 3. \tag{51}$$

As a result, the regression equation can be written as

$$\log(wage) = \beta_3' X_3 + \sum_{l=0,1,2,3} \gamma_l^{DMF2} \lambda_l^{DMF2} + v_3, \tag{52}$$

where $\lambda_2^{DMF2} = M(\pi_2)$, $\lambda_l^{DMF2} = M(\pi_l)\frac{\pi_l}{\pi_l - 1}$ for $l = 0,\ 1,\ 3$, $\gamma_l^{DMF2} = \sigma_3 \rho_l^*$ for $l = 0,\ 1,\ 2,\ 3$, and v_3 is a heteroscedastic random disturbance term with $E(v_3 \mid \Lambda) = 0$. Even though M's do not have closed form solutions, they can be computed using Gauss-Laguerre quadrature. Bourguignon et al. (2007) computes these terms using Davis and Polonsky (1964). Second step estimates of DMF2 are obtained by running a linear regression on Equation (52) after forming consistent estimates of the selectivity correction terms using the maximum likelihood estimates from the first step.

4.3 Dahl's (2002) Model

Using a semi-parametric model, Dahl (2002) solves an ambitious problem dealing with 51 choice outcomes, each of which has its own set of control functions. In the spirit of Ahn and Powell (1993) and Das et al. (2003), Dahl (2002) exploits series approximations to the true control functions. An index sufficiency assumption is invoked to express control functions as polynomials in the choice probabilities. However, the curse of dimensionality precludes use of all choice probabilities. To render the problem tractable, Dahl (2002) proposes restricting the set of choice probabilities to a subset $Q^{'}$ of the full choice set Q [in our case $Q = \{\pi_0,\ \pi_1,\ \pi_2\}$, with $\pi_3 = 1 - (\pi_0 +\ \pi_1 +\ \pi_2)$]. His two assumptions impose the following structure on the conditional joint density function of u_3 and ϵ_2:

$$\varphi(u_3, \epsilon_2 \mid \theta^{'} z) = \varphi(u_3, \epsilon_2 \mid Q) = \varphi(u_3, \epsilon_2 \mid Q^{'}), \tag{53}$$

The control function in turn becomes

$$E(u_3 \mid LFP^{'} = 2, \theta^{'} z) = E(u_3 \mid LFP^{'} = 2, Q^{'}) = \psi(Q^{'}). \tag{54}$$

Here $\psi(.)$ is a unique function, which Dahl (2002) obtains by assuming that the choice probabilities are invertible. Then, the regression function can be expressed as

$$\log(wage) = \beta_3^{'} X_3 + \psi(Q^{'}) + v_3\ , \tag{55}$$

where v_3 is a heteroscedastic random disturbance term with $E(v_3 \mid \Lambda) = 0$. This approach has two shortcomings. First, as stated by Dahl (2002), we generally can not test the index sufficiency assumption. Second, reducing dimensionality is equivalent to restricting the correlations between η_j and u_3 for $j = 0,\ 1,\ 2,\ 3$ (which we term ρ_j) to be a function of the elements of $Q^{'}$(Bourguignon et al. (2007)).

A special case of this model proposed by Dahl (2002) is to assume that the set $Q^{'}$ is a singleton containing only the chosen alternative, in our case $LFP^{'} = 2$. Notice that this assumption is another way of reducing a multiple index problem to a single index problem, which Lee (1983) justified via his first order statistic approach. In the special case, the regression equation becomes

$$\log(wage) = \beta_3^{'} X + \psi(\pi_2) + v_3 \tag{56}$$

Dahl (2002) estimates the selection probabilities nonparametrically, as cell averages. Our data set does not support a nonparametric approach. By relying on the MLM at the first step, we conveniently situate Dahl's approach among the others we have discussed. However, unlike the other Multinomial Logit based selection correction models, Dahl (2002) does not assume a linear conditional expectation function. The form of the regression equation is derived from the behavioral model under the index sufficiency assumption. Following Bourguignon et al. (2007), we consider two variants of Dahl (2002): in the first variant, we express ψ to be a function of the probability of the relevant choice (π_2 in our case) up to fourth order, and in the second, we let ψ to be a function of all choice probabilities, each as a fourth-order polynomial, and with all interactions between them. This latter approach is identified as DAHL2 in our analysis. In both approaches, π_2 is estimated in the first step.

4.4 Application Revisited

We revisit the application given in Section 3 and estimate it using the Multinomial Logit based selection correction models. Conveniently, the estimation steps can be implemented using the STATA module *selmlog* developed by Bourguignon et al. (2007).[8] The results obtained from the selection step are presented in Table 4. Observe that while we set $P_3 = P_1 + P_2$ in our Edgeworth based selection correction following Tunali and Baslevent (2006), Multinolmial Logit based selection correction models cannot accomodate this restriction. As seen in Table 4, coefficients in columns 1 and 2 do not sum to those in column 3. Unlike our model, logit based models use an independent equation for the unemployed. Since selectivity is our main concern, we proceed with the second step.

Once the method is chosen, the selectivity correction terms are formed in line with the expressions given above and the regression estimates for wage workers are obtained. In LEE, DMF, DMF1 and DMF2 weighed least squares was employed to take heteroscedasticity into account. The

[8]The STATA ado file is available at http://www.parisschoolofeconomics.com/gurgand-marc/selmlog/selmlog13.html.

Table 4: Selection Step for Multinomial Logit Based Selection Correction Models

Variable	Self-employed		Wage Worker		Unemployed	
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Age	0.085	0.090	0.575***	0.067	0.128	0.087
Age Squared/100	−0.113	0.126	−0.839***	0.097	−0.264**	0.129
Literate without a Diploma	0.402	0.282	0.371	0.262	0.310	0.295
Elementary School	0.194	0.221	0.500***	0.185	0.287	0.207
Middle School	1.083***	0.311	1.493***	0.233	0.863***	0.282
High School	0.572	0.348	3.226***	0.189	1.642***	0.244
University	1.630***	0.559	5.089***	0.235	1.735***	0.476
Husband Self-employed	0.082	0.160	−1.318***	0.125	−0.863***	0.166
Children Aged $0-2$	−0.362	0.232	−0.390***	0.134	−0.505***	0.174
Children Aged $3-5$	−0.163	0.193	−0.429***	0.118	−0.176	0.147
Female Children aged $6-14$	0.381**	0.178	−0.260**	0.114	−0.086	0.151
Male Children aged $6-14$	0.160	0.180	−0.405***	0.113	0.020	0.152
Extended Household	0.634*	0.343	0.094	0.256	−0.156	0.400
Ext. HH × Children Aged $0-2$	0.220	0.562	0.260	0.338	0.099	0.566
Ext. HH × Children Aged $3-5$	−0.067	0.523	0.446	0.308	−0.450	0.543
Ext. HH × Female Ch. Aged $6-14$	−0.702	0.487	−0.007	0.312	0.183	0.505
Ext. HH × Male Ch. Aged $6-14$	−0.301	0.476	0.293	0.306	−0.090	0.511
Share of Textiles	−0.163	0.539	1.212***	0.433	0.823*	0.498
Share of Agriculture	−1.115	1.248	1.939**	0.841	−1.397	1.039
Share of Finance	−6.208	4.435	−2.910	3.615	−4.184	4.793
Migration Rate	−10.18***	3.356	7.469**	3.070	−3.402	3.406
Aegean	−0.729**	0.310	0.551**	0.227	0.507*	0.260
South	−0.105	0.291	−0.283	0.240	−0.033	0.315
Central	−1.739***	0.361	0.670**	0.292	0.090	0.343
North West	−1.559***	0.552	1.012**	0.398	0.350	0.412
East	−1.386**	0.543	0.495	0.415	0.118	0.481
South East	−0.451	0.468	−0.223	0.361	−1.750***	0.609
North East	−2.790***	1.062	1.037**	0.463	0.347	0.485
Population 200,000 or more	−16.509***	2.978	2.749	2.309	−4.006	2.941
Population 1 million or more	−4.545***	1.240	2.400*	1.346	3.757***	1.465
Share of Welfare Party	0.766***	0.263	0.303*	0.174	0.088	0.230
Share of Left of Center	−0.372	0.259	0.126	0.185	−0.523**	0.249
Constant	−1.993	1.693	−14.708***	1.343	−5.258***	1.605
Number of Observations	8962					
Log-Likelihood Without Covariates	−4603.94					
Log-Likelihood With Covariates	−3528.64					

Notes: Robust standard errors are reported. * is significant at 10%; ** is significant at 5%; *** is significant at 1%. The reference group is non-participation.

relevant weights can be found in the appendix of Bourguignon et al. (2007). Standard errors are estimated using the bootstrap method by drawing with replacement 100 and 400 samples from the subsample of interest, which in our case has 735 observations. Since additional bootstrapping does not lead to big differences in the standard errors of the variables, we only report the estimates of the version with 100 replications.

Table 5: LEE Estimates of the Wage Equation

Variable	Coefficient	Std. Error
Experience	0.033**	0.014
Experience Squared/100	−0.068*	0.041
Literate without a Diploma	−0.110	0.190
Elementary School	−0.003	0.124
Middle School	0.250*	0.143
High School	0.486***	0.184
University	1.067***	0.256
Share of Textiles	−0.391***	0.144
Share of Agriculture	0.156	0.254
Share of Finance	4.211***	1.194
Aegean	−0.132**	0.067
South	0.153	0.095
Central	0.029	0.077
North West	−0.171	0.110
East	0.137	0.130
South East	0.184*	0.112
North East	−0.129	0.125
$\hat{\lambda}_2^{LEE}$	−0.009	0.114
Constant	−0.967***	0.321
σ_{33}	0.279***	0.062
c_2	−0.016	0.178
Number of Observations	735	

Notes: Standard errors are estimated using 100 bootstrap replications. * is significant at 10%; ** is significant at 5%; *** is significant at 1%.

Table 5 gives the results based on LEE. The *selmlog* module provides consistent estimates of σ_3 and c_2 without restricting c_2 to the $[-1, 1]$ interval. The finding from LEE supports random selection since the selectivity correction term has a $p - value = 0.939$. Table 6 reports the second step estimates from DMF, DMF1 and DMF2. The *selmlog* module provides consistent estimates of σ_{33}, ρ_j and ρ_j^* for $j = 0$, 1, 2, 3 without restricting the correlation estimates to their theoretical ranges. The selection correction terms are jointly insignificant in all the models ($p - value = 0.667$ for DMF, 0.8021 for DMF1, and 0.9564 for DMF2). The findings from all four models contradict the results based on trivariate normality and our robust specification. In fact results obtained from these models are very similar to those for random selection, reported in Table 3.

Table 6: DMF, DMF1 and DMF2 Estimates of the Wage Equation

Variable	DMF2		DMF1		DMF	
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Experience	0.031**	0.012	0.028**	0.014	0.029**	0.013
Experience Squared/100	−0.064*	0.034	−0.057	0.036	−0.060*	0.036
Literate without a Diploma	−0.127	0.195	−0.138	0.190	−0.128	0.182
Elementary School	−0.030	0.137	−0.044	0.137	−0.032	0.124
Middle School	0.191	0.176	0.167	0.187	0.208	0.143
High School	0.453**	0.197	0.409**	0.186	0.436***	0.162
University	1.044***	0.299	1.041***	0.285	1.052***	0.225
Share of Textiles	−0.419**	0.166	−0.377*	0.197	−0.329**	0.160
Share of Agriculture	0.176	0.278	0.146	0.271	0.105	0.234
Share of Finance	4.593***	1.728	4.400**	2.166	3.920***	1.247
Aegean	−0.115	0.085	−0.114	0.121	−0.113	0.095
South	0.155*	0.105	0.155	0.134	0.159	0.102
Central	0.045	0.088	0.044	0.107	0.035	0.081
North West	−0.150	0.125	−0.138	0.135	−0.131	0.111
East	0.156	0.143	0.149	0.191	0.145	0.129
South East	0.199	0.140	0.178	0.192	0.140	0.111
North East	−0.101	0.173	−0.077	0.176	−0.090	0.125
$\hat{\lambda}_0^{DMF_m}$	−0.142	0.256	−0.361	0.410	−0.352	0.346
$\hat{\lambda}_1^{DMF_m}$	−0.667	1.415	−0.290	2.475	0.136	0.544
$\hat{\lambda}_2^{DMF_m}$	−0.025	0.974	−0.064	0.060	—	
$\hat{\lambda}_3^{DMF_m}$	0.090	0.546	0.084	0.413	0.273	0.392
Constant	−1.002***	0.325	−1.052**	0.439	−1.001***	0.352
σ_{33}	0.309***	0.132	0.417	0.420	0.432	0.347
ρ_0	—		−0.718	0.529	−0.686	0.476
ρ_1	—		−0.576	2.588	0.266	0.784
ρ_2	—		−0.127	0.096	—	
ρ_3	—		0.168	0.610	0.532	0.574
ρ_0^*	−0.255	0.432	—		—	
ρ_1^*	−1.200	1.932	—		—	
ρ_2^*	−0.046	0.160	—		—	
ρ_3^*	0.163	0.831	—		—	
Number of Observations	735		735		735	

Notes: Standard errors are estimated using 100 bootstrap replications. * is significant at 10%; ** is significant at 5%; *** is significant at 1%.

Table 7 provides the second step estimates obtained from DAHL1 and DAHL2. Unfortunately we cannot test for non-random selection since the *selmlog* module neither provides robust standard errors for correction terms nor computes R^2's. Upon comparing the slope estimates of the education dummies in Table 7 with those obtained from our robust correction in Table 3, we observe that both

Table 7: DAHL 1 and DAHL 2 Estimates of the Wage Equation

Variable	DAHL2		DAHL1	
	Coefficient	Std. Error	Coefficient	Std. Error
Experience	0.027*	0.016	0.033**	0.014
Experience Squared/100	−0.050	0.039	−0.067*	0.039
Literate without a Diploma	−0.204	0.187	−0.140	0.182
Elementary School	−0.090	0.120	−0.041	0.120
Middle School	0.158	0.162	0.182	0.146
High School	0.592***	0.201	0.533***	0.170
University	1.066***	0.291	1.147***	0.240
Share of Textiles	−0.274	0.195	−0.370***	0.121
Share of Agriculture	0.114	0.276	0.140	0.233
Share of Finance	4.474***	1.636	4.138***	1.168
Aegean	−0.122	0.109	−0.142*	0.077
South	0.146	0.112	0.130	0.086
Central	0.098	0.097	0.029	0.073
North West	−0.103	0.125	−0.181**	0.091
East	0.206	0.152	0.113	0.113
South East	0.003	0.165	0.172	0.109
North East	−0.074	0.161	−0.122	0.113
Constant	−267.265	418.920	−1.030***	0.223
Number of Observations	735		735	

Notes:

Standard errors are estimated using 100 bootstrap replications. * is significant at 10%; ** is significant at 5%; *** is significant at 1%.

DAHL1 and DAHL2 yields results close to ours. This may not be surprising given the fact that both approaches involve series approximations. As we argued in Section 3.2, returns to education estimates are likely to be biased unless selectivity is properly accounted. It appears that the control functions in Dahl (2002) do a better job in doing the adjustment for selectivity compared to the other Moultinomial Logit based selection correction alternatives. Note that the point estimates of the other statistically significant variables in DAHL1 differ from ours by up to 20 percent. The discrepancy rises to 30 percent in the case of DAHL2.

5 Conclusion

This paper discusses correction of selectivity bias in models of double selection. The approach proposed here relaxes the conventional trivariate normality assumption that builds on the Heckman-Lee tradition. Instead, we let the distribution of the disturbance term in the regression equation free and assume that the random disturbance terms of the two selection equations are bivariate normally distributed exploiting and Edgeworth expansion. To render our approach tractable, we truncate the terms that provide the series approximation consistent with the literature. Since we allow for skewness and kurtosis, our methodology offers a major improvement over trivariate normality. Besides, we can test whether the trivariate normality specification is tenable. To assess the relative merits of the new method, we compare our methodology with Multinomial Logit based selection correction models.

Arguably the main drawback of our model is the fact that extension beyond two selection equations cannot be handled. Multinomial Logit based correction models do not have this handicap. However these models define an independent selection equation for each possible category and impose the questionable IIA structure. Our model has the advantage of allowing correlation between the random disturbances of the two selection equations. In addition, problems in which the states are not pairwise disjoint are tractable via our specification.

To illustrate the differences among various selection correction models, an empirical example involving the wage determinants of regular and casual female workers in Turkey is presented. Only a fraction of women living in urban areas of Turkey participate and participation probability increases dramatically in response to an increase in education level. Thus, there is reason to believe that wage workers are not a random subsample. Indeed, both the conventional trivariate normality based correction method and our Edgeworth expansion based approach support this conjecture. Notably, the example provides evidence in favor of our new approach, rather than the trivariate normal version. We conclude that it is important to allow for kurtosis and skewness.

As examples of Multinomial Logit based selection correction models, we analyzed the methods provided by Lee (1983), Dubin and McFadden (1984) plus its two variants, and Dahl (2002). Except for Dahl (2002), the alternatives we considered failed to detect any selectivity in an empirical context conducive to selectivity bias. The substantive results from DAHL1 and DAHL2 were closer to those obtained under our Edgewoth based selectivity correction.

Since our selection model is an improvement the conventional trivariate normality assumption builded on the Heckman-Lee tradition, and allows for correlation between selection equations and is capable of accomodating pairwise joint states compared to Multinomial Logit based selection correction models, we are unable to identify any reasons for opting for other methods when a problem can be formulated as a double selection model.

A Appendix

A.1 Derivation of Equations (11) and (12)

Let $f(u_1, u_2, \tilde{u}_3)$ be the joint density of u_1, u_2 and $\tilde{u}_3$. The triplet $(u_{1i}, u_{2i}, \tilde{u}_{3i})$ is assumed to be independently and identically distributed across individuals with zero mean vector and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix},$$

and is independent of X_{ki} for $k = 1, 2, 3$. For brevity, we suppress the conditioning on covariate matrix X and the individual subscript i throughout the section. Our aim is to obtain an expression for $E(\tilde{u}_3 \mid u_1, u_2)$.

We denote the standard trivariate normal density $STVN(\rho_{12}, \rho_{13}, \rho_{23})$ by $g(u_1, u_2, u_3)$ and the standard bivariate normal density $SBVN(\rho_{12})$ by $g(u_1, u_2)$. Let $k(\tilde{u}_3 \mid u_1, u_2)$ be the conditional density of $\tilde{u}_3$ given u_1 and u_2. We assume that $k(\tilde{u}_3 \mid u_1, u_2) g(u_1, u_2) = f(u_1, u_2, \tilde{u}_3)$, i.e. we specify the joint distribution of (u_1, u_2) as standard bivariate normal, but allow the conditional distribution of $\tilde{u}_3$ given u_1 and u_2 to be non-normal. Let

$$A = \sqrt{1 - (\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23})}, \tag{A.1.1}$$

$$a = \frac{1}{\sqrt{1 - \rho_{12}^2}}, \tag{A.1.2}$$

$$b_1 = \rho_{13} - \rho_{12}\rho_{23}, \tag{A.1.3}$$

$$b_2 = \rho_{23} - \rho_{12}\rho_{13}, \tag{A.1.4}$$

$$c = \frac{a}{A}\left(b_1 u_1 + b_2 u_2\right), \tag{A.1.5}$$

$$u_1^*(u_1, u_2) = u_1^* = a(u_1 - \rho_{12} u_2), \tag{A.1.6}$$

$$u_2^*(u_1, u_2) = u_2^* = a(u_2 - \rho_{12} u_1), \tag{A.1.7}$$

$$u_3^* = \frac{u_3}{Aa} - c. \tag{A.1.8}$$

We can write:

$$E\left(\frac{\tilde{u}_3}{Aa} \mid u_1, u_2\right) g(u_1, u_2) = \int_{-\infty}^{\infty} \frac{\tilde{u}_3}{Aa} k(\tilde{u}_3 \mid u_1, u_2) g(u_1, u_2) d\tilde{u}_3. \tag{A.1.9}$$

Equivalently,

$$E(\tilde{u}_3 \mid u_1, u_2) \frac{g(u_1, u_2)}{Aa} = \int_{-\infty}^{\infty} \frac{\tilde{u}_3}{Aa} f(u_1, u_2, \tilde{u}_3) d\tilde{u}_3. \tag{A.1.10}$$

For the Type AAA surface defined in Section 2.3, we may expand $f(u_1, u_2, \tilde{u}_3)$ in terms of a series of derivatives of $g(u_1, u_2, u_3)$ via Equation (7). We truncate the expansion by keeping terms with $r + s + p \leq 4$. This allows us to set $A_{rsp} = K_{rsp}$ where K's denote cumulants (Lahiri and Song (1999)). Thus, the truncated version of Equation $(A.1.10)$ is:

$$\begin{aligned} E(\tilde{u}_3 \mid u_1, u_2) \tfrac{g(u_1, u_2)}{Aa} &= \textstyle\int_{-\infty}^{\infty} \frac{u_3}{Aa} g(u_1, u_2, u_3) du_3 \\ &+ \sum_{r+s+p=3}^{4} \tfrac{(-1)^{r+s+p}}{r!s!p!} K_{rsp} \textstyle\int_{-\infty}^{\infty} \frac{u_3}{Aa} D_{u_1}^r D_{u_2}^s D_{u_3}^p g(u_1, u_2, u_3) du_3. \end{aligned} \tag{A.1.11}$$

Let

$$I_1 = \int_{-\infty}^{\infty} \frac{u_3}{Aa} g(u_1, u_2, u_3)\, du_3, \tag{A.1.12}$$

and

$$I_{rsp} = \int_{-\infty}^{\infty} \frac{u_3}{Aa} D_{u_1}^r D_{u_2}^s D_{u_3}^p g(u_1, u_2, u_3) du_3. \tag{A.1.13}$$

Then, Equation $(A.1.11)$ may be expressed as

$$E(\tilde{u}_3 \mid u_1, u_2) \frac{g(u_1, u_2)}{Aa} = I_1 + \sum_{r+s+p=3}^{4} \frac{(-1)^{r+s+p}}{r!s!p!} K_{rsp} I_{rsp}. \tag{A.1.14}$$

Claim: $g(u_1, u_2, u_3) = \frac{\phi(u_1)\phi(u_2^*)\phi(u_3^*)}{A}$, where ϕ is the univariate standard normal density.

Proof: Let $g(u_2 \mid u_1)$ be the conditional density function of u_2 given u_1 and $g(u_3 \mid u_1, u_2)$ be the conditional density of u_3 given u_1 and u_2. Then, $g(u_1, u_2, u_3)$ can be written as a product of marginal and conditional distributions:

$$g(u_1, u_2, u_3) = g(u_1) g(u_2 \mid u_1) g(u_3 \mid u_1, u_2). \tag{A.1.15}$$

Clearly, $g(u_1) = \phi(u_1)$. We know from Goldberger (1991) that

$$g(u_2 \mid u_1) = a\phi\left(u_2^*\right). \tag{A.1.16}$$

Now, let us find $g(u_3 \mid u_1, u_2)$. Using the formula in Goldberger (1991), we obtain

$$E(u_3 \mid u_1, u_2) = a^2(b_1 u_1 + b_2 u_2), \tag{A.1.17}$$

$$Var(u_3 \mid u_1, u_2) = A^2 a^2. \tag{A.1.18}$$

Then,

$$g(u_3 \mid u_1, u_2) = \frac{\phi\left(u_3^*\right)}{Aa}. \tag{A.1.19}$$

Substituting Equation $(A.1.16)$ and Equation $(A.1.19)$ into Equation $(A.1.15)$, we obtain

$$g(u_1, u_2, u_3) = \frac{\phi(u_1)\phi\left(u_2^*\right)\phi\left(u_3^*\right)}{A} \qquad \text{Q.E.D.} \tag{A.1.20}$$

Inserting Equation $(A.1.20)$ into Equation $(A.1.13)$ and making the transformation $z_3 = u_3^*$, so $Aadz_3 = du_3$, we get

$$I_{rsp} = \int_{-\infty}^{\infty} (z_3 + c)\, D_{u_1}^r D_{u_2}^s D_{u_3}^p \phi(u_1)\phi\left(u_2^*\right)\phi(z_3) a dz_3. \tag{A.1.21}$$

Let $\phi^{(p)}$ denote the p^{th} derivative of the univariate standard normal density. Observe that

$$D_{u_3}^p \phi(z_3) = \phi^{(p)}(z_3)\left(\frac{1}{Aa}\right)^p, \tag{A.1.22}$$

and

$$D_{z_3}^p \phi(z_3) = \phi^{(p)}(z_3). \tag{A.1.23}$$

Hence,

$$\left(\frac{1}{Aa}\right)^p D_{z_3}^p \phi(z_3) = D_{u_3}^p \phi(z_3). \tag{A.1.24}$$

Substituting Equation $(A.1.24)$ into Equation $(A.1.21)$, we obtain

$$I_{rsp} = \frac{1}{A^p a^{p-1}} \int_{-\infty}^{\infty} (z_3 + c)\, D_{u_1}^r D_{u_2}^s D_{z_3}^p \phi(u_1)\phi\left(u_2^*\right)\phi(z_3) dz_3. \tag{A.1.25}$$

Making another transformation $z_3 = u_3$ (abusing the notation), so $D_{z_3}^p \phi(z_3) = D_{u_3}^p \phi(u_3)$ and $dz_3 = du_3$, we have

$$I_{rsp} = \frac{1}{A^p a^{p-1}} \int_{-\infty}^{\infty} (u_3 + c)\, D_{u_1}^r D_{u_2}^s D_{u_3}^p \phi(u_1)\phi\left(u_2^*\right)\phi(u_3) du_3. \tag{A.1.26}$$

The definition of r^{th} order univariate Hermite polynomial can be obtained by setting $s = 0$ in Equation (9). In other words, $D_{u_1}^r \phi(u_1) = (-1)^r H_r(u_1)\phi(u_1)$ which implies that $H_0(u) = u$ and $H_1(u) = u$. Then, we can write

$$u_3 + c = H_1(u_3 + c) = \sum_{i=0}^{1} \begin{pmatrix} 1 \\ i \end{pmatrix} c^i H_{1-i}(u_3), \tag{A.1.27}$$

where $\begin{pmatrix} 1 \\ i \end{pmatrix}$ denotes the binomial coefficient. Incorporating Equation $(A.1.27)$ into Equation $(A.1.26)$, we get

$$I_{rsp} = \frac{(-1)^p}{A^p a^{p-1}} D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right) \sum_{i=0}^{1} \begin{pmatrix} 1 \\ i \end{pmatrix} c^i \int_{-\infty}^{\infty} H_p(u_3) H_{1-i}(u_3)\phi(u_3) du_3. \tag{A.1.28}$$

For any x, the orthogonality property of Hermite polynomials allows us to write

$$\int_{-\infty}^{\infty} H_n(x) H_m(x) \exp\left(-x^2/2\right) dx = n!\sqrt{2\pi}\delta_{n,m}, \tag{A.1.29}$$

where $\delta_{n,m}$ is the Kronecker product, $\delta_{n,m} = \begin{Bmatrix} 1 \text{ if } n = m \\ 0 \text{ if } n \neq m \end{Bmatrix}$. Since $\delta_{p,1-i} = 1 \Longleftrightarrow p = 1 - i \Longleftrightarrow i = 1 - p \Longleftrightarrow p \leq 1$,

$$\begin{aligned}\int_{-\infty}^{\infty} H_p(u_3)H_{1-i}(u_3)\phi(u_3)du_3 &= \int_{-\infty}^{\infty} H_p(u_3)H_{1-i}(u_3)\frac{\exp(-u_3^2/2)}{\sqrt{2\pi}}du_3 \\ &= p!\sqrt{2\pi}\delta_{p,1-i}\frac{1}{\sqrt{2\pi}} = p! \text{ for } p \leq 1 \text{ \& } p = 1-i.\end{aligned} \tag{A.1.30}$$

Also observe that $\int_{-\infty}^{\infty} H_p(u_3)H_{1-i}(u_3)\phi(u_3)du_3 = 0$ for $p > 1$. Returning to Equation $(A.1.28)$, we conclude that

$$I_{rsp} = 0 \text{ for } p > 1. \tag{A.1.31}$$

Moreover, substituting Equation $(A.1.30)$ into Equation $(A.1.28)$, we obtain:

$$I_{rsp} = \frac{(-1)^p}{A^p a^{p-1}} D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right) \begin{pmatrix} 1 \\ 1-p \end{pmatrix} c^{1-p} p! \text{ for } p \leq 1. \tag{A.1.32}$$

Equivalently,

$$I_{rsp} = \frac{(-1)^p}{A^p a^{p-1}} D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right) \frac{c^{1-p}}{(1-p)!} \text{ for } p \leq 1. \tag{A.1.33}$$

Substituting $\frac{a}{A}(b_1u_1 + b_2u_2)$ back instead of c, we obtain

$$I_{rsp} = \frac{(-1)^p}{Aa^{2p-2}} \frac{1}{(1-p)!} D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right) (b_1u_1 + b_2u_2)^{1-p} \text{ for } p \leq 1. \tag{A.1.34}$$

We can conclude from Equations $(A.1.12)$ and $(A.1.13)$ that $I_1 = I_{000}$. Then, we can write I_1 as

$$I_1 = \frac{a^2}{A}\phi(u_1)\phi\left(u_2^*\right)(b_1u_1 + b_2u_2). \tag{A.1.35}$$

Setting $p = 0$ and 1 in Equation $(A.1.34)$, we respectively get:

$$I_{rs0} = \frac{a^2}{A} D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right)(b_1u_1 + b_2u_2), \tag{A.1.36}$$

and

$$I_{rs1} = -\frac{1}{A} D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right). \tag{A.1.37}$$

After substituting Equations $(A.1.31)$, $(A.1.35)$, $(A.1.36)$ and $(A.1.37)$ into Equation $(A.1.16)$ and doing some manipulations, we get

$$\begin{aligned} E(\tilde{u}_3 \mid u_1, u_2) g(u_1, u_2) &= a^3 \phi(u_1) \phi\left(u_2^*\right) \left(b_1 u_1 + b_2 u_2\right) \\ &+ \sum_{r+s=3}^{4} \tfrac{(-1)^{r+s}}{r!s!} K_{rs0} a^3 D_{u_1}^r D_{u_2}^s \phi(u_1) \phi\left(u_2^*\right) \left(b_1 u_1 + b_2 u_2\right) \\ &+ \sum_{r+s=2}^{3} \tfrac{(-1)^{r+s}}{r!s!} K_{rs1} a D_{u_1}^r D_{u_2}^s \phi(u_1) \phi\left(u_2^*\right). \end{aligned} \tag{A.1.38}$$

Since $a\phi(u_1)\phi(u_2^*) = g(u_1, u_2)$, which may be obtained by multiplying Equation $(A.1.16)$ by $\phi(u_1)$, Equation $(A.1.38)$ can be written as:

$$\begin{aligned} E(\tilde{u}_3 \mid u_1, u_2) g(u_1, u_2) &= a^2 g(u_1, u_2) \left(b_1 u_1 + b_2 u_2\right) \\ &+ \sum_{r+s=3}^{4} \tfrac{(-1)^{r+s}}{r!s!} K_{rs0} a^2 D_{u_1}^r D_{u_2}^s g(u_1, u_2) \left(b_1 u_1 + b_2 u_2\right) \\ &+ \sum_{r+s=2}^{3} \tfrac{(-1)^{r+s}}{r!s!} K_{rs1} D_{u_1}^r D_{u_2}^s g(u_1, u_2). \end{aligned} \tag{A.1.39}$$

Now, let us compute $D_{u_1}^r D_{u_2}^s g(u_1, u_2) \left(b_1 u_1 + b_2 u_2\right)$ via the chain rule:

$$\begin{aligned} D_{u_1}^r D_{u_2}^s g(u_1, u_2) \left(b_1 u_1 + b_2 u_2\right) &= \left(b_1 u_1 + b_2 u_2\right) D_{u_1}^r D_{u_2}^s g(u_1, u_2) \\ &+ g(u_1, u_2) D_{u_1}^r D_{u_2}^s \left(b_1 u_1 + b_2 u_2\right). \end{aligned} \tag{A.1.40}$$

Observe that $D_{u_1}^r D_{u_2}^s (b_1 u_1 + b_2 u_2) = 0$ for $r + s \geq 2$. Using the formula for bivariate Hermite polynomials given in Equation (9), we obtain

$$\begin{aligned} D_{u_1}^r D_{u_2}^s g(u_1, u_2) \left(b_1 u_1 + b_2 u_2\right) &= \left(b_1 u_1 + b_2 u_2\right) D_{u_1}^r D_{u_2}^s g(u_1, u_2) \\ &= \left(b_1 u_1 + b_2 u_2\right) (-1)^{r+s} H_{rs}(u_1, u_2) g(u_1, u_2) \\ &\quad \text{for } r + s \geq 2. \end{aligned} \tag{A.1.41}$$

Incorporating Equation $(A.1.41)$ into Equation $(A.1.39)$ and doing some manipulations, we get

$$\begin{aligned} E(\tilde{u}_3 \mid u_1, u_2) &= a^2 \left(b_1 u_1 + b_2 u_2\right) \left[1 + \sum_{r+s=3}^{4} \tfrac{1}{r!s!} K_{rs0} H_{rs}(u_1, u_2)\right] \\ &+ \sum_{r+s=2}^{3} \tfrac{1}{r!s!} K_{rs1} H_{rs}(u_1, u_2). \end{aligned} \tag{A.1.42}$$

Note that

$$1 + \sum_{r+s=3}^{4} \frac{1}{r!s!} K_{rs0} H_{rs}(u_1, u_2) = 1. \tag{A.1.43}$$

since $(u_1, u_2) \sim SBVN(\rho_{12})$, and the third order cumulants of any multivariate normal distribution and the fourth order cumulants of any bivariate normal distribution (corresponds to K_{rso} for $r+s=4$) are equal to 0 (Stuart and Ord (1994)). Then, Equation $(A.1.42)$ becomes

$$\begin{aligned} E(\tilde{u}_3 \mid u_1, u_2) &= a^2(b_1u_1 + b_2u_2) + \tfrac{K_{201}}{2}H_{20}(u_1, u_2) + K_{111}H_{11}(u_1, u_2) \\ &+ \tfrac{K_{021}}{2}H_{02}(u_1, u_2) + \tfrac{K_{301}}{6}H_{30}(u_1, u_2) + \tfrac{K_{211}}{2}H_{21}(u_1, u_2) \\ &+ \tfrac{K_{121}}{2}H_{12}(u_1, u_2) + \tfrac{K_{031}}{6}H_{21}(u_1, u_2). \end{aligned} \tag{A.1.44}$$

When we express a, b_1 and b_2 as in Equations $(A.1.2)$, $(A.1.3)$ and $(A.1.4)$ respectively, we get Equation (12). Using Equations $(A.2.1)$ and $(A.2.2)$ provided in the Section $A.2$, and the cumulant formulas $K_{101} = \rho_{13}$ and $K_{011} = \rho_{23}$ (Stuart and Ord (1994)), it can be shown that

$$a^2\left(b_1u_1 + b_2u_2\right) = K_{101}H_{10}(u_1, u_2) + K_{011}H_{01}(u_1, u_2), \tag{A.1.45}$$

Combining Equation $(A.1.45)$ with Equation $(A.1.44)$, we obtain Equation (11).

A.2 Derivation of $Var(\tilde{u}_3 \mid u_1,\ u_2)$

Since $Var\left(\tilde{u}_3 \mid u_1,\ u_2\right) = E\left(\tilde{u}_3^2 \mid u_1,\ u_2\right) - E^2\left(\tilde{u}_3 \mid u_1,\ u_2\right)$ and we already derived the analytical solution of $E^2\left(\tilde{u}_3 \mid u_1,\ u_2\right)$, we will initially derive $E\left(\tilde{u}_3^2 \mid u_1,\ u_2\right)$ in this subsection. Since $H_2(u) = u^2 = 1$, we can write

$$E\left(\tilde{u}_3^2 \mid u_1,\ u_2\right) = A^2a^2 + E\left(\tilde{u}_3^2 - A^2a^2 \mid u_1,\ u_2\right) = A^2a^2 + A^2a^2E\left(\left(\frac{\tilde{u}_3}{Aa}\right)^2 - 1 \mid u_1,\ u_2\right) = A^2a^2 + A^2a^2E\left(H_2\left(\frac{\tilde{u}_3}{Aa}\right.\right. \tag{A.2.1}$$

Now, let us focus on the second term $E\left(H_2\left(\frac{\tilde{u}_3}{Aa}\right) \mid u_1,\ u_2\right)$. Following the steps in Appendix A.1, we can write

$$E\left(H_2\left(\frac{\tilde{u}_3}{Aa}\right) \mid u_1, u_2\right)g(u_1, u_2) = I_1 + \sum_{r+s+p=3}^{4} \frac{(-1)^{r+s+p}}{r!s!p!}K_{rsp}I_{rsp}. \tag{A.2.2}$$

where

$$I_1 = \int_{-\infty}^{\infty} H_2\left(\frac{u_3}{Aa}\right)g(u_1, u_2, u_3)du_3, \tag{A.2.3}$$

and

$$I_{rsp} = \int_{-\infty}^{\infty} H_2\left(\frac{u_3}{Aa}\right)D_{u_1}^{r}D_{u_2}^{s}D_{u_3}^{p}g(u_1, u_2, u_3)du_3. \tag{A.2.4}$$

Following the same steps used to obtain Equation $(A.1.26)$, we can rewrite Equation $(A.2.4)$ as

$$I_{rsp} = \frac{1}{A^p a^{p-1}} \int_{-\infty}^{\infty} H_2(u_3 + c) D_{u_1}^r D_{u_2}^s D_{u_3}^p \phi(u_1) \phi(u_2^*) \phi(u_3) du_3. \tag{A.2.4}$$

Observe that $H_2(u_3+c) = (u_3+c)^2 - 1 = u_3^2 - 1 + 2cu_3 + c^2 = H_2(u_3) + 2cH_1(u_3) + c^2 H_0(u_3)$ since $H_1(u_3) = u_3$ and $H_0(u_3) = 1$. Hence, we can write

$$H_2(u_3 + c) = \sum_{i=0}^{2} \begin{pmatrix} 2 \\ i \end{pmatrix} c^i H_{2-i}(u_3). \tag{A.2.5}$$

Incorporating Equation $(A.2.5)$ into Equation $(A.2.4)$, we get

$$I_{rsp} = \frac{(-1)^p}{A^p a^{p-1}} D_{u_1}^r D_{u_2}^s \phi(u_1) \phi(u_2^*) \sum_{i=0}^{1} \begin{pmatrix} 2 \\ i \end{pmatrix} c^i \int_{-\infty}^{\infty} H_p(u_3) H_{2-i}(u_3) \phi(u_3) du_3. \tag{A.2.6}$$

Exploiting the orthogonality property of Hermite polynomails provided in Equation $(A.1.29)$, we can conclude that

$$\int_{-\infty}^{\infty} H_p(u_3) H_{2-i}(u_3) \phi(u_3) du_3 = p! \text{ for } p \leq 2 \;\&\; p = 2 - i. \tag{A.2.7}$$

Also observe that $\int_{-\infty}^{\infty} H_p(u_3) H_{2-i}(u_3) \phi(u_3) du_3 = 0$ for $p > 2$, $I_{rsp} = 0$ for $p > 2$. Substituting Equation $(A.2.7)$ into Equation $(A.2.6)$, we get

$$I_{rsp} = \frac{(-1)^p}{A^p a^{p-1}} D_{u_1}^r D_{u_2}^s \phi(u_1) \phi(u_2^*) \begin{pmatrix} 2 \\ 2-p \end{pmatrix} c^{2-p} p! \text{ for } p \leq 2. \tag{A.2.8}$$

Since $I_1 = I_{000}$ and substituting $c = \frac{a}{A}(b_1 u_1 + b_2 u_2)$, we can write

$$I_1 = \frac{a^3}{A^2} \phi(u_1) \phi(u_2^*) (b_1 u_1 + b_2 u_2)^2. \tag{A.2.9}$$

Setting $p = 0$, $p = 1$ and $p = 2$ in Equation $(A.2.8)$, we respectively get:

$$I_{rs0} = \frac{a^3}{A^2} D_{u_1}^r D_{u_2}^s \phi(u_1) \phi(u_2^*) (b_1 u_1 + b_2 u_2)^2, \tag{A.2.10}$$

and

$$I_{rs1} = \frac{-2a}{A^2} D_{u_1}^r D_{u_2}^s \phi(u_1) \phi(u_2^*) (b_1 u_1 + b_2 u_2), \tag{A.2.11}$$

and

$$I_{rs2} = \frac{2}{A^2 a} D_{u_1}^r D_{u_2}^s \phi(u_1) \phi(u_2^*). \tag{A.2.12}$$

Substituting Equations$(A.2.9) - (A.2.12)$ into Equation $(A.2.1)$, we get

$$
\begin{aligned}
E\left(\tilde{u}_3^2 \mid u_1,\ u_2\right) g(u_1,u_2) &= A^2a^2g(u_1,u_2) + a^5\phi(u_1)\phi\left(u_2^*\right)\left(b_1u_1+b_2u_2\right)^2 \\
&+ \sum_{r+s=3}^{4} \tfrac{(-1)^{r+s}}{r!s!} K_{rs0}a^5 D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right)\left(b_1u_1+b_2u_2\right)^2 \\
&- 2\sum_{r+s=2}^{3} \tfrac{(-1)^{r+s}}{r!s!} K_{rs1}a^3 D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right)\left(b_1u_1+b_2u_2\right) \\
&+ 2\sum_{r+s=1}^{2} \tfrac{(-1)^{r+s}}{r!s!} K_{rs2}a D_{u_1}^r D_{u_2}^s \phi(u_1)\phi\left(u_2^*\right).
\end{aligned} \tag{A.2.13}
$$

Since the third and fourth order cumulants of any bivariate normal distribution (corresponds to K_{rso} for $r+s=3,4$) are equal to 0 and $a\phi(u_1)\phi(u_2^*) = g(u_1,u_2)$, Equation $(A.2.13)$ can be written as:

$$
\begin{aligned}
E\left(\tilde{u}_3^2 \mid u_1,\ u_2\right) g(u_1,u_2) &= A^2a^2g(u_1,u_2) + a^4g(u_1,u_2)\left(b_1u_1+b_2u_2\right)^2 \\
&- 2\sum_{r+s=2}^{3} \tfrac{(-1)^{r+s}}{r!s!} K_{rs1}a^2 D_{u_1}^r D_{u_2}^s g(u_1,u_2)\left(b_1u_1+b_2u_2\right) \\
&+ 2\sum_{r+s=1}^{2} \tfrac{(-1)^{r+s}}{r!s!} K_{rs2} D_{u_1}^r D_{u_2}^s g(u_1,u_2).
\end{aligned} \tag{A.2.14}
$$

Using Equation $(A.1.41)$ and doing some manipulations, we get

$$
\begin{aligned}
E\left(\tilde{u}_3^2 \mid u_1,\ u_2\right) &= A^2a^2 + a^4\left(b_1u_1+b_2u_2\right)^2 - 2\left(b_1u_1+b_2u_2\right)\sum_{r+s=2}^{3} \tfrac{1}{r!s!} K_{rs1}H_{rs}(u_1,u_2) \\
&+ 2\sum_{r+s=1}^{2} \tfrac{1}{r!s!} K_{rs2}H_{rs}(u_1,u_2).
\end{aligned} \tag{A.2.15}
$$

A.3 Expressions for Bivariate Hermite Polynomials

The expressions given below were obtained by solving Equation (9) for various values of r and s using Maple.

$$H_{10}(u_1,u_2) = au_1^*, \tag{A.3.1}$$

$$H_{01}(u_1,u_2) = au_2^*, \tag{A.3.2}$$

$$H_{20}(u_1,u_2) = -a^2 + a^2u_1^{*2}, \tag{A.3.3}$$

$$H_{02}(u_1,u_2) = -a^2 + a^2u_2^{*2}, \tag{A.3.4}$$

$$H_{11}(u_1,u_2) = -a^2\rho_{12} - a^2u_1^*u_2^*, \tag{A.3.5}$$

$$H_{30}(u_1,u_2) = au_1^*\left(H_{20} - 2a^2\right), \tag{A.3.6}$$

$$H_{03}(u_1,u_2) = au_2^*\left(H_{02} - 2a^2\right), \tag{A.3.7}$$

$$H_{21}(u_1, u_2) = aH_{11}u_1^* - a^3(u_2^* - \rho_{12}u_1^*), \tag{A.3.8}$$

$$H_{12}(u_1, u_2) = aH_{11}u_2^* - a^3(u_1^* - \rho_{12}u_2^*). \tag{A.3.9}$$

A.4 Moments of the Truncated *SBVN* Distribution

The moment formulas of the truncated *SBVN* distribution up to the second order may be found in Rosenbaum (1961). However, in our independent derivation, we discovered that Rosenbaum (1961) made a sign error while calculating integrals. Our expressions up to the second order agree with those in Henning and Henningsen (2007), who cross-check their formulas using numerical integration and Monte Carlo simulation. We also provide some third order moments of the truncated *SBVN* distribution below. The derivations may be obtained from the authors upon request. We denote the univariate standard normal density and distribution function respectively by $\phi(.)$ and $\Phi(.)$. Let $C_{ti} = \beta_{ti}X_{ti}$ for $t = 1, 2$ for individual i. We suppress the individual subscript throughout the section to avoid notational clutter. We denote the moments of the truncated *SBVN* distribution for the subsample S_4 (defined in Section 2.1) as $m_{j,k} = E(u_1^j u_2^k \mid u_1 > -C_1, u_2 > -C_2)$. Let

$$q = \frac{\sqrt{1-\rho_{12}^2}}{\sqrt{2\pi}} = \frac{1}{a\sqrt{2\pi}}, \tag{A.4.1}$$

$$q' = a^2 q, \tag{A.4.2}$$

$$D(C_1, C_2) \equiv D^2 = C_1^2 - 2\rho_{12}C_1C_2 + C_2^2, \tag{A.4.3}$$

$$C_1^*(C_1, C_2) \equiv C_1^* = a(C_1 - \rho_{12}C_2), \tag{A.4.4}$$

$$C_2^*(C_1, C_2) \equiv C_2^* = a(C_2 - \rho_{12}C_1), \tag{A.4.5}$$

$$C_1^{**}(C_1, -C_2) \equiv C_1^{**} = C_1 + \rho_{12}C_2, \tag{A.4.6}$$

$$C_2^{**}(-C_1, C_2) \equiv C_2^{**} = C_2 + \rho_{12}C_1, \tag{A.4.7}$$

$$\alpha(C_1, C_2) = \phi(-C_1)\Phi(C_2^*), \tag{A.4.8}$$

$$\beta(C_1, C_2) = \phi(-C_2)\Phi(C_1^*), \tag{A.4.9}$$

$$\delta(C_1, C_2) = \phi\left(aD\right), \tag{A.4.10}$$

and

$$L(C_1, C_2; \rho_{12}) = \int_{-C_2}^{\infty}\int_{-C_1}^{\infty} g(u_1, u_2)du_1du_2. \tag{A.4.11}$$

Then,

$$m_{0,1} = \frac{\rho_{12}\alpha(C_1, C_2) + \beta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.4.12}$$

$$m_{1,0} = \frac{\alpha(C_1, C_2) + \rho_{12}\beta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.4.13}$$

$$m_{0,2} = \frac{L(C_1, C_2; \rho_{12}) - \rho_{12}^2 C_1\alpha(C_1, C_2) - C_2\beta(C_1, C_2) + q\rho_{12}\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.4.14}$$

$$m_{2,0} = \frac{L(C_1, C_2; \rho_{12}) - C_1\alpha(C_1, C_2) - \rho_{12}^2 C_2\beta(C_1, C_2) + q\rho_{12}\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.4.15}$$

$$m_{1,1} = \frac{\rho_{12}L(C_1, C_2; \rho_{12}) - \rho_{12}C_1\alpha(C_1, C_2) - \rho_{12}C_2\beta(C_1, C_2) + q\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.4.16}$$

$$\begin{aligned} m_{0,3} &= \tfrac{1}{L(C_1,C_2;\rho_{12})}[2L(C_1, C_2; \rho_{12})m_{0,1} + \left(\rho_{12}\left(1-\rho_{12}^2\right) + \rho_{12}^3 C_1^2\right)\alpha(C_1, C_2) \\ &+ C_2^2\beta(C_1, C_2) - q\rho_{12}C_2^{**}\delta(C_1, C_2)], \end{aligned} \tag{A.4.17}$$

$$\begin{aligned} m_{3,0} &= \tfrac{1}{L(C_1,C_2;\rho_{12})}[2L(C_1, C_2; \rho_{12})m_{1,0} + \left(\rho_{12}\left(1-\rho_{12}^2\right) + \rho_{12}^3 C_2^2\right)\beta(C_1, C_2) \\ &+ C_1^2\alpha(C_1, C_2) - q\rho_{12}C_1^{**}\delta(C_1, C_2)], \end{aligned} \tag{A.4.18}$$

$$\begin{aligned} m_{1,2} &= \tfrac{1}{L(C_1,C_2;\rho_{12})}[2\rho_{12}L(C_1, C_2; \rho_{12})m_{0,1} + \rho_{12}C_2^2\beta(C_1, C_2) \\ &+ (1-\rho_{12}^2 + \rho_{12}^2 C_1^2)\alpha(C_1, C_2) - qC_2^{**}\delta(C_1, C_2)], \end{aligned} \tag{A.4.19}$$

$$\begin{aligned} m_{2,1} &= \tfrac{1}{L(C_1,C_2;\rho_{12})}[2\rho_{12}L(C_1, C_2; \rho_{12})m_{1,0} + \rho_{12}C_1^2\alpha(C_1, C_2) \\ &+ (1-\rho_{12}^2 + \rho_{12}^2 C_2^2)\beta(C_1, C_2) - qC_1^{**}\delta(C_1, C_2)]. \end{aligned} \tag{A.4.20}$$

A.5 Expressions of λ's

We obtain formulas below using Equation (12) and the formulas provided in Appendices A.3 and A.4 via a Maple code.

$$\lambda_1 = \frac{\alpha(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.1}$$

$$\lambda_2 = \frac{\beta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.2}$$

$$\lambda_3 = \frac{-C_1\alpha(C_1, C_2) - q\rho_{12}\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.3}$$

$$\lambda_4 = \frac{-C_2\beta(C_1, C_2) - q\rho_{12}\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.4}$$

$$\lambda_5 = -\frac{q\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.5}$$

$$\lambda_6 = \frac{-\alpha(C_1, C_2) + C_1^2\alpha(C_1, C_2) + q\rho_{12}\left(aC_1^* + C_1\right)\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.6}$$

$$\lambda_7 = \frac{-\beta(C_1, C_2) + C_2^2\beta(C_1, C_2) + q\rho_{12}\left(aC_2^* + C_2\right)\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.7}$$

$$\lambda_8 = -\frac{aqC_1^*\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}, \tag{A.5.8}$$

$$\lambda_9 = -\frac{aqC_2^*\delta(C_1, C_2)}{L(C_1, C_2; \rho_{12})}. \tag{A.5.9}$$

A.6 Expressions of λ's for Tunali and Baslevent's (2006) Problem

Let $\sigma_{12} = \sigma_1\sigma_2\rho_{12}$, $\sigma_{11} = \sigma_1^2$ and $\sigma_{22} = \sigma_2^2$. When $(u_1, u_2) \sim SBVN(\rho_{12})$, $(\sigma_1 u_1, \sigma_1 u_1 + \sigma_2 u_2) \sim BVN(0, 0, \sigma_{11}, a^2, e)$ where $a = \sqrt{\sigma_{11} + \sigma_{22} + 2\sigma_{12}}$ and $e = \sigma_{11} + \sigma_{12}$. Let $b = \sigma_{22} + \sigma_{12}$. Denote the probability of $LFP = j$ by P_j and the standard bivariate normal cumulative distribution function with upper thresholds s_1, s_2 and correlation coefficient r by $\Phi(s_1, s_2; r)$. Assume that state probabilities are given as:

$$P_0 = \Phi(C_{01}, C_{02}; C_{03}), \tag{A.6.1}$$

$$P_1 = \Phi(C_{11}, C_{12}; C_{13}), \tag{A.6.2}$$

$$P_2 = \Phi(C_{21}, C_{22}; C_{23}), \tag{A.6.3}$$

$$P_3 = 1 - P_0. \tag{A.6.4}$$

Unemployment probability follows from the adopted definition. Considering the categorical variable LFP in Equation (20), observe that

$$\begin{aligned} P_0 = P(y_1^* < 0,\ y_1^* + y_2^* < 0) &= P\left(\sigma_1 u_1 < -\beta_1' z, \sigma_1 u_1 + \sigma_2 u_2 < -\left(\beta_1' + \beta_2'\right) z\right) \\ &= P\left(u_1 < \frac{-\beta_1' z}{\sigma_1}, \frac{\sigma_1 u_1 + \sigma_2 u_2}{a} < \frac{-\left(\beta_1' + \beta_2'\right) z}{a}\right). \end{aligned} \tag{A.6.5}$$

$$\begin{aligned} P_1 = P(y_1^* > 0,\ y_2^* < 0) &= P\left(\sigma_1 u_1 > -\beta_1' z, \sigma_2 u_2 < -\beta_2' z\right) \\ &= P\left(u_1 > \frac{-\beta_1' z}{\sigma_1}, u_2 < \frac{-\beta_2' z}{\sigma_2}\right) \\ &= P\left(u_1 < \frac{\beta_1' z}{\sigma_1}, u_2 < \frac{-\beta_2' z}{\sigma_2}\right). \end{aligned} \tag{A.6.6}$$

$$\begin{aligned} P_2 = P(y_2^* > 0, y_1^* + y_2^* > 0) &= P\left(\sigma_2 u_2 > -\beta_2' z, \sigma_1 u_1 + \sigma_2 u_2 > -\left(\beta_1' + \beta_2'\right) z\right) \\ &= P\left(u_2 > \frac{-\beta_2' z}{\sigma_2}, \frac{\sigma_1 u_1 + \sigma_2 u_2}{a} > \frac{-\left(\beta_1' + \beta_2'\right) z}{a}\right) \\ &= P\left(u_2 < \frac{\beta_2' z}{\sigma_2}, \frac{\sigma_1 u_1 + \sigma_2 u_2}{a} < \frac{\left(\beta_1' + \beta_2'\right) z}{a}\right). \end{aligned} \tag{A.6.7}$$

In Equations $(A.6.5)$ through $(A.65.7)$, we rescaled the variables to have unit variances by dividing them to their standard errors. Observe that for $LFP = 0$, correlation will be $\frac{Cov(\sigma_1 u_1, \sigma_1 u_1 + \sigma_2 u_2)}{\sqrt{Var(\sigma_1 u_1)}\sqrt{Var(\sigma_1 u_1 + \sigma_2 u_2)}} = \frac{e}{a\sigma_1}$; for $LFP = 1$, it will be $-\frac{Cov(\sigma_1 u_1, \sigma_2 u_2)}{\sqrt{Var(\sigma_1 u_1)}\sqrt{Var(\sigma_2 u_2)}} = -\frac{\sigma_{12}}{\sigma_1 \sigma_2}$; and for $LFP = 2$, it will be $\frac{Cov(\sigma_2 u_2, \sigma_1 u_1 + \sigma_2 u_2)}{\sqrt{Var(\sigma_2 u_2)}\sqrt{Var(\sigma_1 u_1 + \sigma_2 u_2)}} = \frac{b}{a\sigma_2}$. Hence, we can write

$$C_{01} = \frac{-\beta_1' z}{\sigma_1},\ C_{02} = \frac{-\left(\beta_1' + \beta_2'\right) z}{a},\ C_{03} = \frac{e}{a\sigma_1}, \tag{A.5.8}$$

$$C_{11} = \frac{\beta_1' z}{\sigma_1},\ C_{12} = \frac{-\beta_2' z}{\sigma_2},\ C_{13} = -\frac{\sigma_{12}}{\sigma_1 \sigma_2}, \tag{A.5.9}$$

$$C_{21} = \frac{\beta_2' z}{\sigma_2},\ C_{22} = \frac{\left(\beta_1' + \beta_2'\right) z}{a},\ C_{23} = \frac{b}{a\sigma_2}. \tag{A.5.10}$$

Changing all C_1's to C_{21},all C_2's to C_{22}, and all ρ_{12}'s to C_{23} in the formulas provided in Appendix

$A.4$, we obtain the expressions for λ's in context of Tunali and Baslevent's (2006) problem.

References

Ahn, H. and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *The Annals of Statistics*, 6, 434–451.

Bhattacharya, R. N. and R. R. Rao (1976): *Normal Approximation and Asymptotic Expansions*, John Wiley and Sons Inc., New York.

Bourguignon, F., M. Fournier, and M. Gurgand (2007): "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons," *Journal of Economic Surveys*, 21, 174–205.

Chambers, J. M. (1967): "On Methods of Asymptotic Approximation for Multivariate Distributions," *Biometrika*, 54, 367–383.

Cramer, H. (1937): *Random Variables and Probability Distributions*, Cambridge University Press, Cambridge, UK.

Dahl, G. B. (2002): "Mobility and the Returns to Education: Testing A Roy Model With Multiple Markets," *Econometrica*, 70, 2367–2420.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58.

David, S. T., M. G. Kendall, and A. Stuart (1951): "Some Questions of Distribution in the Theory of Rank Correlation," *Biometrika*, 38, 131–140.

Davis, P. J. and I. Polonsky (1964): "Numerical Interpolation, Differentiation, and Integration," in *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, ed. by M. Abramowitz and I. A. Stegun, National Bureau of Standards.

Dubin, J. A. and D. L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

EDGEWORTH, F. Y. (1905): "The Law of Error," *Transcations of the Cambridge Philosophical Society*, 20, 36–65 and 113–141.

ELDERTON, W. P. (1906): "Notices," *Biometrika*, 5, 206–210.

ERIKSSON, A., L. FORSBERG, AND E. GHYSELS (2004): "Approximating the Probability Distribution of Functions of Random Variables: A New Approach," CIRANO Working Papers 2004s-21, CIRANO.

GOLDBERGER, A. S. (1991): *A Course in Econometrics*, Harvard University Press, Cambridge, MA.

HECKMAN, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selectipn and Limited Dependent Variables and a Simpler Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475–492.

——— (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

HENNING, C. H. AND A. HENNINGSEN (2007): "Modeling Farm Households' Price Responses in the Presence of Transaction Costs and Heterogeneity in Labor Markets," *American Journal of Agricultural Economics*, 89, 665–681.

KEANE, M. P. (1992): "A Note on Identification in the Multinomial Probit Model," *Journal of Business and Economic Statistics*, 10, 193–200.

KOLASSA, J. E. (1997): *Series Approximation Methods in Statistics*, Springer-Verlag, Boston, 2nd ed.

KOLASSA, J. E. AND P. MCCULLAGH (1990): "Edgeworth Series for Lattice Distributions," *The Annals of Statistics*, 18, 981–985.

LAHIRI, K. AND J. G. SONG (1999): "Testing for Normality in a Probit Model with Double Selection," *Economics Letters*, 65, 33–39.

LEE, L. F. (1976): "Estimation of Limited Dependent Variable Models by Two Stage Methods," Ph.D. thesis, Department of Economics, University of Rochester.

——— (1979): "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica*, 47, 977–996.

——— (1982): "Some Approaches to the Correction of Selectivity Bias," *Review of Economic Studies*, 49, 355–372.

——— (1983): "Generalized Econometric Models with Selectivity," *Econometrica*, 51, 505–512.

——— (1984): "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity," *Econometrica*, 52, 843–863.

Leung, S. F. and S. Yu (1996): "On the Choice Between Sample Selection and Two-part Models," *Journal of Econometrics*, 72, 197–229.

Little, R. J. A. (1985): "A Note About Models of Selectivity Bias," *Econometrica*, 53, 1469–1474.

Magnac, T. (1991): "Segmented or Competitive Labor Markets," *Econometrica*, 59, 165–87.

Mardia, K. V. (1970): "Families of Bivariate Distributions," Harner Publishing Company, Darien, CONN.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, Academic Press, New York.

Miller, J. (2011): "Earliest Known Uses of Some of the Words of Mathematics," `http://jeff560.tripod.com/e.html`.

Olsen, R. J. (1980): "A Least Squares Correction for Selectivity Bias," *Econometrica*, 48, 1815–20.

Pretorius, S. J. (1930): "Skew Bivariate Frequency Surfaces, Examined in the Light of Numerical Illustrations," *Biometrika*, 22, 109–223.

Rosenbaum, S. (1961): "Moments of a Truncated Bivariate Normal Distribution," *Journal of the Royal Statistical Society. Series B (Methodological)*, 23, 405–408.

Rothenberg, T. J. (1983): "Asymptotic Properties of Some Estimators in Structural Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, ed. by S. Karlin, T. Amemiya, and L. Goodman, Academic Press, New York.

Ruud, P. A. (1983): "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models," *Econometrica*, 51, 225–228.

Schmertmann, C. P. (1994): "Selectivity Bias Correction Methods in Polychotomous Sample Selection Models," *Journal of Econometrics*, 60, 101–132.

Stoker, T. M. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461–1481.

Stuart, A. and J. K. Ord (1994): *Kendall's Advanced Theory of Statistics: Volume 1: Distribution Theory*, Edward Arnold. London, UK, 6th ed.

Tunali, I. (1986): "A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Re-migration," *Research in Labor Economics*, 8B, 235–282.

Tunali, I. and C. Baslevent (2006): "Female Labor Supply in Turkey," in *The Turkish Economy: The Real Economy, Corporate Governance and Reform and Stabilization Policy*, ed. by S. Altug and A. Filiztekin, Routledge, London, UK.

Vella, F. (1998): "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33, 127–169.