Estimating Heterogeneous Effects in Static Binary Response Panel Data Models

Anastasia Semykina
Department of Economics
Florida State University
Tallahassee, FL 32306-2180, USA
E-mail: asemykina@fsu.edu
Phone: +1 850-644-4557
Fax: +1 850-644-4535

Abstract

This paper considers estimating heterogeneous effects in panel data models when the outcome is binary. We argue that a common practice of splitting the sample and performing estimation separately for each subsample results in inconsistent estimators of heterogeneous parameters. The paper presents methods that account for a possibility of nonrandom sorting and produce consistent estimators of causal effects in two or more heterogeneous sub-populations. Monte Carlo simulations show that considered methods perform well in finite samples. As an empirical application, the paper studies gender differences in job satisfaction by occupation type.

JEL Classifications: C33, C34, C35

Keywords: binary response, heterogeneous effects, nonrandom sorting

The author thanks the participants of the 2019 Latin American Meeting of the Econometric Society, 5th Annual Conference of the International Association for Applied Econometrics, 2018 Midwest Econometrics Group Meeting, and 2017 Southern Economic Association Meeting for helpful comments.

1 Introduction

Researchers are frequently interested in estimating heterogeneous effects on binary outcomes in different population subgroups. For example, one may want to study gender differences in job satisfaction by occupation. The existing literature finds that women on average are as satisfied or more satisfied with their jobs than men, even though women tend to have a lower pay and are generally disadvantaged in the labor market (Clark, 1997; Sloane and Williams, 2000, among others). This phenomenon is known as "the paradox of the contented female worker" (Crosby, 1982). An important research question is whether this paradox manifests itself differently in different types of jobs. Other examples include studying the determinants of high-school dropouts by race, gender, and socio-economic status, and investigating self-employment outcomes by age and education level. Estimating heterogeneous treatment effects is also often of interest.

In the applied literature, it is common to estimate group-specific parameters by dividing the sample into corresponding subsamples and performing the estimation separately for each group. While this approach is intuitively appealing, it generally results in inconsistent estimators when sorting into groups is not random. Similar to linear models (Vella, 1988), consistent estimators of heterogeneous parameters can only be obtained if the full information set is utilized, i.e. when each group is considered as part of the entire population. The estimation is further complicated when considering panel data models, which are characterized by unobserved unit-specific hererogeneity and cross-group transitions over time. The present paper discusses methods that address nonrandom sorting and produce consistent estimators of heterogeneous parameters and partial effects in static binary response panel data models.

The related literature goes back to the studies of linear switching regression models (Goldfeld and Quandt, 1973; Lee 1978; Maddala and Nelson, 1975; Maddala 1983). Such models specify two equations, where the applicability of either equation depends on the endogenous switching from one regime to the other. Another relevant strand of the literature includes studies of program evaluation and estimation of treatment effects. Analogous to switching regression models, program evaluation studies focus on addressing endogenous self-selection into treatment. One parameter of interest is the effect of treatment on the treated, which can be formulated within either a switching regression or self-selection framework (Bjorklund and Moffitt, 1987; Heckman et al., 2006). Furthermore, several studies have proposed methods for estimating heterogeneous treatment effects in linear models using the instrumental variables methodology (Heckman et al., 2006; Basu, 2014, among others).

The problem of nonrandom selection is discussed in studies of sample selection, including the seminal paper by Heckman (1979). Although such models consider homogeneous parameters, the selection problem arises because the dependent variable is not observed for some part of the population. The existing literature discusses methods for addressing sample selection in linear and binary response models; both cross section and panel data models have been considered (Heckman, 1979; Kyriazidou, 1997; Newey, 2009; Semykina and Wooldridge, 2018; Wooldridge, 1995, among others).

Regarding heterogeneous effects in binary response models, several studies discuss a switching probit model, where the endogenous switching is between two regimes, and parameters are regime-specific. For example, Manski et al. (1992) study the cross section case, while Carrasco (2001) proposes an estimator for dynamic panel data models. To the best of our knowledge, estimating heterogeneous effects in models with arbitrary number of groups has not been considered so far. The present paper proposes methods for estimating heterogeneous effects in static binary response panel data models with two or more groups. The correlated random effects approach is used to account for the presence of unobserved unit-specific heterogeneity that may be correlated with explanatory variables. An arbitrary number of ordered or unordered groups is permitted.

The rest of the paper is structured as follows. Section 2 presents binary response models with heterogeneous effects. Estimation of population parameters and partial effects is discussed in Section 3. Section 4 presents asymptotic theory; simulation results are discussed in Section 5. Section 6 contains an empirical application, which estimates gender differences in job satisfaction by occupation type. Section 7 concludes.

2 Heterogeneity in binary response panel data models

2.1 General Setup

Let the population consist of J fixed groups (or subpopulations). Assume that the number of periods, T, is fixed, and $N \to \infty$, where N is the cross section sample size. This paper considers the following binary response model with heterogeneous effects:

$$\begin{aligned} y_{itj}^* &= \mathbf{x}_{it}\boldsymbol{\beta}_j + c_{ij} + u_{itj}, \qquad (1)\\ y_{itj} &= 1[y_{itj}^* > 0], \quad t = 1, \ldots, T, \quad j = 1, \ldots, J, \end{aligned}$$

where y_{itj}^* is a continuous latent variable, y_{itj} is the observed binary outcome for unit i in group j in period t, and $1[\cdot]$ is an indicator function equal to one if the expression in brackets is true. The vector of explanatory variables, $\mathbf{x}_{it}$, is $1 \times K$, and $\boldsymbol{\beta}_j$ is a $K \times 1$ group-specific vector of parameters. Define $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ and make the following assumption:

Assumption 1 $u_{itj} \perp\!\!\!\perp (x_i, c_{ij})$

The assumption states that $\mathbf{x}_{it}$ is independent of the idiosyncratic error, but allows $\mathbf{x}_{it}$ to be correlated with a time-constant group-specific unobserved effect c_{ij}. It also indicates that the observed covariates are strictly exogenous conditional on c_{ij}, i.e. past and future values of $\mathbf{x}_{it}$ do not affect the distribution of y_{it} after accounting for the current values of covariates and the unobserved effect.

Note that a given cross section unit may appear in different groups in different t. Transitions may occur due to changes in both time-varying covariates and idiosyncratic shocks and may be endogenous with respect to y_{itj}. For example, a shock to the main outcome may affect both the probability of success in group j and the probability of belonging to group j in period t. Even if j is constant across t, group sorting is not random if time-constant and/or time-varying unobservables affecting the group assignment are correlated with the unobservables in (1). As discussed in detail below, such endogeneity causes inconsistency in the estimators of $\boldsymbol{\beta_j}$ obtained by estimating (1) separately for each j.

Let d_{it} be a discrete random variable identifying groups, $d_{it} = \{1, 2, \ldots, J\}$. After defining dichotomous indicators for each group as $s_{itj} = 1[d_{it} = j]$, $t = 1, \ldots,$ T, $j = 1, \ldots, J$, the outcome for unit i in a given period can be written as

$$y_{it} = \sum_{j=1}^{J} s_{itj} y_{itj}, \qquad t = 1, \ldots, T. \tag{2}$$

Apart from $\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, parameters of interest include partial effects. These can be of two types. We define the unconditional partial effect (PE_j^U) as a change in the probability of success in group j due to an increase in variable x for a randomly selected unit from the population. In the population, the unconditional partial effect of a continuous explanatory variable is

$$PE_{j,k}^U = \frac{\partial \mathrm{P}(y_{itj} = 1 | \mathbf{x}_{it}, c_{ij})}{\partial x_{itk}}, \qquad j = 1, \ldots, J. \tag{3}$$

On the other hand, we define a conditional partial effect (PE_j^C) as a change in the probability of success due to an increase in x for a unit in group j.[1] For a continuous covariate,

$$PE^C_{j,k} = \frac{\partial \mathrm{P}(y_{it} = 1|d_{it} = j, \mathbf{x}_{it}, c_{ij})}{\partial x_{itk}} = \frac{\partial \mathrm{P}(y_{itj} = 1|d_{it} = j, \mathbf{x}_{it}, c_{ij})}{\partial x_{itk}} \qquad j = 1, \ldots, J. \quad (4)$$

Although both effects may be of interest, PE^U_j is often deemed more suitable for cross-group comparisons, whereas PE^C_j is useful when focusing on a particular group. For example, a commonly used application of PE^C_j is the estimation of the average treatment effect on the treated in the policy evaluation literature.

In practice, c_{ij} is not observed, which makes it impossible to estimate $PE^U_{j,k}$ and $PE^C_{j,k}$. Instead, it is common to estimate average partial effects (APE) that are obtained by 'averaging' over the distribution of the unobserved effect, c_{ij}:

$$\begin{aligned} APE^U_{j,k} &= \mathrm{E}_{c_j}\left[\frac{\partial \mathrm{P}(y_{itj} = 1|\mathbf{x}_{it}, c_{ij})}{\partial x_{itk}}\right], && (5) \\ APE^C_{j,k} &= \mathrm{E}_{c_j}\left[\frac{\partial \mathrm{P}(y_{itj} = 1|d_{it} = j, \mathbf{x}_{it}, c_{ij})}{\partial x_{itk}}\right], \qquad j = 1, \ldots, J. \end{aligned}$$

If group assignment is random, then $\mathrm{P}(y_{itj} = 1|\mathbf{x}_{it}, c_{ij}) = \mathrm{P}(y_{itj} = 1|d_{it} = j, \mathbf{x}_{it}, c_{ij})$, and consistent estimators of model parameters are obtained by estimating (1) separately for each j. However, because of self-selection and other factors sorting into groups may be nonrandom, which causes inconsistency. In this paper, we allow for a possibility that $\mathrm{P}(y_{itj} = 1|d_{it} = j, \mathbf{x}_{it}, c_{ij}) \neq \mathrm{P}(y_{itj} = 1|\mathbf{x}_{it}, c_{ij})$ and discuss how it can be addressed when obtaining consistent estimators of $\boldsymbol{\beta}_j$ and APE. We start by considering a simple case with only two groups and then discuss more general models with $J > 2$, where groups may be ordered or unordered.

[1] Both PE^U_j and PE^C_j are changes in probabilities conditional on $(\mathbf{x}_{it}, c_{ij})$. Hence, the use of "unconditional" partial effects is not fully accurate. Here, we use this terminology for convenience, to distinguish between partial effects conditional on being in group j and partial effects when not conditioning on $d_{it} = j$.

2.2 Model for two groups

Let y_{itj} be determined as in equation (1), where $J = 2$. Applications of such models include, for example, examining labor force participation among married and non-married women, as well as estimating the determinants of dropout incidents among economically disadvantaged and other students. Assume that sorting into groups is determined by the value of a latent variable d_{it}^*,

$$\begin{aligned} d_{it}^* &= \mathbf{z}_{it}\boldsymbol{\delta} + b_i + v_{it}, \\ d_{it} &= 1 \text{ if } d_{it}^* \leq 0, \\ d_{it} &= 2 \text{ if } d_{it}^* > 0, \end{aligned} \tag{6}$$

where $\mathbf{z}_{it}$ is a $1 \times L$ vector of exogenous variables, b_i is a time-constant unobserved effect, and v_{it} is an idiosyncratic error. Setting the cut point at zero is at no cost, as long as z_{it} contains an intercept. Vector $\mathbf{z}_{it} = (\mathbf{x}_{it}, \mathbf{z}_{it1})$ contains at least one additional variable that is not in $\mathbf{x}_{it}$.[2] Similar to the main equation, define $\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT})$ and assume that the following holds:

Assumption 2 $v_{it} \perp\!\!\!\perp (\mathbf{z}_i, b_i)$.

Hence, $\mathbf{z}_{it}$ is strictly exogenous conditional on b_i, but may be correlated with b_i. This correlation causes an omitted variable problem that has to be resolved before addressing nonrandom sorting. Building upon the work by Mundlak (1978) and Chamberlain (1980), unobserved effects can be modeled as

$$\begin{aligned} c_{ij} &= \bar{\mathbf{z}}_i\boldsymbol{\psi}_{cj} + a_{cij}, \quad j = 1, 2, \\ b_i &= \bar{\mathbf{z}}_i\boldsymbol{\psi}_b + a_{bi}, \end{aligned} \tag{7}$$

[2]Strictly speaking, the exclusion restriction is not required for identification. However, when $\mathbf{z}_{it} = \mathbf{x}_{it}$, identification relies exclusively on the functional form of the likelihood function, which is less reliable.

where $\bar{\mathbf{z}}_i = \sum_{t=1}^{T} \mathbf{z}_{it}$, and $(a_{ci1}, a_{ci2}, a_{bi})$ are independent of $\mathbf{z}_i$. This modeling approach has been previously used in both theoretical and applied work (Abrevaya and Dahl, 2008; Jäckle and Himmler, 2010; Papke and Wooldridge, 2008; Semykina, 2018; Semykina and Wooldridge, 2010, 2018; Wooldridge, 1995, among others).[3] Note that although $\mathbf{z}_{it}$ may contain time-constant covariates, equation (7) indicates that their causal effects cannot be distinguished from the impact of c_{ij} and b_i, unless they are independent of the unobserved effects. Nevertheless, it is important to include such variables as controls to prevent inconsistency due to an omitted variable problem.

Note that modeling the unobserved effect as in (7) is different from the approach employed by Carrasco (2001). Carrasco (2001) considers a switching probit model for panel data, but accounts for unobserved heterogeneity by including lagged dependent variables in both main and sorting equations. That approach works well when estimating dynamic models and studying state dependence. In this paper, we focus on estimating static models, which are widely used in applied research, including the estimation of treatment effects. Hence, we model the unobserved effect as in (7).

Using (7), equations (1) and (6) can be written as

$$\begin{aligned} y_{itj} &= 1[\mathbf{x}_{it}\boldsymbol{\beta}_j + \bar{\mathbf{z}}_i\boldsymbol{\psi}_{cj} + \eta_{itj} > 0], \quad t = 1, \ldots, T, \quad j = 1, 2, \\ d_{it} &= 1 \text{ if } \mathbf{z}_{it}\boldsymbol{\delta} + \bar{\mathbf{z}}_i\boldsymbol{\psi}_b + \epsilon_{it} \leq 0, \\ d_{it} &= 2 \text{ if } \mathbf{z}_{it}\boldsymbol{\delta} + \bar{\mathbf{z}}_i\boldsymbol{\psi}_b + \epsilon_{it} > 0. \end{aligned} \tag{8}$$

where $\eta_{itj} = a_{cij} + u_{itj}$, $j = 1, 2$, and $\epsilon_{it} = a_{bi} + v_{it}$. We formulate the following assumption:

Assumption 3

(i) (η_{tj}, ϵ_t) *are independent of* $(\mathbf{z}_1, \ldots, \mathbf{z}_T)$, $j = 1, 2$, $t = 1, \ldots, T$.

[3] Mundlak (1978) has shown that when this method is used for estimating linear models, the resulting $\hat{\boldsymbol{\beta}}_j$ is identical to the fixed effects estimator of $\boldsymbol{\beta}_j$.

(ii) For each t,

$$\begin{pmatrix} \eta_{tj} \\ \epsilon_t \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ \rho_j & 1 \end{bmatrix} \right), \quad j = 1, 2. \tag{9}$$

(iii) $0 < \frac{1}{T}\sum_{t=1}^{T} \mathrm{P}(d_t = j) < 1,\ j = 1, 2.$

The assumption is stated for the underlying population; hence, subscript i is dropped. However, by random sampling, it also holds for a randomly selected unit from the population. Part (i) imposes strict exogeneity of covariates in (8). The normality assumption in (ii), is rather standard in the literature and permits obtaining formulae for conditional probabilities and partial effects. Because each i can only belong to one group in a given t, $\mathrm{Corr}(\eta_{it1}, \eta_{it2})$ is not defined. Moreover, note that $\mathrm{Corr}(\eta_{it1}, \eta_{is2})$ and $\mathrm{Corr}(\eta_{itj}, \epsilon_{is})$, $t \neq s$, are not specified, but can be (and likely are) different from zero. Finally, part (iii) ensures that there are cross section units in each group in at least some periods in the population.

Under Assumption 3, the two-group model is a switching probit model (Carrasco, 2001), which is analogous to a linear switching regression model discussed in the literature (Lee 1978; Maddala and Nelson, 1975; Maddala 1983; Manski et al., 1992). In the linear case, nonrandom group assignment is usually addressed by constructing a correction term. In binary response models, however, this approach is inapplicable because of the nonlinearity of the conditional mean. Instead, using the properties of normal distributions, we can write

$$\begin{aligned} \eta_{itj} &= \rho_j \epsilon_{it} + e_{itj}, \\ e_{itj}|\mathbf{z}_i, \epsilon_{it} &\sim Normal(0, 1 - \rho_j^2), \end{aligned} \tag{10}$$

so that $y_{itj} = 1[\mathbf{x}_{it}\boldsymbol{\beta}_j + \bar{\mathbf{z}}_i\boldsymbol{\psi}_{cj} + \rho_j\epsilon_{it} + e_{itj} > 0]$, $t = 1, \ldots, T$, $j = 1, 2$. Clearly, any likelihood function that ignores the error correlation would be misspecified and would result in inconsistent estimators.

Define $\mathbf{w}_{it} = (\mathbf{x}_{it}, \bar{\mathbf{z}}_i)$, $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j', \boldsymbol{\psi}_{cj}')'$, $\mathbf{q}_{it} = (\mathbf{z}_{it}, \bar{\mathbf{z}}_i)$, and $\boldsymbol{\pi} = (\boldsymbol{\delta}', \boldsymbol{\psi}_{bj}')'$. Then, the conditional probability for $j = 1$, period t, can be written as

$$\begin{aligned} \mathrm{P}(y_{it} = 1 | d_{it} = 1, \mathbf{z}_i) &= \frac{\mathrm{P}(-e_{it1} < \mathbf{w}_{it}\boldsymbol{\theta}_1 + \rho_1 \epsilon_{it}, \epsilon_{it} \leq -\mathbf{q}_{it}\boldsymbol{\pi} | \mathbf{z}_i)}{\mathrm{P}(\epsilon_{it} \leq -\mathbf{q}_{it}\boldsymbol{\pi} | \mathbf{z}_i)} \\ &= \frac{\int_{-\infty}^{-\mathbf{q}_{it}\boldsymbol{\pi}} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_1 + \rho_1 \epsilon}{\sqrt{1-\rho_1^2}}\right) \phi(\epsilon) d\epsilon}{1 - \Phi(\mathbf{q}_{it}\boldsymbol{\pi})}, \end{aligned} \tag{11}$$

and the corresponding conditional probability for $j = 2$ is

$$\mathrm{P}(y_{it} = 1 | d_{it} = 2, \mathbf{z}_i) = \frac{\int_{-\mathbf{q}_{it}\boldsymbol{\pi}}^{\infty} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_2 + \rho_2 \epsilon}{\sqrt{1-\rho_2^2}}\right) \phi(\epsilon) d\epsilon}{\Phi(\mathbf{q}_{it}\boldsymbol{\pi})}, \tag{12}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal density and cumulative distribution functions, respectively. Note that $\mathrm{P}(y_{itj} = 1 | \mathbf{z}_i) = \Phi(\mathbf{w}_{it}\boldsymbol{\theta}_j)$ and is the same regardless of the number of groups and ordering. These probabilities can be used to obtain APE_j^U and APE_j^C, as will be discussed in Section 3.

2.3 Model for multiple ordered groups

Let the total number of groups, J, exceed two. Using the unobserved effects model in (7), define vectors $\mathbf{q}_{it}$ and $\boldsymbol{\pi}$ as in the previous section. Also, define d_{it}^* and d_{it} as

$$\begin{aligned} & d_{it}^* = \mathbf{q}_{it}\boldsymbol{\pi} + \epsilon_{it}, \\ & d_{it} = j \quad \text{if} \quad C_{j-1} < d_{it}^* \leq C_j, \quad j = 1, \ldots, J, \\ & C_0 = -\infty, \quad \text{and} \quad C_J = \infty. \end{aligned} \tag{13}$$

Such a model is applicable, for example, when the goal is to study the probability of self-employment by age or education level.

Similar to a two-group case, it is convenient to assume that the distribution of ϵ_{it} is normal, which results in an ordered probit model that now has to be combined with a main binary response outcome. Formally, let Assumption 3 hold for $j = 1, 2, \ldots, J$, so that the errors are independent of $\mathbf{z}_i$ and have a joint normal distribution. Then, using the argument similar to the one in Section 2.2, we can write

$$y_{itj} = 1[\mathbf{w}_{it}\boldsymbol{\theta}_j + \rho_j \epsilon_{it} + e_{itj} > 0], \quad t = 1, \ldots, T, \quad j = 1, \ldots, J, \tag{14}$$
$$e_{itj}|\mathbf{w}_{it}, \epsilon_{it} \sim Normal(0, 1 - \rho_j^2),$$

where $\mathbf{w}_{it}$ and $\boldsymbol{\theta}_j$ are defined as in Section 2.2.

From (13) and (14), the conditional probabilities for each group are

$$\mathrm{P}(y_{it} = 1|d_{it} = j, \mathbf{z}_i) = \frac{\int_{C_{j-1}-\mathbf{q}_{it}\boldsymbol{\pi}}^{C_j-\mathbf{q}_{it}\boldsymbol{\pi}} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_j+\rho_j\epsilon}{\sqrt{1-\rho_j^2}}\right)\phi(\epsilon)d\epsilon}{\Phi(C_j - \mathbf{q}_{it}\boldsymbol{\pi}) - \Phi(C_{j-1} - \mathbf{q}_{it}\boldsymbol{\pi})}, \quad j = 2, \ldots, J-1,$$
$$C_0 = -\infty, \qquad C_J = \infty.$$

2.4 Model for unordered multiple groups

In some cases, there may be more than two groups that are not ordered. For example, one might want to study the determinants of job satisfaction or promotion among workers in different occupations. Then, the choice of $d_{it} = j$ can be described in the context of a multinomial response model. To formalize ideas, define

$$d^*_{itj} = \mathbf{q}_{it}\boldsymbol{\pi}_j + \epsilon_{itj}, \quad t = 1, \ldots, T, \quad j = 1, \ldots, J, \tag{15}$$

where the parameter vector and error term now vary by group.

Following the standard formulation of a multinomial response model, the cross-section unit i will be in group j in period t if it has the highest chance of belonging to that group.

In the case of self-selection, choice j is the best option in the available set:

$$d_{it} = j \quad \text{if} \quad d^*_{itj} = \max\{d^*_{it1}, d^*_{it2}, \dots, d^*_{itJ}\} \tag{16}$$

The choice in (16) will be made if $\mathbf{q}_{it}\boldsymbol{\pi}_j + \epsilon_{itj} > \mathbf{q}_{it}\boldsymbol{\pi}_l + \epsilon_{itl}$ for all $l \neq j$. It is clear that only differences between d^*_{itj} are identified, so that a reference category needs to be assigned – a feature that is common to all multinomial response models. We formulate the following assumption:

Assumption 4

(i) $(\eta_{tj}, \epsilon_{t1}, \dots, \epsilon_{tJ})$ *are independent of* $(\mathbf{z}_1, \dots, \mathbf{z}_T)$*, for* $j = 1, \dots, J$*,* $t = 1, \dots, T$*.*

(ii) For each t*,*

$$\begin{pmatrix} \eta_{tj} \\ \epsilon_{t1} \\ \dots \\ \epsilon_{tj} \\ \dots \\ \epsilon_{tJ} \end{pmatrix} \sim Normal \left(\begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \dots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \dots & \rho_j & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_j & 0 & \dots & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \right), \quad j = 1, \dots, J. \tag{17}$$

(iii) $0 < \frac{1}{T}\sum_{t=1}^{T} \mathrm{P}(d_t = j) < 1$*,* $j = 1, \dots, J$*.*

Note that Assumption 4 imposes restrictions on the variance-covariance matrix. Theoretically, one could allow the variance in part (ii) to be completely unrestricted. However, in practice it is usually necessary to impose restrictions to ensure feasibility of the estimation. In the present context, the imposed restrictions are reasonable. Specifically, $\mathrm{Cov}(\epsilon_{tj}, \epsilon_{tl}) = 0$ for $j \neq l$ is effectively true by independence across i because each cross section unit can only belong to one group in a given t. For the same reason, $\mathrm{Cov}(\eta_{tj}, \epsilon_{tl}) = 0$

for $j \neq l$ holds. Importantly, $\text{Cov}(\epsilon_{tj}, \epsilon_{sl})$ and $\text{Cov}(\eta_{tj}, \epsilon_{sl})$, $s \neq l$, are left completely unrestricted, which is consistent with what would be observed in the population. Indeed, these covariances are likely different from zero because of transitions across groups over time.

Define $\tilde{\epsilon}_{itl} = \epsilon_{itj} - \epsilon_{itl}$, and $\tilde{\boldsymbol{\pi}}_l = \boldsymbol{\pi}_j - \boldsymbol{\pi}_l$, for $l \neq j$. Then, under Assumption 4, for group $j = 1$, for example, we obtain

$$\begin{aligned} \mathrm{P}(y_{it} = 1, d_{it} = 1|\mathbf{z}_i) &= \int_{-\mathbf{w}_{it}\boldsymbol{\theta}_1}^{\infty} \int_{-\mathbf{q}_{it}\tilde{\boldsymbol{\pi}}_2}^{\infty} \cdots \int_{-\mathbf{q}_{it}\tilde{\boldsymbol{\pi}}_J}^{\infty} \phi(e_{it1}, \tilde{\epsilon}_2 \ldots \tilde{\epsilon}_J; \Sigma) du_1 d\tilde{\epsilon}_2 \ldots d\tilde{\epsilon}_J, \quad (18) \\ \mathrm{P}(d_{it} = 1|\mathbf{z}_i) &= \int_{-\mathbf{q}_{it}\tilde{\boldsymbol{\pi}}_2}^{\infty} \cdots \int_{-\mathbf{q}_{it}\tilde{\boldsymbol{\pi}}_J}^{\infty} \phi(\tilde{\epsilon}_2, \ldots, \tilde{\epsilon}_J; \tilde{\Sigma}) d\tilde{\epsilon}_2 \ldots d\tilde{\epsilon}_J, \end{aligned}$$

where Σ and $\tilde{\Sigma}$ are variance-covariance matrices of vectors $(e_{it1}, \tilde{\epsilon}_2 \ldots \tilde{\epsilon}_J)'$ and $(\tilde{\epsilon}_2 \ldots \tilde{\epsilon}_J)'$, respectively. Using (18), the conditional probability is obtained as $P(y_{it} = 1|d_{it} = 1, \mathbf{z}_i) = \frac{P(y_{it}=1, d_{it}=1|\mathbf{z}_i)}{P(d_{it}=1|\mathbf{z}_i)}$. Probabilities $P(y_{it} = 1|d_{it} = j, \mathbf{z}_i)$, $j = 2, \ldots, J$, are obtained similarly.

Because equation (18) does not have a closed form solution, one would need to numerically evaluate a J-dimensional integral. Although simulated likelihood methods have been helpful in addressing computational difficulties, the estimation may still be problematic if there are more than four groups, especially when the number of observations is not very large. Both the number of parameters and dimension of the integral grow with the number of groups, which may cause computational issues (e.g. nonconvergence) when performing the optimization. Therefore, we also discuss a different approach.

The unordered multiple groups case can be considered in the context of selection models, where the choice is made between the best option (observed choice) and the second best alternative. Define a binary indicator for group j in period t as

$$\begin{aligned} \omega_{itj} &= 1[\mathbf{q}_{it}\boldsymbol{\pi}_j + \epsilon_{itj} > \bar{d}_{itj}], \quad (19) \\ \bar{d}_{itj} &= \max_{l \neq j}\{\mathbf{q}_{it}\boldsymbol{\pi}_l + \epsilon_{itl}\}, \end{aligned}$$

which can be re-written as

$$\omega_{itj} = 1[\mathbf{q}_{it}\bar{\boldsymbol{\pi}}_j + \bar{\epsilon}_{itj} > 0], \quad t = 1, \ldots, T, \quad j = 1, \ldots, J, \tag{20}$$

where $\bar{\boldsymbol{\pi}}_j$ is the difference between $\boldsymbol{\pi}_j$ and the vector of parameters that correspond to $\bar{d}_{itj}$, and $\bar{\epsilon}_{itj}$ is the difference between ϵ_{itj} and the error corresponding to $\bar{d}_{itj}$. Because in the unordered case the second best option is not known, $\bar{\boldsymbol{\pi}}_j$ is a weighted average of $\boldsymbol{\pi}_j - \boldsymbol{\pi}_l$, $l \neq j$, where weights depend on the probability that group l is the best alternative to j. Notice that in this model it is not possible to estimate $\boldsymbol{\pi}_j$. Correspondingly, conditional APE cannot be estimated. However, the consistent estimators of $\boldsymbol{\theta}$ and APE^U can be obtained.

Assumption 5

(i) $(\eta_{tj}, \bar{\epsilon}_{tj})$ *are independent of* $(\mathbf{z}_1, \ldots, \mathbf{z}_T)$, $j = 1 \ldots, J$, $t = 1, \ldots, T$.

(ii) For each t,

$$\begin{pmatrix} \eta_{tj} \\ \bar{\epsilon}_{tj} \end{pmatrix} \sim Normal\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ \rho_j & 1 \end{bmatrix} \right), \quad j = 1, \ldots, J. \tag{21}$$

(iii) $0 < \frac{1}{T}\sum_{t=1}^{T} \mathrm{P}(d_t = j) < 1$, $j = 1, \ldots, J$.

Under Assumption 5, conditional probabilities for each j and t are obtained as

$$\mathrm{P}(y_{it} = 1 | d_{it} = j, \mathbf{z}_i) = \mathrm{P}(y_{it} = 1 | \omega_{itj} = 1, \mathbf{z}_i) = \frac{\int_{-\infty}^{\mathbf{q}_{it}\bar{\boldsymbol{\pi}}_j} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}j + \rho_j \bar{\epsilon}}{\sqrt{1-\rho_j^2}}\right) \phi(\bar{\epsilon}) d\bar{\epsilon}}{\Phi(\mathbf{q}_{it}\bar{\boldsymbol{\pi}}_j)}. \tag{22}$$

3 Estimation

To estimate the models presented in Section 2, one can use the maximum likelihood estimator (MLE). Full MLE would be an efficient estimator, but it requires specifying the conditional density of $(y_{i1}, ..., y_{iT}, d_{i1}, ..., d_{iT})$, given $\mathbf{z}_i$. Because $\{y_{i1}, ..., y_{iT}\}$ and $\{d_{i1}, ..., d_{iT}\}$ are likely serially dependent (largely due to the presence of time-constant unobserved effects), the joint density function would generally be very complicated. That would increase the computational cost and could make the estimation infeasible, unless additional restrictions on the error variance-covariance matrix are imposed. In this paper, we use a more feasible partial MLE estimator, which only requires specifying the conditional density in a given t.

Consider a general model with $J \geq 2$, but for the moment ignore the second unordered groups model discussed at the end of Section 2.4. Denote $\mathbf{Y}_{it} = (y_{it}, d_{it})$. For each t, the joint density of $\mathbf{Y}_{it}$ conditional on $\mathbf{z}_i$ for observation i is

$$f_t(\mathbf{Y}_{it}|\mathbf{z}_i, \boldsymbol{\gamma}) = \mathrm{P}_{it,11}^{y_{it}s_{it1}} \cdot \mathrm{P}_{it,01}^{(1-y_{it})s_{it1}} \cdot \ldots \cdot \mathrm{P}_{it,1J}^{y_{it}s_{itJ}} \cdot \mathrm{P}_{it,0J}^{(1-y_{it})s_{itJ}}, \tag{23}$$

where $\mathrm{P}_{it,1j} = \mathrm{P}(y_{it} = 1, d_{it} = j|\mathbf{z}_i; \boldsymbol{\gamma})$, $\mathrm{P}_{it,0j} = \mathrm{P}(y_{it} = 0, d_{it} = j|\mathbf{z}_i; \boldsymbol{\gamma})$, $j = 1, ..., J$, $\boldsymbol{\gamma} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_J', \boldsymbol{\pi}, \rho_1, \ldots, \rho_J)'$ for the models in Sections 2.2 and 2.3, and $\boldsymbol{\gamma} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_J', \bar{\boldsymbol{\pi}}_1, \ldots, \bar{\boldsymbol{\pi}}_J, \rho_1, \ldots, \rho_J)'$ for the first model in Section 2.4.

Joint probabilities for the two-group model are

$$\begin{aligned} \mathrm{P}_{it,1j} &= \int_{-\infty}^{-\mathbf{q}_{it}\boldsymbol{\pi}} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_j + \rho_j\epsilon}{\sqrt{1-\rho_j^2}}\right)\phi(\epsilon)d\epsilon, \\ \mathrm{P}_{it,0j} &= \int_{-\infty}^{-\mathbf{q}_{it}\boldsymbol{\pi}} \left[1 - \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_j + \rho_j\epsilon}{\sqrt{1-\rho_j^2}}\right)\right]\phi(\epsilon)d\epsilon, \quad j = 1, 2. \end{aligned} \tag{24}$$

For ordered multiple groups, the probabilities are

$$
\begin{aligned}
\mathrm{P}_{it,1j} &= \int_{C_{j-1}-\mathbf{q}_{it}\boldsymbol{\pi}}^{C_j-\mathbf{q}_{it}\boldsymbol{\pi}} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_j+\rho_j\epsilon}{\sqrt{1-\rho_j^2}}\right)\phi(\epsilon)d\epsilon, &(25)\\
\mathrm{P}_{it,0j} &= \int_{C_{j-1}-\mathbf{q}_{it}\boldsymbol{\pi}}^{C_j-\mathbf{q}_{it}\boldsymbol{\pi}} \left[1-\Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}_j+\rho_j\epsilon}{\sqrt{1-\rho_j^2}}\right)\right]\phi(\epsilon)d\epsilon,\\
& C_0=-\infty,\quad C_J=\infty,\quad j=1,\ldots,J.
\end{aligned}
$$

In the unordered case, the first equation in (18) specifies $\mathrm{P}_{it,1j}$ for the first model in Section 2.4. Probabilities for $j=2,\ldots,J$ are obtained similarly. Changing the limits of integration permits computing $\mathrm{P}_{it,0j}$.

Denote $l_{it}(\boldsymbol{\gamma}) = \ln[f_t(\mathbf{Y}_{it}|\mathbf{z}_i,\boldsymbol{\gamma})]$. Partial MLE is obtained by solving the following maximization problem:

$$\max_{\boldsymbol{\gamma}} \sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\boldsymbol{\gamma}). \tag{26}$$

Optimization can be performed using usual optimization techniques, such as the Newton-Raphson method and Berndt, Hall, Hall, and Hausman algorithm (Berndt, Hall, Hall, and Hausman, 1974).

Section 2.4 discusses estimation of parameters in the unordered multiple groups model under Assumption 4.2. For that model, for each group j the conditional density for observation i in period t can be written as

$$f_t(\mathbf{Y}_{it}|\mathbf{z}_i,\boldsymbol{\gamma}) = \mathrm{P}_{it,11}^{y_{it}\omega_{itj}}\cdot \mathrm{P}_{it,01}^{(1-y_{it})\omega_{itj}}\cdot \mathrm{P}_{it,0}^{(1-\omega_{itj})}, \tag{27}$$

where

$$\mathrm{P}_{it,11} = \int_{-\infty}^{\mathbf{q}_{it}\bar{\boldsymbol{\pi}}_j} \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta}j+\rho_j\bar{\epsilon}}{\sqrt{1-\rho_j^2}}\right)\phi(\bar{\epsilon})d\bar{\epsilon}, \tag{28}$$

$$
\begin{aligned}
\mathrm{P}_{it,01} &= \int_{-\infty}^{\mathbf{q}_{it}\bar{\boldsymbol{\pi}}_j} \left[1 - \Phi\left(\frac{\mathbf{w}_{it}\boldsymbol{\theta} j + \rho_j \bar{\epsilon}}{\sqrt{1-\rho_j^2}}\right)\right] \phi(\bar{\epsilon}) d\bar{\epsilon}, \\
\mathrm{P}_{it,0} &= 1 - \Phi(\mathbf{q}_{it}\bar{\boldsymbol{\pi}}_j).
\end{aligned}
$$

Then, $\boldsymbol{\gamma}_j$ can be estimated separately for each j by partial MLE.

As mentioned in Section 2, estimating APEs is often of interest. For all models, unconditional APE of a continuous variable x_k can be estimated as

$$
\widehat{APE}_{j,k}^{U} = \hat{\beta}_{jk} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi(\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_j), \quad j = 1, \dots, J. \tag{29}
$$

To obtain the estimator of APE_j^C, note that from (7) we can write

$$
\mathrm{P}(y_j = 1 | d = j, \mathbf{x}, c_j) = \mathrm{P}(y_j = 1 | d = j, \mathbf{z}, \bar{\mathbf{z}}, a_{cj}, a_{bj}, v), \tag{30}
$$

where we use the fact that d is a deterministic function of $(\mathbf{z}, \bar{\mathbf{z}}, a_{bj}, v)$ in the population. After interchanging the integration and differentiation, it follows that conditional APE of a continuous variable x_k in group j is

$$
\frac{\partial \mathrm{E}_{\bar{\mathbf{z}}, a_{cj}, a_{bj}, v}[\mathrm{P}(y_j = 1 | d = j, \mathbf{z}, \bar{\mathbf{z}}, a_{cj}, a_{bj}, v)]}{\partial x_k}. \tag{31}
$$

Hence, for $j = 2$ in a two-group model, the time-averaged conditional APE can be estimated as

$$
\begin{aligned}
\widehat{APE}_{2,k}^{C} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Bigg[& \hat{\delta}_k \cdot \frac{\phi(\mathbf{q}_{it}\hat{\boldsymbol{\pi}})}{\Phi(\mathbf{q}_{it}\hat{\boldsymbol{\pi}})} \cdot \Phi\left(\frac{\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2 + \hat{\rho}_2 \mathbf{q}_{it}\hat{\boldsymbol{\pi}}}{\sqrt{1-\hat{\rho}_2^2}}\right) \\
& + \frac{1}{\Phi(\mathbf{q}_{it}\hat{\boldsymbol{\pi}})} \cdot \frac{\hat{\beta}_{2k}}{\sqrt{1-\hat{\rho}_2^2}} \cdot \int_{-\infty}^{\mathbf{q}_{it}\hat{\boldsymbol{\pi}}} \phi\left(\frac{\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2 + \hat{\rho}_2 \epsilon}{\sqrt{1-\hat{\rho}_2^2}}\right) \phi(\epsilon) d\epsilon - \hat{\delta}_k \cdot \frac{\phi(\mathbf{q}_{it}\hat{\boldsymbol{\pi}})}{\Phi(\mathbf{q}_{it}\hat{\boldsymbol{\pi}})} \cdot \hat{\mathrm{P}}_{it,12}^{C} \Bigg],
\end{aligned} \tag{32}
$$

where $\hat{\delta}_k$, $\hat{\beta}_{2k}$, $\hat{\boldsymbol{\pi}}$, $\hat{\boldsymbol{\theta}}_2$, and $\hat{\rho}_2$, are the estimators of δ_k, β_{2k}, $\boldsymbol{\pi}$, $\boldsymbol{\theta}_2$, and ρ_2, respectively, and $\hat{\mathrm{P}}^{C}_{it,12}$ is the estimator of $\mathrm{P}(y_{it} = 1|d_{it} = 2, \mathbf{z}_i)$, which is defined in equation (12). Correspondingly, $\widehat{APE}_{1,k}$ is obtained by replacing $\hat{\boldsymbol{\pi}}$ and $\hat{\delta}_k$ with $-\hat{\boldsymbol{\pi}}$ and $-\hat{\delta}_k$, respectively, and changing $\hat{\boldsymbol{\theta}}$ and $\hat{\rho}$ subscripts to one. Notice that when the group assignment is random, the partial effects on the conditional probabilities are the same as the unconditional partial effects. However, they are different when $\rho_j \neq 0$.

For the ordered groups model, the time-averaged conditional APE for each j can be estimated using

$$
\begin{aligned}
\widehat{APE}^{C}_{j,k} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{\hat{\delta}_k}{\Phi(\hat{\alpha}_j)-\Phi(\hat{\alpha}_{j-1})}\cdot\left[\phi(\hat{\alpha}_{j-1})\Phi\left(\frac{\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_j+\hat{\rho}_j\hat{\alpha}_{j-1}}{\sqrt{1-\hat{\rho}_j^2}}\right)\right.\right. \qquad (33)\\
&- \left.\phi(\hat{\alpha}_j)\Phi\left(\frac{\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_j+\hat{\rho}_j\hat{\alpha}_j}{\sqrt{1-\hat{\rho}_j^2}}\right)\right]\\
&+ \frac{1}{\Phi(\hat{\alpha}_j)-\Phi(\hat{\alpha}_{j-1})}\cdot\frac{\hat{\beta}_{jk}}{\sqrt{1-\hat{\rho}_j^2}}\cdot\int_{-\hat{\alpha}_{j-1}}^{\hat{\alpha}_j}\phi\left(\frac{\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_j+\hat{\rho}_j\epsilon}{\sqrt{1-\hat{\rho}_j^2}}\right)\phi(\epsilon)d\epsilon\\
&+ \left.\hat{\delta}_k\cdot\frac{\phi(\hat{\alpha}_j)-\phi(\hat{\alpha}_{j-1})}{\Phi(\hat{\alpha}_j)-\Phi(\hat{\alpha}_{j-1})}\cdot\hat{\mathrm{P}}^{C}_{it,1j}\right\},\\
&\quad \hat{C}_0=-\infty,\quad \hat{C}_J=\infty,\quad \hat{\alpha}_j=C_j-\mathbf{q}_{it}\hat{\boldsymbol{\pi}},
\end{aligned}
$$

where $\hat{C}_j$ and $\hat{\alpha}_j$ are the estimators of C_j and α_j, respectively, and $\hat{\mathrm{P}}^{C}_{it,1j}$ is the estimator of $\mathrm{P}(y_{it} = 1|d_{it} = j, \mathbf{z}_i)$ defined in equation (15).

Given the complexity of the conditional probability function for multiple unordered groups, APE^C_j for the first model in Section 2.4 would have to be obtained by numerically computing the derivative. However, APE^C_j for the second unordered groups model can be estimated using (32) after replacing $\hat{\boldsymbol{\pi}}$ with $\hat{\boldsymbol{\pi}}_j$.

From (32) and (33), it is seen that the sign of $\widehat{APE}^{C}_{j,k}$ does not necessarily coincide with the sign of β_{jk}. When x_k increases, it affects not only the probability that $y_j = 1$, but also the probability that $d = j$. Consequently, more or fewer units are induced into group j, so that the size and composition of the group change. Hence, the direction of the change in $\mathrm{P}(y_j = 1|d = j)$ generally depends on both β_{jk} and δ_k (or, δ_{jk}) and is uncertain. However, if x_k has no effect on sorting ($\delta_k = 0$), only the second term in (32) and (33) is different from zero, and the signs of $\widehat{APE}^C_{j,k}$ and β_{jk} coincide.

Average partial effects of discrete variables (e.g. binary indicators) are obtained as average changes in estimated probabilities. For a discrete variable h in group j

$$\widehat{APE}^M_{j,h} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\widehat{\mathrm{P}}^{M1}_{it,1j} - \widehat{\mathrm{P}}^{M0}_{it,1j}\right], \quad M = U, C, \tag{34}$$

where $\widehat{\mathrm{P}}^{U1}_{it,1j} = \mathrm{P}(y_{itj} = 1|\mathbf{z}^l_i;\hat{\boldsymbol{\gamma}})$, $\widehat{\mathrm{P}}^{C1}_{it,1j} = \mathrm{P}(y_{itj} = 1|d_{it} = j, \mathbf{z}^l_i;\hat{\boldsymbol{\gamma}})$, $l = 0, 1$, $\mathbf{z}^l_i = (\mathbf{x}^l_{it}, \mathbf{z}_{it,1}, \bar{\mathbf{z}}_i)$, $\mathbf{x}^1_{it} = (x_{it,1}, \ldots, x_{it,h-1}, x^1_{it,h}, x_{it,h+1}, \ldots, x_{it,K})$, and $\mathbf{x}^0_{it} = \big(x_{it,1}, \ldots, x_{it,h-1}, x^0_{it,h}, x_{it,h+1}, \ldots, x_{it,K}\big)$.

Note that (34) can also be used to obtain conditional APE for continuous variables. One can simply consider a particular (e.g. one unit) increase in x_k from a given value, such as the sample mean of x_k. Given the complexity of formulas in (32) and (33), using (34) may be preferred. It appears to be especially attractive when obtaining APE^C_j in a model with multiple unordered groups.

In (34), estimators of APE are obtained by averaging over the distribution of all covariates other then the one whose effect is being estimated. Alternatively, one can obtain APE evaluated at particular values of other explanatory variables ($\widetilde{\mathbf{z}}$), such as sample means or median values. Then, equation (34), for example, would become

$$\widetilde{APE}^M_{j,h} = \widetilde{\mathrm{P}}^{M1}_{1j} - \widetilde{\mathrm{P}}^{M0}_{1j}, \quad M = U, C, \tag{35}$$

where $\widetilde{\mathrm{P}}^{U1}_{1j} = \mathrm{P}(y_j = 1|\widetilde{\mathbf{z}}^l;\hat{\boldsymbol{\gamma}})$, $\widetilde{\mathrm{P}}^{C1}_{1j} = \mathrm{P}(y_j = 1|d = j, \widetilde{\mathbf{z}}^l;\hat{\boldsymbol{\gamma}})$, $l = 0, 1$, for some fixed values $\widetilde{\mathbf{z}}^l = (\widetilde{\mathbf{x}}^l, \widetilde{\mathbf{z}}_1, \widetilde{\bar{\mathbf{z}}})$, $\widetilde{\mathbf{x}}^1 = (\widetilde{x}_1, \ldots, \widetilde{x}_{h-1}, x^1_h, \widetilde{x}_{h+1}, \ldots, \widetilde{x}_K)$ and $\widetilde{\mathbf{x}}^0 = (\widetilde{x}_1, \ldots, \widetilde{x}_{h-1}, x^0_h, \widetilde{x}_{h+1}, \ldots, \widetilde{x}_K)$. APE of continuous covariates are obtained similarly.

4 Consistency and asymptotic normality

A necessary condition for the consistency of proposed estimators is that the true parameter vector is identified. Let $\boldsymbol{\gamma}_0$ be the true parameter vector. Vector $\boldsymbol{\gamma}_0$ is identified if it uniquely maximizes the expected value of the log likelihood function in the population. For the likelihood function formulated in Section 3, this condition will fail if there is perfect collinearity among covariates. On the other hand, the exclusion restriction is not required. The parameters are identified even if covariates in the main and sorting equations are the same. However, when $\mathbf{z}_{it} = \mathbf{x}_{it}$, identification relies on the nonlinearity of the likelihood function, which makes the estimation less reliable. Hence, we follow the literature and assume that $\mathbf{z}_{it}$ contains at least one variable that is not in $\mathbf{x}_{it}$.

In addition to identification, several other conditions are needed for consistency. Specifically, we make the following assumption:

Assumption 6

(i) For each model, the corresponding assumptions in Section 2 hold, so that $f_t(\cdot|\mathbf{z}, \boldsymbol{\gamma})$ is the true density function for each t.

(ii) For some $\boldsymbol{\gamma}_0 \in \boldsymbol{\Gamma}$, $\boldsymbol{\gamma}_0$ is the unique solution to the maximization problem

$$\max_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} \mathrm{E}[l_i(\boldsymbol{\gamma})], \qquad l_i(\boldsymbol{\gamma}) = \sum_{t=1}^{T} l_{it}(\boldsymbol{\gamma}). \tag{36}$$

(iii) $\boldsymbol{\Gamma}$ is a compact set.

(iv) $|l(\mathbf{Y}, \mathbf{z}, \boldsymbol{\gamma})| \leq b(\mathbf{Y}, \mathbf{z})$, all $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, and $\mathrm{E}[b(\mathbf{Y}, \mathbf{z})] < \infty$.

Assumptions (i) and (ii) state that the model is correctly specified, and $\boldsymbol{\gamma}_0$ is identified. In addition to compactness, continuity of the log-likelihood function is required. Because $l_{it}(\boldsymbol{\gamma})$ in Section 3 is continuous in $\boldsymbol{\gamma}$, we do not state this assumption separately, as it holds if Assumption 6 (i) holds. Finally, assumption (iv) requires the expected value of $l(\mathbf{Y}, \mathbf{z}, \boldsymbol{\gamma})$ to be bounded across $\boldsymbol{\gamma}$.

Theorem 1. If Assumption 6 holds, then $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$ as $N \to \infty$, T fixed.

Proof: For a proof, we follow Newey and McFadden (1994). Assumption 6 parts (iii) and (iv) and the continuity of the likelihood function satisfy the conditions of Lemma 2.4 in Newey and McFadden (1994) for a given t. Then, by Lemma 2.4 in Newey and McFadden (1994), $\frac{1}{N}\sum_{i=1}^{N} l_{it}(\hat{\boldsymbol{\gamma}})$ uniformly converges in probability to $\mathrm{E}[l_{it}(\boldsymbol{\gamma}_0)]$ for each t, so that $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\hat{\boldsymbol{\gamma}})$ uniformly converges in probability to $\mathrm{E}[\sum_{t=1}^{T} l_{it}(\boldsymbol{\gamma}_0)] = \mathrm{E}[l_i(\boldsymbol{\gamma}_0)]$ for a fixed T. Assumption 6 parts (ii) and (iii), the continuity of the likelihood function, and uniform convergence satisfy the conditions of Theorem 2.1 in Newey and McFadden (1994). By Theorem 2.1 in Newey and McFadden (1994), $\hat{\gamma} \xrightarrow{p} \gamma$. ■

To show asymptotic normality, we note that for the conditional density in Section 3, $l_i(\boldsymbol{\gamma})$ is twice continuously differentiable in the interior of $\boldsymbol{\Gamma}$. Hence, we can define

$$\begin{aligned}
\mathbf{S}(\mathbf{Y}_i, \mathbf{z}_i, \boldsymbol{\gamma}) &= \mathbf{S}_i(\boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} l_i(\boldsymbol{\gamma})' = \sum_{t=1}^{T} \nabla_{\boldsymbol{\gamma}} l_{it}(\boldsymbol{\gamma})' = \sum_{t=1}^{T} \mathbf{S}_{it}(\boldsymbol{\gamma}), \qquad (37) \\
\mathbf{H}(\mathbf{Y}_i, \mathbf{z}_i, \boldsymbol{\gamma}) &= \mathbf{H}_i(\boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} \mathbf{S}_i(\boldsymbol{\gamma}) = \nabla^2_{\boldsymbol{\gamma}} l_i(\boldsymbol{\gamma}) = \sum_{t=1}^{T} \mathbf{H}_{it}(\boldsymbol{\gamma}).
\end{aligned}$$

Formulate an additional assumption:

Assumption 7

(i) $\boldsymbol{\gamma}_0$ *is in the interior of* $\boldsymbol{\Gamma}$.

(ii) Each element of $\mathbf{H}(\mathbf{Y}, \mathbf{z}, \boldsymbol{\gamma})$ *is bounded in absolute value by a function* $b(\mathbf{Y}, \mathbf{z})$, *where* $\mathrm{E}[b(\mathbf{Y}, \mathbf{z})] < \infty$.

(iii) $\mathrm{E}[\mathbf{H}(\mathbf{Y}, \mathbf{z}, \boldsymbol{\gamma}_0)]$ *is negative definite.*

(iv) $\mathrm{E}[\mathbf{S}(\mathbf{Y}, \mathbf{z}, \boldsymbol{\gamma}_0)] = 0$.

(v) Each element of $\mathbf{S}(\mathbf{Y}, \mathbf{z}, \boldsymbol{\gamma}_0)$ *has finite second moment.*

Theorem 2. If Assumptions 6 and 7 hold, then

$$\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} Normal(0, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}), \tag{38}$$

where

$$\begin{aligned} \mathbf{A}_0 &= -\mathrm{E}\left[\mathbf{H}_i(\boldsymbol{\gamma}_0)\right] = -\sum_{t=1}^{T} \mathrm{E}\left[\mathbf{H}_{it}(\boldsymbol{\gamma}_0)\right], \\ \mathbf{B}_0 &= \mathrm{E}\left[\mathbf{S}_i(\boldsymbol{\gamma}_0)\mathbf{S}_i(\boldsymbol{\gamma}_0)'\right] = \mathrm{E}\left[\left(\sum_{t=1}^{T} \mathbf{S}_{it}(\boldsymbol{\gamma}_0)\right)\left(\sum_{t=1}^{T} \mathbf{S}_{it}(\boldsymbol{\gamma}_0)'\right)\right]. \end{aligned} \tag{39}$$

Proof: By construction, estimator $\hat{\boldsymbol{\gamma}}$ solves the first order condition

$$\sum_{i=1}^{N} \mathbf{S}_i(\hat{\boldsymbol{\gamma}}) = 0. \tag{40}$$

Using the mean-value expansion about $\boldsymbol{\gamma}_0$, obtain

$$\sum_{i=1}^{N} \mathbf{S}_i(\hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{N} \mathbf{S}_i(\boldsymbol{\gamma}_0) + \left(\sum_{i=1}^{N} \ddot{\mathbf{H}}_i\right)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0), \tag{41}$$

where $\ddot{\mathbf{H}}_i$ is evaluated at a different mean value. Using (40), we can set the left-hand side in (41) to zero, multiply by $\sqrt{N}$ and rearrange to obtain

$$\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = -\left(\frac{1}{N}\sum_{i=1}^{N} \ddot{\mathbf{H}}_i\right)^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \mathbf{S}_i(\boldsymbol{\gamma}_0)\right]. \tag{42}$$

By Central Limit Theorem,

$$\sqrt{N}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} Normal(0, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}). \tag{43}$$

which completes the proof. ■

The asymptotic variance of $\hat{\boldsymbol{\gamma}}$ can be estimated as

$$\begin{aligned} \widehat{\text{Avar}}(\hat{\boldsymbol{\gamma}}) &= \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}/N, \\ \hat{\mathbf{A}} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{H}_{it}(\hat{\boldsymbol{\gamma}}), \\ \hat{\mathbf{B}} &= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T}\mathbf{S}_{it}(\hat{\boldsymbol{\gamma}})\right)\left(\sum_{t=1}^{T}\mathbf{S}_{it}(\hat{\boldsymbol{\gamma}})'\right)\right]. \end{aligned} \tag{44}$$

Regarding statistical inference, several null hypotheses may be of particular interest. For example, to check whether sorting is random (so that the usual group-by-group estimation is valid), one can test $H_0 : \rho_j = 0$ or $H_0 : \rho_1 = \ldots = \rho_J = 0$ using a fully-robust Wald test. Equality of the coefficients in two or more groups can be tested either for each explanatory variable separately, or for the entire vector of parameters, $\boldsymbol{\theta}_j$.

5 Monte Carlo Simulations

To study the performance of proposed estimators in finite samples we conduct limited Monte Carlo experiments. In addition to the methods discussed in Section 3, parameters in each group were also estimated by pooled probit, which is a commonly used method in applied research. In every regression, the list of covariates was augmented by variable time means, $\bar{z}_i$. Hence, the focus is on assessing the benefits of accounting for nonrandom sorting.

Data were simulated for a two-group model, a model with three ordered groups, and the one with three unordered groups. Explanatory variables are $(1, x_{it}, \bar{x}_i, \bar{z}_i)$ in the main equations, and $(1, x_{it}, z_{it}, \bar{x}_i, \bar{z}_i)$ in the sorting equations. The covariates are generated as

$$x_{it} = b_{i1} + \zeta_{it1}, \tag{45}$$

$$
\begin{aligned}
z_{it} &= b_{i2} + \zeta_{it2}, \\
\bar{x}_i &= \sum_{t=1}^{T} x_{it}, \quad \bar{z}_i = \sum_{t=1}^{T} z_{it},
\end{aligned} \tag{46}
$$

where b_{ij} are independent across i, $b_{ij} \sim Normal(0, \sigma_b^2)$, $j = 1, 2$, and $\text{Corr}(b_{i1}, b_{i2}) = 0.25$. Correspondingly, $zeta_{itj}$ are independent across i and t, $\zeta_{itj} \sim Normal(0, \sigma_\zeta^2)$, $j = 1, 2$, $\sigma_b^2 + \sigma_\zeta^2 = 1$, and $\frac{\sigma_b^2}{\sigma_b^2 + \sigma_\zeta^2} = 0.5$.

The coefficients on time means are set at $\xi_{cj} = (-0.3, -0.3)'$, $j = 1, 2$, $\xi_b = (0.3, 0.3)'$ in all models. However, other population parameters vary by the model to ensure that cross-section units are approximately equally distributed across groups. In a two-group model, y_{itj} and d_{it} are generated as in (8), using $\beta_1 = (1, -1)'$, $\beta_2 = (0.5, 1)'$, $\delta = (0.1, 0.5, 1)'$. The response variables for the model with three ordered groups are created using $\beta_1 = (-0.5, -1)'$, $\beta_2 = (0.2, -2)'$, $\beta_3 = (-0.2, 2)'$, $\delta = (0.5, 0.5, 1)'$, and cut points $C_1 = -0.3$, $C_2 = 1.2$. In the model with three unordered groups, the parameters are set at $\beta_1 = (-0.5, -1)'$, $\beta_2 = (1, -2)'$, $\beta_3 = (1, 2)'$, $\delta_2 = (-0.5, 0.5, 1)'$, $\delta_2 = (-0.5, -0.5, 1.2)'$, and $j = 1$ is a base group.

For each j, error terms were generated as $\eta_{itj} = a_{cij} + u_{itj}$, $\epsilon_{it} = a_{bi} + v_{it}$, where $a_{cij} \sim Normal(0, \sigma_a^2)$, $a_{bi} \sim Normal(0, \sigma_a^2)$, $u_{itj} \sim Normal(0, \sigma_u^2)$, $v_{it} \sim Normal(0, \sigma_v^2)$, $\sigma_a^2 + \sigma_u^2 = \sigma_a^2 + \sigma_v^2 = 1$, $\frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2} = 0.5$, and $\text{Corr}(a_{cij}, a_{bi}) = \text{Corr}(u_{itj}, v_{it}) = \rho_j$. In the two-group model, data are simulated using $\rho_1 = -0.5$, $\rho_2 = 0.5$, while in both three-group models the correlations were $\rho_1 = 0.5$, $\rho_2 = 0.5$, $\rho_3 = -0.5$. We also generated data for $\rho_j = 0$, $\forall j$. Simulations were done for $T = 3$, $N = 300$, 500, and $1,000$, using $1,000$ replications.

Simulation results for $N = 300$ are presented in Tables 1-4. In all tables, PMLE is the partial MLE estimator discussed in Section 3. As seen in Table 1, joint PMLE estimators have slightly smaller biases than probit in a two-group model when $\rho_j = 0$. However, probit estimators have smaller average standard errors and, correspondingly, smaller root mean-square errors (RMSE). When ρ_j are different from zero, probit estimators have sizable biases, which is not the case for joint PMLE. As a result, RMSE tend to be larger for probit.

For the model with three ordered groups (Table 2), results are similar. Joint PMLE outperforms probit in terms of smaller biases and RMSE when $\rho_j \neq 0$, although it has larger standard errors, on average. In the model with three unordered groups (Table 3), both joint PMLE and PMLE for the second ("best alternative") model in Section 2.4 have smaller biases than probit when error correlations are different form zero. A notable exception is a relatively poor performance of the "best alternative" PMLE for estimating parameters in group $j = 1$ (noticeably larger biases and average standard errors). Thus, it appears that the "best alternative" PMLE is less reliable than the joint PMLE.

Table 4 displays simulation outcomes for the parameters in the sorting equations in all three models. All results are obtained using the joint PMLE method. In all cases biases are very small. Average standard errors and RMSE of the estimators of slope parameters are also small. The estimators of the correlation coefficients tend to have larger standard errors suggesting that it is hard to estimate ρ_j with a high degree of precision.

Simulation results for APE^U are presented in Table 5. The bias is computed as a difference between the APE evaluated at the estimated coefficients and APE evaluated at true θ. Standard errors are obtained using the delta method. Similar to the findings for estimated coefficients, probit is the preferred estimation method when errors are not correlated, as it has lower standard errors and small biases. However, when the error correlation is present, the joint PMLE estimator has the smallest computed bias and reasonable standard errors. Table 6 contains simulation results for APE^C. In the Table, the bias tends to be small. However, standard errors are noticeably higher than for APE^U, especially in the case of three unordered groups. The increase in standard errors is not surprising, given that correlation coefficients cannot be estimated precisely.

To study the properties of the estimators when the normality assumption fails, we consider an alternative data generating process, where error u_{itj} has zero-mean chi-square distribution with three degrees of freedom. The variance of u_{itj} is set equal to that in the normal distribution case. The results are presented in Tables 7-12. As seen from the Tables, the computed biases are larger for all estimators of model parameters when the normality assumption is violated. However, the joint PMLE still has less bias than a single-equation probit when errors are correlated, and its bias is reasonably small. Similarly, misspecification of the likelihood function leads to an increase in the bias of APE^U and APE^C estimators, but the increase is minor.

Simulations were also preformed for $N = 500$ and $N = 1,000$. The corresponding results are reported in the Online Appendix. As expected, computed biases, average standard errors, and RMSE tend to decrease as N grows. Otherwise, the results are qualitatively unchanged.

6 Empirical Application

To illustrate the presented theoretical argument with an empirical example, we study gender differences in job satisfaction in different types of occupations. Using Blau and Kahn (2017) as our guide, we consider three types of jobs: (i) "male" professional occupations (professional jobs excluding nurses and K-12 and other non-college teachers), (ii) "female" professional occupations (nurses, K-12 teachers, and other non-college teachers), and (iii) other occupations. The effect of gender may vary by occupation. Moreover, self-selection into a particular occupation type is likely nonrandom and depends on personal preferences. The unobserved factors that influence occupation choice may also affect job satisfaction, which implies that the errors in the main and sorting equations may be correlated, so that the methodology presented in this paper should be helpful. Because there are three unordered groups, we use estimators presented in Section 2.4.

To perform the analysis, we employ data from the National Longitudinal Survey of Youth, 1979 (NLSY79). The initial sample is representative of individuals born between

1957-1964. We utilize 1987-1992 waves of the survey that are characterized by a relatively high response rate (90% or higher). Respondents in supplemental samples, military samples, self-employed, and agricultural workers are excluded. Limiting the data to respondents who participated and were employed in all six waves of the survey results in a balanced panel of 2,874 workers. Observations with missing information on any of the variables used in the analysis and individuals younger than 25 years old were dropped. The final sample includes 2,729 workers, and a total of 14,750 person-year observations. About 22.4% of the respondents transitioned from one occupation type to another at least once during the considered period.

The dependent variable in the main equation is an indicator equal to one if the worker is very satisfied with his or her job, and zero otherwise. Explanatory variables include female indicator, age, race and ethnicity indicators, college education indicator, and graduate education indicator. To control for individual differences in cognitive ability we include the Armed Forces Qualification Test (AFQT) score, which was administered in 1979. We also include two personality variables, locus of control and self-esteem, which are from years 1979 and 1980, respectively. The self-esteem measure (Rosenberg, 1965) is used to assess the degree of approval or disapproval toward oneself, while locus of control (Rotter, 1966) is used to evaluate how strongly the respondent believes he/she can control own life outcomes (internal locus of control) rather than outcomes being determined by fate or luck (external locus of control). In the data, the value of this variable is larger if the person has a more external locus of control (smaller for internal). The AFQT score and personality measures were standardized to have a zero mean and unit variance in the sample. Vector $\bar{z}_i$ includes the individual time means of time-varying covariates – college and graduate education indicators. The time mean of age was omitted from $\bar{z}_i$ to avoid perfect collinearity with year dummies, as age increases by one every year for all respondents. Year indicators are included in all equations.

The determinants of the occupation choice include all variables from the job satisfaction equation and parental occupation at the time when the respondent was 14 years old. Specifically, we include an indicator for father (father figure) working in a “male” professional occupation, an indicator for father (father figure) working in a “female” professional occupation, and similar indicators for the mother (mother figure). Parents employed in other occupations and non-working parents comprise the reference group.[4]

Summary statistics are presented in Table 13. As seen in the Table, the percent of workers who are very satisfied with their job is the highest among those in “female” professional occupations. As expected, the proportion of females is the highest among those who have a “female” professional job and lowest in “male” professional occupations. The percent of minorities is the largest in “other occupations” group. Workers in “male” and “female” professional occupations tend to have more education, higher cognitive ability, higher self-esteem, and a more internal locus of control. Respondents in “female” professional jobs are more likely to have a mother (mother figure) who worked in a “female” professional occupation at the time when the respondent was 14 years old. Similarly, those in “male” professional occupations are more likely to have a father (father figure) who had a “male” professional job.

To obtain main results, the job satisfaction equation was first estimated separately for each occupation type by pooled probit. Subsequently, the same equations were estimated using the two methods described in Section 2.4. Estimated average partial effects and standard errors are presented in Table 14. The first three columns display estimated unconditional APE. As expected, results vary by estimation method. The differences are particularly large if looking at “female” professional occupations. For example, the estimated APE^U of the female indicator is much larger for the two PMLE methods than for probit. After accounting for the nonrandom occupation choice using joint PMLE (third column in Table 14), the predicted probability of being very satisfied is about 10

[4]When indicators for working parents were included, they were never significant and caused multicollinearity problems. Therefore, these variables were excluded from the model.

percentage points higher for women, as compared to men, which is more than double of the probit estimate. On the other hand, the estimated racial difference for this occupation type becomes substantially smaller after accounting for self selection. The dissimilarities between probit and PMLE are not surprising, given that the estimated error correlation for "female" professional occupations is very large (0.952) and highly statistically significant.

The differences between PMLE and probit estimates of APE^U for "male" professional occupations are much smaller, while they are almost non-existent for "other occupations." These similarities can be explained by smaller error correlations, which are insignificant in both cases. Notably, estimation results reveal no gender difference in job satisfaction if working in a "male" professional occupation, but a moderate (four percentage points) significant positive effect of being a woman for "other occupations." Thus, there is substantial heterogeneity in the gender effect by occupation.

The last column in Table 14 reports the APE conditional on being in a given occupation type. The discrepancies between APE^U and APE^C are rather minor for the "male" professional and "other" occupations, which is as expected. However, the corresponding differences for "female" professional jobs (where the error correlation is very high) are sizable. Among individuals who choose to work in such occupations women are only 4.5 percentage points more likely to be very satisfied with their jobs than men, which is roughly half of the effect under the random occupation assignment. Moreover, APE^C of the female indicator is statistically insignificant, indicating no systematic gender differences in job satisfaction for workers in this group. This finding is likely due to the positive error correlation. Both male and female workers who self-select into "female" professional occupations tend to be happier with their jobs, which reduced gender differences in job satisfaction.

Estimated average partial effects in the occupation equations are displayed in Table 15. "Male" professional jobs is the base group. Not surprisingly, women are more likely to be in " female" professional occupations than in the other two occupation types, while higher cognitive ability is positively associated with the probability of being in "male" professional jobs. Importantly, parental occupation at age 14 influences the person's occupation choice later in life. Mother having a "female" professional job has a positive effect on the individual's probability of being in a "female" professional occupation. On the other hand, the probability that the person is in a "male" professional occupation is higher if their father had a "male" or "female" professional job. The effects are statistically significant at the 5% significance level or better.

Going back to cross-occupation comparisons of gender differences in job satisfaction, the obtained results are reasonable. Women often choose "female" professional jobs because such jobs tend to have characteristics valued by women, e.g. social interaction. On the other hand, "male" professional jobs appear to be least likely to offer such intrinsic benefits, which leads to the lower job satisfaction among women and reduced probability of them choosing these jobs.

7 Conclusion

This paper discusses the methodology for estimating heterogeneous effects in static binary response panel data models. In addition to a two-group case, we consider estimating parameters for multiple heterogeneous groups, which may be ordered or unordered. Under the formulated assumptions, the proposed methods produce consistent and $\sqrt{N}$-asymptotically normal estimators of heterogeneous parameters. Simulations show that the methods perform well in finite samples. The computed biases remain small when the correlation between errors in the main and sorting equations increases. The RMSE are also smaller than probit RMSE when error correlations are different from zero.

As an empirical application, we use NLSY79 data to estimate heterogeneous gender differences in job satisfaction by occupation type. We find that accounting for nonrandom occupation choice produces different results as compared to simple probit estimation when error correlation is high. Once the non-random self-selection is accounted for, the predicted gender difference in job satisfaction is about 10 percentage points in "female" professional occupations, where women tend to feel happier than men. In contrast, there is no gender difference for "male" professional occupations.

The proposed methods can be used for estimating heterogeneous effects using cross-section data. It would correspond to a special case with $T = 1$. Obviously, the Mundlak-Chamberlain model of the unobserved effect cannot be used in such a setting. Instead, one would need to include a sufficient set of controls to avoid inconsistencies resulting from an omitted variable problem. Possible venues for future research include considering the estimation of heterogeneous effects in dynamic binary response panel data models with an arbitrary number of groups.

References

Abrevaya, J., and C.M. Dahl, 2008, The Effects of Birth Inputs on Birthweight. *Journal of Business and Economic Statistics* 26, 379-97.

Basu, A., 2014, Estimating Person-Centered Treatment (PeT) Effects Using Instrumental Variables: An Application to Evaluating Prostate Cancer Treatments. *Journal of Applied Econometrics* 29, 671-691.

Berndt, E.R., B.H. Hall, R.E. Hall, and J.A. Hausman, 1974, Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement* 3, 653-666.

Bjorklund, A., and R. Moffitt, 1987, The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *Review of Economics and Statistics* 69(1), 42-49.

Blau, F. D., and L. M. Kahn, 2017, The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature* 55(3), 789–865.

Carrasco, R., 2001, Binary Choice With Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation, *Journal of Business and Economic Statistics* 19(4): 385-94.

Chamberlain, G., 1980, Analysis of Covariance with Qualitative Data, *Review of Economic Studies* 47, 225-238.

Clark, A., 1997. Job satisfaction and gender: why are women so happy at work? *Labour Economics* 4, 341-372.

Crosby, F. J., 1982, *Relative deprivation and working women.* New York: Oxford University Press.

Fairlie, R., 2002, Drug Dealing and Legitimate Self-Employment. *Journal of Labor Economics* 20(3), 538-567.

Goldfeld, S.M. and R.E. Quandt, 1973, *Nonlinear Methods in Econometrics.* Amsterdam: North Holland.

Heckman, J.J., 1979, Sample Selection Bias as a Specification Error. *Econometrica* 47(1), 153-61.

Heckman, J.J., Sergio Urzua and Edward Vytlacil, 2006, Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics* 88(3), 389-432.

Jäckle, R., and O. Himmler, 2010, Health and Wages: Panel Data Estimates Considering Selection and Endogeneity. *Journal of Human Resources* 45(2), 364-406.

Kyriazidou, E., 1997, Estimation of a panel data sample selection model. *Econometrica* 65, 1335-1364.

Lee, L.F., 1978, Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables. *International Economic Review* 19, 415-433.

Maddala G.S. and F. Nelson, 1975, Switching Regression Models with Exogenous and Endogenous Switching. *Proceedings of the American Statistical Association* (Business and Economics Section), 423-426.

Maddala, G.S., 1983, *Limited Dependent Variable and Qualitative Variables in Econometrics*. Cambridge, U.K.: Cambridge University Press.

Manski, C., D. Sandefur, S. McLanahan, and D. Powers, 1992, Alternative Estimates of Family Structure During Adolescence on High School Graduation, *Journal of the American Statistical Association* 87, 25-37.

Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data. *Econometrica* 46(1), 69-85.

Newey, W.K., 2009, Two-step series estimation of sample selection models. *Econometrics Journal* 12, S217-S229.

Newey, W.K. and D. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics*, Volume 4, ed. R.F.Engle and D.McFadden. Amsterdam: North Holland, 2111-2245.

Semykina, A., 2018, Self-Employment among Women: Do Children Matter More Than We Previously Thought? *Journal of Applied Econometrics* 33, 416-434.

Semykina, A. and J.M. Wooldridge, 2018, Binary Response Panel Data Models with Sample Selection and Self Selection. *Journal of Applied Econometrics* 33, 179-197.

Sloane, P. and H. Williams, 2000, Job satisfaction, comparison earnings and gender. *Labour* 14, 473-501.

Vella, F., 1988, Generating Conditional Expectations from Models with Selectivity Bias. *Economics Letters* 28, 97-103.

Wooldridge, J.M., 1995, Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions. *Journal of Econometrics* 68, 115–132.

Table 1: Simulation Results for $J = 2$ ($T = 3$, $N = 300$)

	$\rho_1 = \rho_2 = 0$			$\rho_1 = -0.5$, $\rho_2 = 0.5$		
	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
β_{01}						
Probit	0.018	0.116	0.114	-0.274	0.105	0.295
Joint PMLE	0.003	0.207	0.220	-0.012	0.158	0.163
β_{11}						
Probit	-0.020	0.131	0.132	0.019	0.127	0.134
Joint PMLE	-0.005	0.136	0.147	-0.008	0.123	0.128
β_{02}						
Probit	0.006	0.102	0.098	-0.365	0.098	0.379
Joint PMLE	-0.005	0.209	0.219	-0.011	0.172	0.177
β_{12}						
Probit	0.020	0.127	0.129	-0.030	0.124	0.129
Joint PMLE	0.005	0.133	0.147	0.002	0.120	0.120

Table 2: Simulation Results for $J = 3$, Ordered Groups ($T = 3$, $N = 300$)

	$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$		
	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
β_{01}						
Probit	-0.021	0.147	0.155	-0.615	0.170	0.641
Joint PMLE	-0.005	0.386	0.335	-0.040	0.422	0.361
β_{11}						
Probit	-0.027	0.166	0.171	-0.201	0.180	0.278
Joint PMLE	-0.010	0.191	0.176	-0.034	0.229	0.208
β_{02}						
Probit	0.011	0.114	0.118	0.008	0.121	0.126
Joint PMLE	0.008	0.116	0.117	0.003	0.115	0.115
β_{12}						
Probit	-0.063	0.263	0.287	-0.457	0.309	0.569
Joint PMLE	-0.024	0.302	0.290	-0.048	0.420	0.379
β_{03}						
Probit	-0.004	0.151	0.157	-0.571	0.167	0.598
Joint PMLE	0.001	0.416	0.361	-0.038	0.373	0.370
β_{13}						
Probit	0.054	0.241	0.263	0.269	0.263	0.393
Joint PMLE	0.012	0.283	0.267	0.039	0.327	0.324

Table 3: Simulation Results for $J = 3$, Unordered Groups ($T = 3$, $N = 300$)

	$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$		
	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
β_{01}						
Probit	-0.015	0.106	0.107	0.348	0.099	0.362
Best alt. PMLE	0.100	0.509	0.717	0.215	0.466	0.736
Joint PMLE	0.045	0.418	0.437	0.060	0.386	0.386
β_{11}						
Probit	-0.026	0.165	0.175	-0.074	0.166	0.182
Best alt. PMLE	0.161	0.235	0.268	0.139	0.245	0.258
Joint PMLE	0.042	0.192	0.181	0.028	0.205	0.179
β_{02}						
Probit	0.044	0.182	0.191	0.479	0.211	0.531
Best alt. PMLE	0.010	0.379	0.384	0.034	0.433	0.432
Joint PMLE	0.009	0.370	0.375	0.042	0.416	0.421
β_{12}						
Probit	-0.087	0.263	0.292	-0.224	0.280	0.369
Best alt. PMLE	-0.032	0.284	0.283	-0.042	0.320	0.320
Joint PMLE	-0.034	0.282	0.283	-0.048	0.315	0.317
β_{03}						
Probit	0.033	0.168	0.180	-0.324	0.153	0.363
Best alt. PMLE	0.002	0.340	0.335	0.010	0.280	0.282
Joint PMLE	0.001	0.331	0.327	0.003	0.273	0.280
β_{13}						
Probit	0.070	0.260	0.300	0.070	0.259	0.284
Best alt. PMLE	0.028	0.273	0.298	0.045	0.260	0.272
Joint PMLE	0.029	0.272	0.297	0.044	0.261	0.271

Table 4: Simulation Results for Parameters in Sorting Equations ($T = 3$, $N = 300$)

	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
$J = 2$						
		$\rho_1 = \rho_2 = 0$			$\rho_1 = -0.5$, $\rho_2 = 0.5$	
δ_0	0.004	0.065	0.066	0.002	0.066	0.067
δ_1	0.004	0.086	0.093	0.006	0.086	0.087
δ_2	0.005	0.098	0.115	0.008	0.099	0.096
ρ_1	-0.006	0.276	0.290	0.027	0.234	0.249
ρ_2	-0.001	0.246	0.258	-0.014	0.196	0.210
$J = 3$, ordered						
		$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$	
δ_1	0.001	0.070	0.063	0.001	0.072	0.065
δ_2	0.009	0.084	0.076	0.007	0.088	0.077
ρ_1	0.007	0.322	0.284	-0.020	0.283	0.244
ρ_2	0.003	0.258	0.232	-0.002	0.215	0.191
ρ_3	-0.001	0.368	0.317	0.026	0.275	0.276
$J = 3$, unordered						
		$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$	
δ_{02}	-0.007	0.102	0.106	-0.009	0.102	0.099
δ_{12}	0.004	0.120	0.124	0.007	0.120	0.118
δ_{22}	0.007	0.139	0.139	0.016	0.140	0.139
δ_{03}	-0.007	0.102	0.102	-0.002	0.102	0.104
δ_{13}	-0.008	0.120	0.117	-0.007	0.120	0.116
δ_{23}	-0.011	0.139	0.137	-0.017	0.140	0.139
ρ_1	-0.032	0.550	0.581	-0.079	0.531	0.530
ρ_2	0.011	0.417	0.444	-0.020	0.382	0.389
ρ_3	0.016	0.379	0.377	0.017	0.339	0.353

Estimation was performed using joint PMLE estimator.

Table 5: Simulation Results for APE^U ($T=3$, $N=300$)

	Bias	Avg. Std. Err.	Bias	Avg. Std. Err.
$J=2$				
	$\rho_1=\rho_2=0$		$\rho_1=-0.5,\ \rho_2=0.5$	
APE_1^U: Probit	-0.0005	0.0240	-0.0190	0.0244
APE_1^U: Joint PMLE	0.0001	0.0301	0.0003	0.0312
APE_2^U: Probit	-0.0014	0.0254	0.0096	0.0227
APE_2^U: Joint PMLE	-0.0056	0.0316	-0.0013	0.0285
$J=3$, ordered				
	$\rho_1=\rho_2=\rho_3=0$		$\rho_1=\rho_2=0.5,\ \rho_3=-0.5$	
APE_1^U: Probit	0.0010	0.0309	0.0244	0.0302
APE_1^U: Joint PMLE	0.0054	0.0357	0.0045	0.0370
APE_2^U, Probit	-0.0005	0.0424	0.0096	0.0436
APE_2^U: Joint PMLE	0.0008	0.0365	0.0016	0.0448
APE_3^U: Probit	-0.0001	0.0492	-0.0183	0.0597
APE_3^U: Joint PMLE	-0.0078	0.0592	-0.0046	0.0850
$J=3$, unordered				
	$\rho_1=\rho_2=\rho_3=0$		$\rho_1=\rho_2=0.5,\ \rho_3=-0.5$	
APE_1^U: Probit	0.0012	0.0305	-0.0212	0.0310
APE_1^U: Best alt. PMLE	0.0445	0.0386	0.0296	0.0419
APE_1^U: Joint PMLE	0.0139	0.0478	0.0085	0.0461
APE_2^U: Probit	0.0000	0.0468	0.0200	0.0481
APE_2^U: Best alt. PMLE	0.0035	0.0422	0.0043	0.0476
APE_2^U: Joint PMLE	0.0043	0.0509	0.0041	0.0473
APE_3^U: Probit	-0.0008	0.0352	0.0206	0.0361
APE_3^U: Best alt. PMLE	-0.0056	0.0411	-0.0059	0.0462
APE_3^U: Joint PMLEv	-0.0057	0.0501	-0.0061	0.0460

Table 6: Simulation Results for APE^C ($T=3$, $N=300$)

	Bias	Avg. Std. Err.	Bias	Avg. Std. Err.
$J=2$				
	$\rho_1=\rho_2=0$		$\rho_1=-0.5,\ \rho_2=0.5$	
APE_1^C: Joint PMLE	-0.0010	0.0338	0.0011	0.0360
APE_2^C: Joint PMLE	-0.0020	0.0353	0.0011	0.0255
$J=3$, ordered				
	$\rho_1=\rho_2=\rho_3=0$		$\rho_1=\rho_2=0.5,\ \rho_3=-0.5$	
APE_1^C: Joint PMLE	0.0012	0.0488	-0.0020	0.0295
APE_2^C: Joint PMLE	-0.0002	0.0470	-0.0005	0.0329
APE_3^C: Joint PMLE	-0.0008	0.0599	0.0022	0.1175
$J=3$, unordered				
	$\rho_1=\rho_2=\rho_3=0$		$\rho_1=\rho_2=0.5,\ \rho_3=-0.5$	
APE_1^C: Joint PMLE	0.0470	0.1370	0.0653	0.0739
APE_2^C: Joint PMLE	-0.0002	0.1565	-0.0044	0.0765
APE_3^C: Joint PMLE	0.0138	0.0515	0.0041	0.1050

Table 7: Simulation Results for Chi-Square Error Distribution, $J = 2$ ($T = 3$, $N = 300$)

	$\rho_1 = \rho_2 = 0$			$\rho_1 = -0.5$, $\rho_2 = 0.5$		
	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
β_{01}						
Probit	0.057	0.120	0.138	-0.274	0.107	0.293
Joint PMLE	0.016	0.233	0.237	0.001	0.161	0.153
β_{11}						
Probit	-0.092	0.137	0.167	-0.017	0.130	0.133
Joint PMLE	-0.070	0.146	0.162	-0.044	0.126	0.136
β_{02}						
Probit	-0.002	0.104	0.108	-0.398	0.099	0.410
Joint PMLE	-0.011	0.215	0.217	-0.036	0.168	0.172
β_{12}						
Probit	0.105	0.133	0.177	0.014	0.128	0.132
Joint PMLE	0.092	0.140	0.174	0.044	0.123	0.133

Table 8: Simulation Results for Chi-Square Error Distribution, $J = 3$, Ordered Groups ($T = 3$, $N = 300$)

	$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$		
	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
β_{01}						
Probit	-0.077	0.154	0.174	-0.664	0.179	0.690
Joint PMLE	-0.093	0.898	0.318	-0.199	0.566	0.413
β_{11}						
Probit	-0.083	0.172	0.198	-0.217	0.185	0.294
Joint PMLE	-0.074	0.503	0.199	-0.086	0.307	0.230
β_{02}						
Probit	-0.002	0.113	0.117	0.008	0.120	0.124
Joint PMLE	-0.005	0.132	0.115	0.006	0.140	0.115
β_{12}						
Probit	-0.108	0.276	0.312	-0.505	0.322	0.615
Joint PMLE	-0.068	0.367	0.309	-0.126	0.507	0.410
β_{03}						
Probit	-0.032	0.157	0.171	-0.592	0.174	0.618
Joint PMLE	-0.051	0.450	0.336	-0.163	0.421	0.384
β_{13}						
Probit	0.145	0.260	0.313	0.327	0.276	0.447
Joint PMLE	0.114	0.568	0.301	0.149	0.365	0.354

Table 9: Simulation Results for Chi-Square Error Distribution, $J = 3$, Unordered Groups ($T = 3$, $N = 300$)

	$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$		
	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
β_{01}						
Probit	-0.051	0.109	0.122	0.327	0.101	0.342
Best alt. PMLE	-0.038	0.456	0.646	0.146	0.433	0.691
Joint PMLE	-0.100	0.392	0.436	-0.012	0.360	0.406
β_{11}						
Probit	-0.038	0.168	0.179	-0.107	0.170	0.205
Best alt. PMLE	0.141	0.215	0.270	0.119	0.228	0.274
Joint PMLE	0.038	0.195	0.191	0.023	0.192	0.199
β_{02}						
Probit	0.037	0.179	0.191	0.481	0.208	0.529
Best alt. PMLE	-0.019	0.411	0.420	-0.018	0.449	0.469
Joint PMLE	-0.024	0.390	0.436	-0.029	0.410	0.482
β_{12}						
Probit	-0.088	0.272	0.302	-0.234	0.287	0.389
Best alt. PMLE	-0.021	0.299	0.301	-0.021	0.330	0.358
Joint PMLE	-0.009	0.293	0.318	-0.012	0.311	0.360
β_{03}						
Probit	0.047	0.167	0.179	-0.320	0.152	0.358
Best alt. PMLE	0.018	0.351	0.369	0.002	0.271	0.280
Joint PMLE	0.008	0.338	0.374	0.003	0.262	0.282
β_{13}						
Probit	0.141	0.273	0.323	0.114	0.269	0.303
Best alt. PMLE	0.089	0.292	0.309	0.086	0.269	0.283
Joint PMLE	0.083	0.286	0.315	0.076	0.265	0.289

Table 10: Simulation Results for Parameters in Sorting Equations, Chi-Square Error Distribution ($T = 3$, $N = 300$)

	Bias	Avg. Std. Err.	RMSE	Bias	Avg. Std. Err.	RMSE
$J = 2$						
		$\rho_1 = \rho_2 = 0$			$\rho_1 = -0.5,\ \rho_2 = 0.5$	
δ_0	-0.002	0.065	0.065	0.001	0.065	0.067
δ_1	0.006	0.086	0.085	0.005	0.085	0.088
δ_2	0.011	0.099	0.100	0.013	0.098	0.101
ρ_1	0.032	0.308	0.327	0.007	0.238	0.229
ρ_2	-0.003	0.253	0.258	-0.005	0.188	0.196
$J = 3$, ordered						
		$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5,\ \rho_3 = -0.5$	
δ_1	0.005	0.082	0.066	0.006	0.106	0.066
δ_2	0.006	0.127	0.077	0.012	0.137	0.074
ρ_1	-0.023	0.667	0.257	-0.104	0.402	0.272
ρ_2	0.010	0.335	0.227	-0.034	0.350	0.204
ρ_3	0.021	0.459	0.283	0.106	0.353	0.284
$J = 3$, unordered						
		$\rho_1 = \rho_2 = \rho_3 = 0$			$\rho_1 = \rho_2 = 0.5,\ \rho_3 = -0.5$	
δ_{02}	-0.002	0.100	0.114	0.006	0.091	0.117
δ_{12}	-0.003	0.120	0.121	0.000	0.108	0.123
δ_{22}	0.002	0.135	0.150	-0.009	0.122	0.151
δ_{03}	-0.005	0.100	0.107	0.007	0.091	0.112
δ_{13}	-0.008	0.120	0.116	0.004	0.108	0.119
δ_{23}	-0.006	0.135	0.146	-0.002	0.122	0.146
ρ_1	0.128	0.539	0.611	0.000	0.486	0.564
ρ_2	0.029	0.439	0.503	0.038	0.371	0.479
ρ_3	0.015	0.386	0.435	0.021	0.323	0.372

Estimation was performed using joint PMLE estimator.

Table 11: Simulation Results for Chi-Square Error Distribution, APE^U ($T = 3$, $N = 300$)

	Bias	Avg. Std. Err.	Bias	Avg. Std. Err.
$J = 2$				
	$\rho_1 = \rho_2 = 0$		$\rho_1 = -0.5$, $\rho_2 = 0.5$	
APE_1^U: Probit	-0.0059	0.0246	-0.0253	0.0248
APE_1^U: Joint PMLE	-0.0063	0.0309	-0.0045	0.0317
APE_2^U: Probit	0.0160	0.0266	0.0168	0.0230
APE_2^U: Joint PMLE	0.0113	0.0326	0.0066	0.0289
$J = 3$, ordered				
	$\rho_1 = \rho_2 = \rho_3 = 0$		$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$	
APE_1^U: Probit	-0.0026	0.0312	0.0268	0.0308
APE_1^U: Joint PMLE	0.0013	0.0366	0.0086	0.0379
APE_2^U: Probit	-0.0004	0.0443	0.0085	0.0449
APE_2^U: Joint PMLE	0.0007	0.0911	0.0009	0.0550
APE_3^U: Probit	0.0016	0.0504	-0.0200	0.0686
APE_3^U: Joint PMLE	-0.0044	0.1029	-0.0069	0.0625
$J = 3$, unordered				
	$\rho_1 = \rho_2 = \rho_3 = 0$		$\rho_1 = \rho_2 = 0.5$, $\rho_3 = -0.5$	
APE_1^U: Probit	0.0004	0.0312	-0.0230	0.0312
APE_1^U: Best alt. PMLE	0.0383	0.0399	0.0315	0.0426
APE_1^U: Joint PMLE	0.0182	0.0496	0.0125	0.0478
APE_2^U: Probit	-0.0005	0.0422	0.0193	0.0447
APE_2^U: Best alt. PMLE	0.0016	0.0447	0.0009	0.0493
APE_2^U: Joint PMLE	0.0029	0.0535	0.0009	0.0491
APE_3^U: Probit	0.0031	0.0369	0.0237	0.0360
APE_3^U: Best alt. PMLE	-0.0013	0.0437	0.0004	0.0464
APE_3^U: Joint PMLE	-0.0015	0.0525	-0.0007	0.0484

Table 12: Simulation Results for Chi-Square Error Distribution, APE^C ($T = 3$, $N = 300$)

	Bias	Avg. Std. Err.	Bias	Avg. Std. Err.
$J = 2$				
	$\rho_1 = \rho_2 = 0$		$\rho_1 = -0.5,\ \rho_2 = 0.5$	
APE_1^C: Joint PMLE	-0.0058	0.0357	-0.0059	0.0398
APE_2^C: Joint PMLE	0.0153	0.0478	0.0075	0.0249
$J = 3$, ordered				
	$\rho_1 = \rho_2 = \rho_3 = 0$		$\rho_1 = \rho_2 = 0.5,\ \rho_3 = -0.5$	
APE_1^C: Joint PMLE	-0.0028	0.0733	-0.0010	0.0393
APE_2^C: Joint PMLE	-0.0002	0.0487	-0.0012	0.0393
APE_3^C: Joint PMLE	0.0012	0.0933	0.0033	0.0797
$J = 3$, unordered				
	$\rho_1 = \rho_2 = \rho_3 = 0$		$\rho_1 = \rho_2 = 0.5,\ \rho_3 = -0.5$	
APE_1^C: Joint PMLE	0.0535	0.0515	0.0780	0.0786
APE_2^C: Joint PMLE	0.0000	0.7188	-0.0042	0.3031
APE_3^C: Joint PMLE	0.0161	0.0530	0.0078	0.1078

Table 13: Summary Statistics

Variable	"Male" prof. occupations	"Female" prof. occupations	Other occupations
Very satisfied with the job (%)	51.23	64.25	42.16
Female (%)	44.22	85.38	48.21
Black (%)	5.70	8.63	12.05
Hispanic (%)	5.39	4.75	6.50
Age	28.98	29.26	28.83
	(2.46)	(2.40)	(2.48)
College education (%)	54.70	59.88	38.62
Graduate education (%)	34.36	32.13	5.72
AFQT score	0.73	0.49	-0.10
	(0.84)	(0.82)	(0.98)
Self-esteem	0.31	0.14	-0.04
	(0.96)	(0.95)	(1.00)
Locus of control	-0.20	-0.27	0.02
	(1.05)	(1.03)	(0.99)
Mother in "male" prof. occupation (%)	5.24	4.50	1.79
Mother in "female" prof. occupation (%)	10.48	14.88	3.83
Father in "male" prof. occupation (%)	19.26	10.25	6.64
Father in "female" prof. occupation (%)	3.16	2.75	1.17
Number of observations	1,298	800	12,652

Table 14: Estimated Partial Effects for Probability of Being Very Satisfied with the Job

	Probit APE^U	Best alt. PMLE APE^U	Joint PMLE APE^U	Joint PMLE APE^C
		"Male" professional occupations		
Female	-0.018	-0.026	0.007	-0.037
	(0.038)	(0.027)	(0.120)	(0.080)
Black	0.047	0.034	0.038	0.030
	(0.083)	(0.056)	(0.056)	(0.032)
Hispanic	0.106	0.104*	0.064	0.054
	(0.090)	(0.059)	(0.103)	(0.052)
Age	-0.009	-0.009***	-0.005	-0.006
	(0.009)	(0.003)	(0.020)	(0.017)
College education	0.382	0.356	0.288	0.244
	(0.338)	(0.305)	(0.294)	(0.164)
Graduate education	0.348	0.310	0.271	0.227
	(0.359)	(0.317)	(0.296)	(0.177)
AFQT score	-0.025	-0.001	-0.040	-0.017
	(0.027)	(0.036)	(0.101)	(0.060)
Self-esteem	0.037*	0.033**	0.024	0.021
	(0.020)	(0.013)	(0.038)	(0.022)
Locus of control	0.003	0.006	-0.003	0.002
	(0.020)	(0.013)	(0.029)	(0.017)
ρ		0.290	-0.469	
		(0.443)	(2.449)	
		"Female" professional occupations		
Female	0.042	0.072***	0.097**	0.045
	(0.072)	(0.018)	(0.047)	(0.098)
Black	-0.161*	-0.031***	-0.049**	-0.173
	(0.093)	(0.010)	(0.020)	(14.367)
Hispanic	-0.039	-0.007	-0.014	-0.026
	(0.076)	(0.016)	(0.025)	(1.296)
Age	-0.002	0.003	0.003	-0.003
	(0.013)	(0.004)	(0.007)	(0.014)
College education	0.003	0.008	0.017	0.033
	(0.301)	(0.069)	(0.102)	(0.267)
Graduate education	-0.150	-0.011	-0.016	-0.117
	(0.327)	(0.058)	(0.083)	(0.301)
AFQT score	-0.042	-0.009*	-0.017	-0.050
	(0.040)	(0.005)	(0.012)	(0.036)
Self-esteem	0.015	0.000	0.001	0.017
	(0.027)	(0.004)	(0.009)	(0.026)
Locus of control	-0.062**	-0.015***	-0.024**	-0.070**
	(0.026)	(0.004)	(0.011)	(0.029)
ρ		0.831***	0.952***	
		(0.110)	(0.289)	

Standard errors in parentheses; *** significant at the 1% level, ** at the 5% level, * at the 10% level.
All equations include year indicators and individual time means for college and graduate education.

Table 14. *Continued.*

	Probit APE^U	Best alt. PMLE APE^U	Joint PMLE APE^U	Joint PMLE APE^C
		Other occupations		
Female	0.036***	0.038***	0.040**	0.026
	(0.014)	(0.011)	(0.017)	(0.017)
Black	-0.051**	-0.051***	-0.052**	-0.048**
	(0.023)	(0.015)	(0.023)	(0.023)
Hispanic	0.040	0.040**	0.041	0.033
	(0.029)	(0.018)	(0.029)	(0.030)
Age	-0.002	-0.002	-0.002	-0.002
	(0.003)	(0.002)	(0.003)	(0.003)
College education	-0.029	-0.028	-0.027	-0.033
	(0.049)	(0.054)	(0.050)	(0.051)
Graduate education	0.008	0.009	0.011	0.004
	(0.070)	(0.084)	(0.071)	(0.072)
AFQT score	-0.003	-0.001	0.000	-0.009
	(0.009)	(0.007)	(0.011)	(0.011)
Self-esteem	0.049***	0.049***	0.049***	0.047***
	(0.007)	(0.005)	(0.008)	(0.008)
Locus of control	-0.007	-0.007	-0.007	-0.008
	(0.007)	(0.005)	(0.007)	(0.008)
ρ		-0.078	-0.185	
		(0.251)	(0.432)	

Standard errors in parentheses; *** significant at the 1% level, ** at the 5% level, * at the 10% level.
All equations include year indicators and individual time means for college and graduate education.

Table 15: Estimated Average Partial Effects in Occupation Equations

	"Female" professional occupations	Other occupations
Female	0.381***	0.001
	(0.046)	(0.007)
Black	-0.027	0.002
	(0.059)	(0.013)
Hispanic	-0.072	-0.023
	(0.069)	(0.018)
Age	0.021*	0.001
	(0.012)	(0.001)
College education	-0.026	-0.020
	(0.094)	(0.026)
Graduate education	0.029	-0.013
	(0.115)	(0.026)
AFQT score	-0.082***	-0.029***
	(0.022)	(0.006)
Self-esteem	-0.024	-0.003
	(0.017)	(0.003)
Locus of control	-0.026	-0.005
	(0.018)	(0.004)
Mother in "male" professional occupation	0.027	-0.010
	(0.037)	(0.007)
Mother in "female" professional occupation	0.097**	-0.032**
	(0.039)	(0.009)
Father in "male" professional occupation	-0.078**	-0.017**
	(0.033)	(0.008)
Father in "female" professional occupation	-0.065**	-0.045***
	(0.033)	(0.010)

Standard errors in parentheses; *** significant at the 1% level, ** at the 5% level, * at the 10% level. "Male" professional occupations is the base group. All equations include year indicators and individual time means for college and graduate education.