Econometric Methods for Multiple Fractional Response Variables with a Binary Endogenous Covariate: An Application to Time-Use Data

Benjamin I. Miller*

Department of Economics, Michigan State University, East Lansing, MI 48824-1038, United States

Abstract

I develop a model and an estimator for a panel data setting with multiple fractional response variables with a binary endogenous covariate. I develop a two-step technique to obtain consistent estimates of the average partial effects. Then, I provide a variable addition test for endogeneity. I demonstrate using simulations that if the chosen conditional mean function is incorrect, it is still possible to obtain estimates of the average partial effects that are close to the true values. Data from the NLSY97 survey is used to estimate the average partial effect of marriage on how individuals allocate their time within a year.

Keywords— Fractional response, Instrumental variables, Multiple simultaneous equation models, Panel data

Acknowledgements

The author thanks Kyoo-Il Kim, John Mullahy, Stephen Woodbury, and Jeffrey Wooldridge for helpful comments and suggestions. I have no conflicts of interest to report.

*Corresponding author: tel 1+954 873 8091, email address: mill3104@msu.edu

1 Introduction

Fractional response variables are often treated as having a linear conditional mean. This can be done for the purpose of demonstrating that there is a strong association between the fractional random variable and some covariates of interest. The largest contrast to this would be to take a structural approach and try to derive the exact form of the conditional mean function. The decision to use a fractional response model can serve as a point between a linear approximation and a structural approach. When there is a single fractional response variable of interest, the use of a nonlinear conditional mean function can serve to provide a closer approximation to the true relationship between the covariate and the conditional mean compared to a linear function. When there are multiple fractional response variables of interest and those variables represent shares in a bundle, then the linear conditional mean function no longer reflects the choices of an agent. If a structural approach is to be taken, then the structural assumptions that might have reflected an agent's decision problem in a single time period no longer hold over multiple periods. There could be heterogeneity amongst agents that influences their decisions. More assumptions or a more complex structural approach might be needed to identify the parameters of interest.

In this paper, I develop a reduced form estimator of the average partial effects for a fixed number multiple fractional responses. First, I expand upon Mullahy (2015), which is an extension of Papke and Wooldridge (2008), to develop a panel data estimator with a binary endogenous explanatory variable (EEV). This is done using a quasi-maxiumum likelihood approach, combiing this with a Mundlack-Chamberlain device, and then integrating out the unobserved heterogeneity in an approach that is similar to Heckman (1979). This paper is in the spirit of Nam (2013), which handles the case of multiple fractional responses in a panel data setting with a continuous endogenous covariate.

Then, I show the asymptotic properties of the estimates of the average partial effect. Although it may not be possible to recover the unscaled coefficients on the covariates, it is possible to estimate and performance inference on the average partial effects.

After this, I develop a variable addition test (VAT) for endogeneity. This test is based upon a simpler version of the VAT found in Lin and Wooldridge (2017). The methods in this paper can be computationally intense to implement. Without endogenous covariates, the method of Mullahy (2015) would be applied. The test allows for detection of endogeneity across all of the fractional response variables.

I also show that if one of the key identification assumptions fails, it is still possible to obtain estimates of the average partial effects (APE) that are reasonably close to the true values. I break this down into several cases, looking at when the distribution of the unobservables are incorrectly specified, thus causing the conditional mean to be incorrectly specified. The cases include when the unobservable variables are asymmetric.

Finally, as an application of my estimator I use data from the 1997 National Longitudinal Survey of Youth (NLSY97). This is done to estimate how changes in a binary endogenous variable affect the fraction of an individual's time within a year devoted to work, sleep, and leisure. The marital status of each survey participant is reversible, and the religiosity of the parent is used as an instrument coupled with controls that might be correlated with parental religiosity. This specific data and the problem of estimating how marital status affects individuals' use of their time is explored within this paper, while demonstrating how the fractional nature of the dependent variable can be exploited.

The paper is organized as follows. Section 2 presents the model and identification assumptions. Section 3 presents the proposed method for obtaining consistent estimates of the average partial effects. Section 4 presents the average partial effects, as well as their asymptotic distribution. Section 5 presents a VAT for endogeneity. Under the null hypothesis of this test there is no binary EEV, so that the parameters can be estimated using a standard multinomial logit QMLE framework. Section 6 presents simulation results under model misspecification. Under the model misspecification, the multinomial logit specification is incorrect, yet the results show that proposed method will still provide a good approximation of the average partial effects. Section 7 contains the empirical application of this estimator to the 1997 National Longitudinal Survey of Youth. Section

8 concludes the paper and considers alternatives to the estimator presented in this paper.

2 Model and Assumptions

I assume that I have a panel of data consisting of a random sample of N subjects across T time periods with L fractional dependent outcomes, with each dependent outcome denoted as y_{itl}. I also assume strict exogeneity of the structural conditional mean. x_{it} is a 1x K vector of covariates, separate from the 1 x M vector of instruments z_{it}. t_{it} is the binary EEV. e_{it} represents omitted variables that change over time and with each subject i. The structural conditional mean is,

$$E[y_{itl}|\mathbf{x}_i, t_i, \mathbf{z}_i, c_i, e_i] = E[y_{itl}|\mathbf{x}_{it}, t_{it}, c_i, e_{it}] = G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it}) \tag{1}$$

where

$$t_{it} = 1[\mathbf{z}_{it}\delta_z + \mathbf{x}_{it}\delta_x + u_{it} \geq 0] \tag{2}$$

$$c_i = \mathbf{h}_i\pi + a_i \tag{3}$$

$$\mathbf{h}_i = (\bar{\mathbf{x}}_i\ \bar{\mathbf{z}}_i\ \bar{\mathbf{t}}_i) \tag{4}$$

$$0 \leq G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it}) \leq 1 \tag{5}$$

$$r_{itl} := \lambda_l a_i + e_{it} \tag{6}$$

$$\sum_{l=1}^{L} G_{itl} = 1 \tag{7}$$

and ξ_{tl} is a time-varying intercept, where u_{it} is independent of x_{it} (and z_{it}). This is denoted by $u_{it} \perp\!\!\!\perp x_{it}, z_{it}$. G is what I call the structural mean function. It is unknown to the researcher. It is assumed that the distribution of $u_{it}, D(u_{it})$, is known to the researcher. For example, the researcher would know that $u_{it} \sim Normal(0,1)$, and so $\hat{\delta}$ would be obtained from a pooled probit MLE. Equation (3) represents the use of the Chamberlain (1980) device to write the correlated random effect c_i as a projection onto the time-averaged values of the explanatory variables and the instruments. $\mathbf{h}_i$ is the vector of these time-averaged values, where $\bar{\mathbf{x}}_i = \sum_{t=1}^{T} \frac{x_{it}}{T}$, $\bar{\mathbf{z}}_i = \sum_{t=1}^{T} \frac{z_{it}}{T}$, and $\bar{\mathbf{t}}_i = \sum_{t=1}^{T} \frac{t_{it}}{T}$. Equation (5) represents the restriction that the value of the conditional mean for any values of x_{it}, t_{it}, c_i, and e_{it} must be between zero and one, inclusive. Equation (7) represents the constraint that since each y_{itl} is bounded between zero and one and must sum to one, then the structural mean values must sum to one for each time period and subject. This constraint rules out the application of a probit model when $L \geq 3$, since there is no guarantee that the choice of a probit conditional mean function would satisfy the constraint. In Equation (6), r_{itl} is being defined as the sum of λ_l multiplied by the error from the projection of c_i on h_i plus e_{it}. It is important to note that though is tempting to write y_{itl} as equal to $G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it})$, this may rule out the case in which y_{itl} takes the values 0 or 1. It should also be noted that, in contrast to Becker (2014), the correlated random effects have different coefficients across l. This is similar to how previous approaches have handled a single source of heterogeneity in multinomial logit model (see section 16.2.4 of Wooldridge (2010)).

Now, let,

$$r_{itl} = \zeta_l u_{it} + v_{it} \tag{8}$$

In effect, I am noting that r_{itl} can be written as a linear function of u_{it} and v_{it}, where v_{it} is the error arising from regressing r_{itl} on u_{it}.This extends naturally from the model. Since the correlated random effects approach eliminates inconsistency from the explanatory variables, then any endogeneity that arises must be through r_{itl} due to u_{it}. Then, $v_{it} \perp\!\!\!\perp x_{it}, z_{it}, u_{it}$. I am implying that if the unobservables are projected onto the random variable that is the source of the endogeneity, then whatever remains should be random noise. I am also going to assume that,

$$E(y_{itl}|\mathbf{h}_i, \mathbf{x}_i, \mathbf{z}_i, t_i, u_i) = E(y_{itl}|\mathbf{h}_i, \mathbf{x}_{it}, t_{it}, u_{it}) = \Lambda(d^v_{itl} + u_{it}\zeta^v_l) \tag{9}$$

where

$$d_{itl} \coloneqq \xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l \mathbf{h}_i \pi \tag{10}$$

,

$$d^v_{itl} \coloneqq \xi^v_{tl} + \mathbf{x}_{it}\beta^v_l + t_{it}\alpha^v_l + \lambda^v_l \mathbf{h}_i \pi^v \tag{11}$$

,and The multinomial logit function that is of interest here is,

$$\Lambda(d^v_{itl} + u_{it}\zeta^v_l) = \frac{e^{d^v_{itl} + u_{it}\zeta^v_l}}{(1 + \sum_{k=1}^{L-1} e^{(d^v_{itk} + \zeta^v u_{it})})} \tag{12}$$

Equation (11) represents the scaled valued of d_{itl}.Equation (9) represents the assumption that once v_{it} is integrated out, then the conditional mean function is equal to the multinomial logit conditional mean function, and the values of the parameters are related, though it is unknown exactly how, to the original structural parameters. The first equality of equation (9) represents the strict exogeneity assumption. The original structural parameters cannot be identified, but these transformed parameters can be used to estimate the APEs. While I have not made any direct assumptions about the distribution of v_{it}, I have made a somewhat strong assumption, though one that is not out of place in the econometrics literature. In effect, I am placing a restriction upon the combination of G and the random variable v_{it} such that (9) holds; however, it is not known what restrictions must be placed upon G and v_{it} so that (9) is true. It could also be that the original structural model is incompatible with (9), but (9) is assumed to hold for the sake of convenience. This is analogous to assumptions in the generalized estimation equations (GEE) literature, where when heterogeneity is averaged out, a convenient functional form is chosen as in Zeger, Liang, and Albert (1988). Petrin and Train (2010) apply this approach in the context of consumer choice by dividing the structural error in the consumer utility function into two parts. Distributional assumptions are made on each part in order to form a mixed logit conditional mean function. A version of this assumption is guaranteed to hold for any fractional variable, and the method that I am proposing can still be applied. As noted by Mullahy (2015) from Woodland (1979), it is always the case for fractional response variables that the conditional mean function should have a form that is based upon an underlying Dirichlet random variable. Then the conditional mean has the form,

$$E[y_{itl}|x_{it}, z_{it}, t_{it}, u_{it}] = \frac{z_l(x_{it}, z_{it}, t_{it}, u_{it}) - z_L(x_{it}, z_{it}, t_{it}, u_{it})}{1 + \sum_{l=1}^{L-1} z_l(x_{it}, z_{it}, t_{it}, u_{it}) - z_L(x_{it}, z_{it}, t_{it}, u_{it})} \tag{13}$$

When (12) is combined with (13), this is equivalent to setting $z_l(x_{it}, z_{it}, t_{it}, u_{it}) = e^{d^v_{itl}+u_{it}\zeta^v_l}$.

The previous assumptions are summarized as follows:

Assumption ID.1.

$$\begin{aligned} E[y_{itl}|\mathbf{x}_i, t_i, \mathbf{z}_i, c_i, e_i] = E[y_{itl}|\mathbf{x}_{it}, t_{it}, c_i, e_{it}] &= G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it}) \\ t_{it} &= 1[\mathbf{z}_{it}\delta_z + \mathbf{x}_{it}\delta_x + u_{it} \geq 0] \\ c_i &= \mathbf{h}_i\pi + a_i \end{aligned}$$

Assumption ID.2.

$$\begin{aligned} &0 \leq G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it}) \leq 1 \\ &\sum_{l=1}^{L} G_{itl} = 1 \end{aligned}$$

Assumption ID.3. $E(y_{itl}|\mathbf{h}_i, \mathbf{x}_i, \mathbf{z}_i, t_i, u_i) = E(y_{itl}|\mathbf{h}_i, \mathbf{x}_{it}, t_{it}, u_{it}) = \Lambda(d^v_{itl} + u_{it}\zeta^v_l)$

Assumption ID.4. $u_{it} \perp\!\!\!\perp x_{it}, z_{it}$, where $D(u_{it})$ is known.

Assumption ID.5. $E[e_{it}|x_{it}] = 0$, $E[e_{it}|z_{it}] = 0$, and $cov(z_{it}, t_{it} \neq 0)$.

Though Assumption ID.3 is not out of place in the econometrics literature, it can be tested. For example, a choice could be made between $e^{d^v_{itl}+u_{it}\zeta^v_l}$ and $z_l(x_{it}, z_{it}, t_{it}, u_{it})$. Then, a test for misspecified moments based upon Rivers and Vuong (2002) can be applied. This is important because equation 12 always holds, so Assumption ID.3 can be tested using the Rivers and Vuong test, and an alternative model based upon such a test could be provided if the assumption does not hold.

Now, given the previous assumptions of the model the following is obtained:

$$E(y_{itl}|\mathbf{h}_i, \mathbf{x}_{it}, \mathbf{z}_{it}, t_{it} = 1) = \int_{-q_{it}}^{\infty} \Lambda(d_{itl}^v + u_{it}\zeta_l^v) f(u_{it}) du_{it} \tag{14}$$

$$E(y_{itl}|\mathbf{h}_i, \mathbf{x}_{it}, \mathbf{z}_{it}, t_{it} = 0) = \int_{-\infty}^{-q_{it}} \Lambda(d_{itl}^v + u_{it}\zeta_l^v) f(u_{it}) du_{it} \tag{15}$$

where $q_{it} = \mathbf{z}_{it}\delta_z^\sigma + \mathbf{x}_{it}\delta_x^\sigma$. δ_x^σ and δ_z^σ denote the scaled coefficients of $\mathbf{x}$ and $\mathbf{z}$ in the Bernoulli MLE problem, where, for a given parametric assumption on u_{it}, $P(t_{it} = 1|\mathbf{x}_{it}, \mathbf{z}_{it}) = g(\mathbf{x}_{it}, \mathbf{z}_{it}, ; \delta_x^\sigma, \delta_z^\sigma)$.While u_{it} is not independent of t_{it}, the independence assumption on u_{it} with respect to x_{it} and z_{it} implies that the observed value of t_{it} only communicates information about the support of $u_{it}|t_{it} = t, x_{it}, z_{it}$. For example, if u_{it} is normal and independent of $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$, then $P(t_{it} = 1|\mathbf{x}_{it}, \mathbf{z}_{it}) = \Phi(\mathbf{x}_{it}\delta_z^\sigma + \mathbf{z}_{it}^\sigma)$.

3 Estimation Method

The estimator will be derived from a QMLE problem by pooling across time for each subject. The serial dependence across $\{y_{itl}\}$ is unrestricted. Based upon (14) and (15), the following QMLE problem is,

$$\begin{aligned}
&\max_{\beta^v \in \mathbb{R}^{K\times L}, \pi \in \mathbb{R}^{K+M+1}, \alpha_l^v \in \mathbb{R}, \lambda^v \in \mathbb{R}^L, \xi^v \in \mathbb{R}^{T\times L}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Bigg[t_{it} \left(\sum_{l=1}^{L-1} y_{itl} log \int_{-q_{it}}^{\infty} \Lambda(d_{itl}^v + u_{it}\zeta_l^v) f(u_{it}) du_{it} \right) \\
&+ t_{it} \left(1 - \sum_{l=1}^{L-1} y_{itl} \right) log \int_{-q_{it}}^{\infty} (1 - \sum_{l=1}^{L-1} \Lambda(d_{itl}^v + u_{it}\zeta_l^v)) f(u_{it}) du_{it} \\
&+ (1 - t_{it}) \left(\sum_{l=1}^{L-1} y_{itl} log \int_{-\infty}^{-q_{it}} \Lambda(d_{itl}^v + u_{it}\zeta_l^v) f(u_{it}) du_{it} \right) \\
&+ (1 - t_{it}) \left(1 - \sum_{l=1}^{L-1} y_{itl} \right) log \int_{-\infty}^{-q_{it}} (1 - \sum_{l=1}^{L-1} \Lambda(d_{itl}^v + u_{it}\zeta_l^v)) f(u_{it}) du_{it}) \Bigg]
\end{aligned}$$

The maximization problem represents the maximization of the likelihood of a multinomial random variable, where choice l is made with probability $\int_{-q_{it}}^{\infty} \Lambda(d_{itl}^v + u_{it}\zeta_l^v) f(u_{it}) du_{it}$ or $\int_{-\infty}^{-q_{it}} \Lambda(d_{itl}^v +$

$u_{it}\zeta_l^v)f(u_{it})du_{it}$, given the covariates and the observed value of t_{it}. The above objective function handles the endogenous switching problem that arises from the endogenous binary variable. This method is applied by Wooldridge (2014) in the context of a probit conditional mean function when there exists two fractional dependent variables and a binary endogenous variable.

A just-identified GMM system can be set up using the score functions based upon the QMLE problem. Let,

$$\theta = (\xi_{tl}^{v0}, \beta_l^{v0}, \alpha_l^{v0}, \lambda_l^{v0}, \pi^{v0}, \zeta_l^{v0}) \tag{16}$$

$$\delta = (\delta_z^{\sigma}, \delta_x^{\sigma}) \tag{17}$$

$$\psi_i(\theta, \delta) = \begin{bmatrix} \mathbf{s}_{\delta i}(\delta) \\ s_{\xi^v i}(\theta; \delta) \\ s_{\beta^v i}(\theta; \delta) \\ s_{\alpha^v i}(\theta; \delta) \\ s_{\lambda^v i}(\theta; \delta) \\ s_{\pi^v i}(\theta; \delta) \\ s_{\zeta^v i}(\theta; \delta) \end{bmatrix} \tag{18}$$

For example, if I were to assume u_{it} follows a normal distribution and using a logit specification for the conditional mean function, the score functions are

$$\mathbf{s}_{\delta i}(\delta) = \sum_{t=1}^{T} \frac{(\mathbf{z}_{it}\ \mathbf{x}_{it})^{\top}\phi((\mathbf{z}_{it}\ \mathbf{x}_{it})\delta^{\top})[t_{it} - \Phi((\mathbf{z}_{it}\ \mathbf{x}_{it})\delta^{\top})]}{\Phi((\mathbf{z}_{it}\ \mathbf{x}_{it})\delta^{\top})[1 - \Phi((\mathbf{z}_{it}\ \mathbf{x}_{it})\delta^{\top})]}$$

$$\begin{aligned} s_{\zeta_l^v}(\theta; \delta) = \sum_{t=1}^{T} \Bigg[& t_{it} \left(\sum_{k=1, k\neq l}^{L-1} y_{itk} \frac{\int_{-q_{it}}^{\infty} M(d_{itk}^v + \zeta_k^v u_{it}, d_{itl}^v + \zeta_l^v u_{it}) f(u_{it}) du_{it}}{\int_{-q_{it}}^{\infty} \Lambda(d_{itk}^v + \zeta_k^v u_{it}) f(u_{it}) du_{it}} \right) \\ & + t_{it} y_{itl} \frac{\int_{-q_{it}}^{\infty} \Lambda'(d_{itl}^v + \zeta_l^v u_{it}) u_{it} f(u_{it}) du_{it}}{\int_{-q_{it}}^{\infty} \Lambda(d_{itl}^v + \zeta_l^v u_{it}) f(u_{it}) du_{it}} \\ & + t_{it} \left(1 - \sum_{k=1}^{L-1} y_{itk}\right) \frac{\int_{-q_{it}}^{\infty} P(d_{itl}^v + \zeta_l^v u_{it}) u_{it} f(u_{it}) du_{it}}{\int_{-q_{it}}^{\infty} (1 - \sum_{k=1}^{L-1} \Lambda(d_{itk}^v + \zeta_k^v u_{it})) f(u_{it}) du_{it}} \end{aligned}$$

$$
\begin{aligned}
&+ (1-t_{it})y_{itl}\frac{\int_{-\infty}^{-q_{it}} \Lambda'(d_{itl}^{v} + \zeta_l^{v} u_{it})u_{it} f(u_{it})du_{it}}{\int_{-\infty}^{-q_{it}} \Lambda(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}} \\
&+ (1-t_{it})\left(\sum_{k=1,k\neq l}^{L-1} y_{itk}\frac{\int_{-\infty}^{-q_{it}} M(d_{itk}^{v} + \zeta_k^{v} u_{it}, d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-\infty}^{-q_{it}} \Lambda(d_{itk}^{v} + \zeta_k^{v} u_{it}) f(u_{it})du_{it}}\right) \\
&+ (1-t_{it})\left(1-\sum_{k=1}^{L-1} y_{itl}\right)\frac{\int_{-\infty}^{-q_{it}} P(d_{itl}^{v} + \zeta_l^{v} u_{it})u_{it} f(u_{it})du_{it}}{\int_{-\infty}^{-q_{it}}(1-\sum_{k=1}^{L-1}\Lambda(d_{itk}^{v} + \zeta_k^{v} u_{it}))f(u_{it})du_{it}}\Bigg]
\end{aligned}
$$

while for each intercept,

$$
\begin{aligned}
s_{\xi_{tl}^{v}}(\theta;\delta) = &\left(\sum_{k=1,k\neq l}^{L-1} y_{itk}\frac{\int_{-q_{it}}^{\infty} M(d_{itk}^{v} + \zeta_k^{v} u_{it}, d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-q_{it}}^{\infty} \Lambda(d_{itk}^{v} + \zeta_k^{v} u_{it}) f(u_{it})du_{it}}\right) \\
&+ t_{it}y_{itl}\frac{\int_{-q_{it}}^{\infty} \Lambda'(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-q_{it}}^{\infty} \Lambda(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}} \\
&+ t_{it}\left(1-\sum_{k=1}^{L-1} y_{itk}\right)\frac{\int_{-q_{it}}^{\infty} P(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-q_{it}}^{\infty}(1-\sum_{k=1}^{L-1}\Lambda(d_{itk}^{v} + \zeta_k^{v} u_{it}))f(u_{it})du_{it}} \\
&+ (1-t_{it})y_{itl}\frac{\int_{-\infty}^{-q_{it}} \Lambda'(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-\infty}^{-q_{it}} \Lambda(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}} \\
&+ (1-t_{it})\left(\sum_{k=1,k\neq l}^{L-1} y_{itk}\frac{\int_{-\infty}^{-q_{it}} M(d_{itk}^{v} + \zeta_k^{v} u_{it}, d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-\infty}^{-q_{it}} \Lambda(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}\right) \\
&+ (1-t_{it})\left(1-\sum_{k=1}^{L-1} y_{itl}\right)\frac{\int_{-\infty}^{-q} P(d_{itl}^{v} + \zeta_l^{v} u_{it}) f(u_{it})du_{it}}{\int_{-\infty}^{-q_{it}}(1-\sum_{k=1}^{L-1}\Lambda(d_{itk}^{v} + \zeta_k^{v} u_{it}))f(u_{it})du_{it}}
\end{aligned}
$$

Note that

$$
\begin{aligned}
\Lambda'(d_{itl}^{v} + \zeta^{v} u_{it}) &= \frac{e^{(d_{itl}^{v}+\zeta^{v}u_{it})}(1+\sum_{k=1}^{L-1} e^{(d_{itk}^{v}+\zeta^{v}u_{it})}) - e^{2(d_{itl}^{v}+\zeta^{v}u_{it})}}{(1+\sum_{k=1}^{L-1} e^{(d_{itk}^{v}+\zeta^{v}u_{it})})^2} \\
M(d_{itk}^{v} + \zeta^{v} u_{it}, d_{itl}^{v} + \zeta^{v} u_{it}) &= \frac{-e^{(d_{itl}^{v}+\zeta^{v}u_{it})}(e^{(d_{itk}^{v}+\zeta^{v}u_{it})})}{(1+\sum_{k=1}^{L-1} e^{(d_{itk}^{v}+\zeta^{v}u_{it})})^2} \\
P(d_{itk}^{v} + \zeta^{v} u_{it}) &= \frac{(-e^{(d_{itk}^{v}+\zeta^{v}u_{it})})}{(1+\sum_{k=1}^{L-1} e^{(d_{itk}^{v}+\zeta^{v}u_{it})})^2}
\end{aligned}
$$

These scores, along with the scores that are used to estimate δ, are used to generate corresponding sample moments. Now, the standard regularity assumptions from Newey and McFadden (1994) are made.

Assumption C.1. $\Delta \times \Theta$ is compact.

Assumption C.2. $\{(x_{i1}, ..., x_{iT}), (z_{i1}, ..., z_{iT}), (t_{i1}, ..., t_{iT}), [(y_{i11}, ...y_{ilL}), ..., (y_{iT1}, ..., y_{iTL})]\}_{i=1}^{N}$ is iid

Assumption C.3. There exists a unique $(\delta_0, \theta_0) \in \Delta \times \Theta$ such that $E(\psi_i(\theta_0, \delta_0)) = 0$.

Assumption C.4. $E(\sup_{\delta\times\theta\in\Delta\times\Theta} ||\psi_i(\theta, \delta)||) < \infty$

Assumption C.5. $E(||\psi_i(\theta_0, \delta_0)||^2) < \infty$ and $E(||\frac{\partial\psi_i(\theta_0,\delta_0}{\partial(\delta,\theta)^\top})) < \infty$

Then,

$$\sqrt{N}\begin{bmatrix}\hat{\delta} - \delta \\ \hat{\theta} - \theta\end{bmatrix} \to N(0, (D^\top D)^{-1} D^\top \Delta D (D^\top D)^{-1})$$

where $D = E[\frac{\partial\psi_i(\theta_0,\delta_0)}{\partial\theta^\top}]$ and $\Delta = E[\psi_i(\theta_0, \delta_0)\psi_i(\theta_0, \delta_0)^\top]$. The asymptotic variance can be estimated using $\hat{D} = N^{-1}\sum_{i=1}^{N}\frac{\partial\psi_i(\hat{\theta},\hat{\delta})}{\partial\theta^\top}$ as a consistent estimator for D and $\hat{\Delta} = N^{-1}\sum_{i=1}^{N}[\psi_i(\theta, \delta)\psi_i(\theta, \delta)^\top]$ as a consistent estimator for Δ.

4 Average Partial Effects

In order to separate out the average partial effect of the continuous variables from the discrete variable, let $\Upsilon_{it} = (\mathbf{h}_i, \mathbf{x}_{it}, \mathbf{z}_{it})$. Let $\upsilon^j_{(\cdot)}$ denote element j corresponding to the continuous covariate $(\cdot)$. Then (see the appendix),

$$\begin{aligned} E_{u_{it}}[E[y_{itl}|\boldsymbol{\Upsilon}, t_{it} = 1, u_{it}] - E[y_{itl}|\boldsymbol{\Upsilon}, t_{it} = 0, u_{it}]] &= \int \Lambda(d_{tl}^{v,t=1} + \zeta^v u_{it}) f(u_{it}) du_{it} - \int \Lambda(d_{tl}^{v,t=0} + \zeta^v u_{it}) f(u_{it}) du_{it} \\ &= E_{u_{it}}[\Psi_{itl}(\mathbf{x}^0, \mathbf{h}^0; \theta)] = \Sigma_{tl}(\mathbf{x}^0, \mathbf{h}^0) \end{aligned}$$

$$d_{tl}^{v,t=1} = \xi_{tl}^{v} + \mathbf{x}^0\beta_l^v + \alpha_l^v + \lambda_l^v \mathbf{h}^0 \pi^{\mathbf{v}}$$

$$d_{tl}^{v,t=0} = \xi_{tl}^{v} + \mathbf{x}^0\beta_l^v + \lambda_l^v \mathbf{h}_i^0 \pi^{\mathbf{v}}$$

$$\Psi_{tl}(\mathbf{x}^0, \mathbf{h}^0; \theta) = \Lambda(d_{tl}^{v,t=1} + \zeta^v u_{it}) - \Lambda(d_{tl}^{v,t=0} + \zeta^v u_{it})$$

Similarly, the average partial effect for a continuous explanatory variable is,

$$E_{u_{it}}[\partial E[y_{itl}|\mathbf{\Upsilon}, t_{it}, \tilde{u}_{it}]/\partial \upsilon_{(.)}^{j}] = \left[\int_{\mathbb{R}} \partial\Lambda(d_{itl}^{v} + \zeta_l^v u_{it})/\partial \upsilon_{(.)}^{j} f(u_{it}) du_{it}\right] = \Xi_{tl}(\mathbf{x}^0, \mathbf{h}^0, \mathbf{t}^0)$$

In either case, the estimator is similar to the Average Structural Function (ASF) of Blundell and Powell (2001). The expected value of the marginal change in the explanatory variable is taken with respect to the distribution of the unobserved error, the source of the endogeneity that is still present. This leads to the following two theorems,

Theorem 1. Suppose assumptions ID.1-1D.5 and C.1-C.5 hold and $E(\sup_{\delta\times\theta\in\Delta\times\Theta} ||\nabla_{\delta,\theta}\partial(\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial \upsilon_{(.)}^{j})||) < \infty$. Then,

$$\sqrt{N}(\hat{\Xi}_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0) - \Xi_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0)) \rightarrow N(0, V_{tl})$$

where $\hat{\Xi}_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0) = N^{-1}[\sum_{i=1}^{N} \partial\Lambda(\hat{d}_{tl}^{v0} + \hat{\zeta}_l^v u_{it})/\partial \upsilon_{(.)}^{j}]$,

$d_{itl}^{v0} = \xi_{tl}^{v} + \mathbf{x}^0\beta_l^v + t_{it}^0\alpha_l^v + \lambda_l^v \mathbf{h}^0 \pi^v$,

$\hat{d}_{itl}^{v0} = \hat{\xi}_{tl}^{v} + \mathbf{x}^0\hat{\beta}_l^v + t_{it}^0\hat{\alpha}_l^v + \hat{\lambda}_l^v \mathbf{h}^0 \hat{\pi}^v$

$V_{tl} = E[V_{tli}^{\top} V_{tli}]$

$V_{tli} = \partial\Lambda(d_{tl}^{v0} + \zeta^v u_{it})/\partial \upsilon_{(.)}^{j}) - \Xi_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0) + E([\nabla_{\delta,\theta}(\partial\Lambda(d_{tl}^{v0} + \zeta^v u_{it})/\partial \upsilon_{(.)}^{j})])\mathbf{K}_i$

$\mathbf{K}_i = -\mathbf{B}_0^{-1}\mathbf{G}^{\top}{}_0\psi_i(\theta_0, \delta_0)$

$\mathbf{B}_0 = \mathbf{G}^{\top}{}_0\mathbf{G}_0$

$\mathbf{G}_0 = E[\nabla_{\delta,\theta}\psi_i(\theta_0, \delta_0)]$

Theorem 2. Suppose assumptions ID.1-1D.5 and C.1-C.5 hold and $E(\sup_{\delta\times\theta\in\Delta\times\Theta} ||\nabla_{\delta,\theta}\Psi_{tl}(\mathbf{x}^0, \mathbf{h}^0; \theta)||) <$

∞. Then,

$$\sqrt{(}N)(\hat{\Sigma}_{tl}(\mathbf{x}^0, \mathbf{h}^0) - \Sigma_{tl}(\mathbf{x}^0, \mathbf{h}^0))) \rightarrow N(0, J_{tl})$$

where $\hat{\Sigma}_{tl}(\mathbf{x}^0, \mathbf{h}^0) = N^{-1} \sum_{i=1}^{N} [\Lambda(d_{tl}^{\widehat{v,t=1}} + \hat{\zeta}^v u_{it}) - \Lambda(d_{tl}^{\widehat{v,t=0}} + \hat{\zeta}^v u_{it})]$,
$d_{tl}^{\widehat{v,t=1}} = \hat{\xi}_{tl}^v + \mathbf{x}^0\hat{\beta}_l^v + \hat{\alpha}_l^v + \hat{\lambda}_l^v\mathbf{h}^0\hat{\pi}^{\mathbf{v}}$,
$d_{tl}^{v,t=0} = \xi_{tl}^v + \mathbf{x}^0\beta_l^v + \lambda_l^v\mathbf{h}^{\mathbf{0}}\pi^{\mathbf{v}}$
$J_{tl} = E[J_{tli}^\top J_{tli}]$
$J_{tli} = \Psi_{tl}(\mathbf{x}^0, \mathbf{h}^0; \theta) - \hat{\Sigma}_{tl}(\mathbf{x}^0, \mathbf{h}^0) + [E(\nabla_{\delta,\theta}\Psi_{itl}(\mathbf{x}^0, \mathbf{h}^0; \theta))]\mathbf{K}_i$

Theorems 1 and 2 represent the average partial effect taken while holding $\mathbf{h}_i$ at some fixed value. I have left the notation u_{it} to emphasize that this estimator is particular to a specific time period, and to emphasize a single u_{it} is generated for each i. Usually, the average partial effect is taken over the distribution of the unobserved heterogeneity. This is done below. The computations required to obtain this average partial effect lead to the generation of multiple u_{ik}. In contrast to a single u_{it} drawn for each i, multiple draws must be made for each i. The average partial effect when the expectation is taken over the joint distribution of c_i and e_{it}, conditioning on their proxies $\mathbf{h}_i$ and c_i at the fixed values $\mathbf{x}^0$ and t^0 is,

$$E_{\mathbf{h}_i,u_{it}}[\partial E[y_{itl}|\mathbf{x}_{it}^0, t_{it}^0, \mathbf{h}_i, u_{it}]/\partial \mathbf{v}_{it}^j] = \int_{\mathbb{R}}\int_{\mathbb{R}} \partial\Lambda(\xi_{tl}+\mathbf{x}_{it}^0\beta_l+t_{it}^0\alpha_l+\lambda_l\mathbf{h}_i\pi+\zeta_l u_{it})/\partial\mathbf{v}_{it}^j f(u_i|\mathbf{h}_i)f(\mathbf{h}_i)d\mathbf{h}_i\, du_{it} = \Omega_{tl}(\mathbf{x}_t^0, t_t^0)$$

The corresponding estimator is,

$$N^{-1}R^{-1}\sum_{i=1}^{N}\sum_{r=1}^{R} \partial\Lambda(\hat{\xi}_{tl} + \mathbf{x}^0\hat{\beta}_l + t^0\hat{\alpha}_l + \hat{\lambda}_l\mathbf{h}_i\hat{\pi} + \hat{\zeta}_l u_{itr})/\partial\mathbf{v}_{it}^j = \hat{\Omega}_{tl}(\mathbf{x}^0, t^0)$$

Similarly, the APE in this case for the binary EEV is

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \Lambda(\xi_{tl} + \mathbf{x}^0_{it}\beta_l + \alpha_l + \lambda_l h_i \pi + \zeta_l u_{it}) f(u_{it}|\mathbf{h}_i) f(\mathbf{h}_i) dh_i \, du_{it} - \int_{\mathbb{R}}\int_{\mathbb{R}} \Lambda(\xi_{tl} + \mathbf{x}^0_{it}\beta_l + \lambda_l h_i \pi + \zeta_l u_{it}) f(u_{it}|\mathbf{h}_i) f(\mathbf{h}_i) dh_i \, du_{it} = \Gamma_{tl}(\mathbf{x}^0_t)$$

In this case, the corresponding estimator is,

$$N^{-1}R^{-1}\sum_{i=1}^{N}\sum_{r=1}^{R}[\Lambda(\hat{\xi}_{tl} + \mathbf{x}^0_{it}\hat{\beta}_l + \hat{\alpha}_l + \hat{\lambda}_l \mathbf{h}_i \hat{\pi} + \hat{\zeta}_l u_{itr}) - \Lambda(\hat{\xi}_{tl} + \mathbf{x}^0_{it}\hat{\beta}_l + \hat{\lambda}_l \mathbf{h}_i \hat{\pi} + \hat{\zeta}_l u_{itr})] = \hat{\Gamma}_{tl}(\mathbf{x}^0_t)$$

Next, I will present the asymptotic distribution of these APEs. First, I will apply a mean value expansion to $\sqrt{N}\Omega_{tl}(x_0, t_0)$. Let

$$\sqrt{N}\hat{\Omega}_{tl}(\mathbf{x}^0, t^0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) + \nabla \tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \tilde{\theta})(\hat{\theta} - \theta_0)] \tag{19}$$

Where $\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta) = R^{-1}\sum_{r=1}^{R} \partial(\Lambda(\xi_{tl} + \mathbf{x}^0\beta_l + t^0\alpha_l + \lambda_l \mathbf{h}_i \pi + \zeta^v u_{itr})/\partial \upsilon^j_{(.)})$, and $f(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta) = E[\partial(\Lambda(\xi_{tl} + \mathbf{x}^0\beta_l + t^0\alpha_l + \lambda_l \mathbf{h}_i \pi + \zeta^v u_{itr})/\partial \upsilon^j_{(.)})|\mathbf{h}_i]$. Now, note that if $R \to \infty$ along with $N \to \infty$, then by the consistency assumptions and $E(\sup_{\delta\times\theta\in\Delta\times\Theta} ||\nabla_{\delta,\theta}\partial(\Lambda(\xi_{tl} + \mathbf{x}^0\beta_l + t^0\alpha_l + \lambda_l \mathbf{h}_i \pi + \zeta^v u_{it})/\partial \upsilon^j_{(.)})||) < \infty$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nabla_{\delta,\theta}\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \tilde{\theta})(\hat{\eta} - \eta_0) = E[\nabla_{\delta,\theta}\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)]\mathbf{K_N} + o_p(1) \tag{20}$$

Then,

$$\sqrt{N}(\hat{\Omega}_{tl}(\mathbf{x}^0, t^0) - \Omega_{tl}(\mathbf{x}^0, t^0)) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) - \Omega_{tl}(\mathbf{x}^0, t^0) + E[\nabla_{\delta,\theta}\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)]\mathbf{K}_i + o_p(1)] \tag{21}$$

Now, note that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) - \Omega_{tl}(\mathbf{x}^0, t^0) + E[\nabla_{\delta,\theta}\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta)]\mathbf{K}_i$ has mean 0. If this is not immediately obvious, note that unlike in Hajivassiliou and Ruud (1994) the simulations are being carried out directly upon the partial derivative of the multinomial logit function with respect to some explanatory variable for each $\mathbf{h}_i$, as opposed to the simulations being carried out on the likelihood and then taking the gradient. $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)] = \sum_{i=1}^{N} f(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) + A_N + B_N$, where

$$A_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) - E_{u|\mathbf{h}}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)]], \tag{22}$$

$$B_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[E_{u|\mathbf{h}}[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)] - f(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)] \tag{23}$$

By the definition of $\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)$ and $f(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0)$, A_N and B_N both have zero expectation. Then since $E(\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) = E[E[\tilde{f}_R(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0|h_i)]] = \Omega_{tl}(\mathbf{x}^0, t^0)$, the central limit theorem implies,

$$\sqrt{NR}(\hat{\Omega}_{tl}(\mathbf{x}^0, t^0) - \Omega_{tl}(\mathbf{x}^0, t^0)) \xrightarrow{N\to\infty, R\to\infty} N(0, J_{tl}) \tag{24}$$

$$J_{tl} = E(J_{tli}^{\top} J_{tli}) \tag{25}$$

$$J_{tli} = \tilde{f}(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) - \Omega_{tl}(\mathbf{x}^0, t^0) + E[\nabla_{\delta,\theta}\tilde{f}(\mathbf{x}^0, t_0, \mathbf{h}_i; \theta_0)]\mathbf{K}_i \tag{26}$$

$$\tilde{f}(\mathbf{x}^0, t^0, \mathbf{h}_i; \theta_0) = \partial(\Lambda(\xi_{tl} + \mathbf{x}^0\beta_l + t^0\alpha_l + \lambda_l\mathbf{h}_i\pi + \zeta^v u_{itr})/\partial\upsilon^j_{(.)}) \tag{27}$$

Similarly,

$$\sqrt{NR}(\hat{\Gamma}_{tl}(\mathbf{x}_t^0) - (\Gamma_{tl}(\mathbf{x}_t^0)) \xrightarrow{N\to\infty, R\to\infty} N(0, L_{tl}) \tag{28}$$

$$L_{tl} = E(L_{tli}^{\top} L_{tli}) \tag{29}$$

$$L_{tli} = f_{\text{diff}}(\mathbf{x}^0, \mathbf{h}_i; \theta_0) - \Gamma_{tl}(\mathbf{x}_t^0) + E[\nabla_{\delta,\theta}\tilde{f}_{\text{diff}}(\mathbf{x}^0, \mathbf{h}_i; \theta_0)]\mathbf{K}_i \tag{30}$$

$$f_{\text{diff}}(\mathbf{x}^0, \mathbf{h}_i; \theta_0) = \Lambda(\xi_{tl} + \mathbf{x}^0\beta_l + \alpha_l + \lambda_l\mathbf{h}_i\pi + \zeta^v u_{itr}) - \Lambda(\xi_{tl} + \mathbf{x}^0\beta_l + \lambda_l\mathbf{h}_i\pi + \zeta^v u_{itr}) \tag{31}$$

It important to note that even though it would appear that the estimators of the true APEs are limited by the sample size, this is not the case. Since the distribution of u_{it} is known, a researcher could simply simulate u_{it} enough until the researcher feels that the estimators are sufficiently close to the true APEs. It is important to note that these theorems are true for a correctly specified model. That is to say, when equation (9) holds, except for one special case, which will be explored in the simulation section.

5 Test for Endogeneity

The method described relies upon distributional assumptions that are separate from the correct choice of the functional form for the conditional mean function. The distribution of u_{it} would have to be correctly chosen. Furthermore, the method itself is computationally intensive; however, a variable addition test (VAT) based upon Wooldridge (2014) could serve as a test of whether the variable t_{it} is endogenous. Given the choice of for the distribution of u_{it}, the test will utilize a generalized residual as proposed in Gourieroux et al. (1987). The test will rely upon $\zeta_l = 0$ for $l = 1, ..., L-1$. The VAT on the generalized residual will be shown in this section to be asymptotically equivalent to Lagrange Multiplier test under the null hypothesis of no endogeneity. I will show that the Lagrange Multiplier(LM) test statistic has an asymptotic distribution of χ^2_{L-1}, under $H_0 : \zeta_1 = ... = \zeta_{L-1} = 0$. Then, I will show that the VAT is asymptotically equivalent to the LM test, and therefore the VAT statistic also has a χ^2_{L-1} distribution under H_0. It should be noted that the specific test here is based upon Lin and Wooldridge (2017), though the LM test is not infeasible, since there is no additional error arising from a continuous endogenous variable.

Let d^{vr}_{itl} denote the value of d^{v}_{itl} when $\zeta_l = 0$ and θ^r be the estimate of θ based upon the restricted model. Note that in the restricted model t_{it} is exogenous. Let $\hat{gr}_{it}(\hat{\delta})$ denote the estimate of the generalized residual, which is to be used as a consistent estimator of $gr := E(u_{it}|t_{it}, \mathbf{x}_i, \mathbf{z}_i)$. Note that I have written the estimate of the generalized residual as a function of $\hat{\delta}$ to emphasize that the estimates are a function of the estimated parameters from the latent variable equation for t_{it}. Consider the LM statistic, where the estimates of the restricted model will be plugged into the score from the unrestricted model,

$$LM = (\sum_{i=1}^{N} \tilde{\mathbf{S}}_{i,\zeta})^\top \tilde{\mathbf{A}}_{22}[\tilde{\mathbf{V}}_{22}]^{-1}\tilde{\mathbf{A}}_{22}(\sum_{i=1}^{N} \tilde{\mathbf{S}}_{i,\zeta})/N \tag{32}$$

where

$$\tilde{\mathbf{S}}_{i,\zeta} := \frac{\partial lnL_i}{\partial \zeta}\Big|_{\substack{\theta=\theta^r \\ \zeta=0}} = \begin{bmatrix} \sum_{t=1}^{T}[y_{it1}[(1-\Lambda(d_{it1}^{vr})]u_{it} - \sum_{k=1,k\neq 1}^{L-1} y_{itk}\Lambda(d_{it1}^{vr}) - (1-\sum_{k=1}^{L-1} y_{itk})\Lambda(d_{it1}^{vr})u_{it} \\ \vdots \\ \sum_{t=1}^{T}[y_{itL-1}[(1-\Lambda(d_{itL-1}^{vr})]u_{it} - \sum_{k=1,k\neq L-1}^{L-1} y_{itk}\Lambda(d_{itL-1}^{vr} - (1-\sum_{k=1}^{L-1} y_{itk})\Lambda(d_{itL-1}^{vr})u_{it} \end{bmatrix},$$

$$\tilde{\mathbf{A}} := \frac{-1}{N}\begin{bmatrix} \sum_{i=1}^{N}\hat{E}(\frac{\partial^2 lnL_i}{\partial\theta\theta^\top}|t_{it},\mathbf{x}_i,\mathbf{z}_i)\Big|_{\substack{\theta=\theta^r \\ \zeta=0}} & \sum_{i=1}^{N}\hat{E}(\frac{\partial^2 lnL_i}{\partial\theta\zeta^\top}|t_{it},\mathbf{x}_i,\mathbf{z}_i)\Big|_{\substack{\theta=\theta^r \\ \zeta=0}} \\ \sum_{i=1}^{N}\hat{E}(\frac{\partial^2 lnL_i}{\partial\zeta\theta^\top}|t_{it},\mathbf{x}_i,\mathbf{z}_i)\Big|_{\substack{\theta=\theta^r \\ \zeta=0}} & \sum_{i=1}^{N}\hat{E}(\frac{\partial^2 lnL_i}{\partial\zeta\zeta^\top}|t_{it},\mathbf{x}_i,\mathbf{z}_i)\Big|_{\substack{\theta=\theta^r \\ \zeta=0}} \end{bmatrix},$$

$$\tilde{\mathbf{A}}^{-1} = \begin{bmatrix} \tilde{\mathbf{A}}_{11} & \tilde{\mathbf{A}}_{12} \\ \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} \end{bmatrix},$$

$$\tilde{\mathbf{V}} = \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{A}}^{-1} = \begin{bmatrix} \tilde{\mathbf{V}}_{11} & \tilde{\mathbf{V}}_{12} \\ \tilde{\mathbf{V}}_{21} & \tilde{\mathbf{V}}_{22} \end{bmatrix},$$

$$\tilde{\mathbf{B}} := \frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{S}}_{i,\zeta}\tilde{\mathbf{S}}_{i,\zeta}^\top$$

In $\tilde{A}$, the expectation is taken with respect to u_{it}. It is important to note the use of $\hat{E}(\cdot)|t_{it},\mathbf{x}_i,\mathbf{z}_i)$, which is a function of $\hat{gr}_{it}(\hat{\delta})$. While the distribution of $u_{it}|\mathbf{x}_i,\mathbf{z}_i$ is known, this does not imply knowledge of the distribution of $u_{it}|t_{it},\mathbf{x}_i,\mathbf{z}_i$. If this distribution was known, then each element of $\tilde{\mathbf{A}}$ would be a function of the conditional expectations themselves, as opposed to estimators that are functions of the generalized residuals. The summation represents the application of iterated expectations. Applying the summation over i and dividing by N serves as a consistent estimator of the unconditional expected value of the second partial derivatives of the likelihood function.

Alternatively, the log likelihood function for an individual i when implementing the VAT is

$$L_i = \sum_{t=1}^{T}\Big[\left(\sum_{l=1}^{L-1} y_{itl} log\Lambda(d_{itl}^{v} + gr_{it}\tau_l)\right) + \left(1-\sum_{l=1}^{L-1} y_{itl}\right) log\left(1-\sum_{l=1}^{L-1}\Lambda(d_{itl}^{v} + gr_{it}\tau_l)\right)\Big]$$

To implement the VAT, the procedure is as follows:

Procedure 5.1

1. Generate the generalized residuals $\hat{gr}_{it}$ from a first stage pooled maximum likelihood estimation (e.g. probit estimation) of t_{it} on $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$.

2. Obtain the maximum likelihood estimates of the parameters using the individual log-likelihood function L_i .

3. Use a Wald test to determine the joint significance of $\tau_1, \tau_2, ..., \tau_{L-1}$.

Under the null hypothesis, $\tau_l = 0$ for l=1,...,L-1, so the estimation of $\hat{gr}_{it}$ does not affect the asymptotic distribution of the test statistic. The score vector is,

$$\mathbf{S}_{i,\tau} = \begin{bmatrix} \frac{\partial L_i}{\partial \theta} \\ \frac{\partial L_i}{\partial \tau} \end{bmatrix} = \begin{bmatrix} s_{\xi L_i} \\ s_{\beta L_i} \\ s_{\alpha L_i} \\ s_{\lambda L_i} \\ s_{\pi L_i} \\ s_{\tau L_i} \end{bmatrix}$$

Now, since

$$\sqrt{N} \begin{bmatrix} \hat{\theta} - \theta \\ \hat{\tau} - \tau \end{bmatrix} = N^{-1/2} \mathbf{A}^{-1} \sum_{i=1}^{N} \mathbf{S}_i + o_p(1)$$

Under $H_0 : \tau_1 = \ldots = \tau_{L-1} = 0$, the Wald test statistic is

$$W = (\hat{\tau} - \tau)^{\top} (\hat{\mathbf{V}}_{22}/N)^{-1} (\hat{\tau} - \tau) = \sqrt{N} (\hat{\tau} - \tau)^{\top} (\hat{\mathbf{V}}_{22}/N)^{-1} \sqrt{N} (\hat{\tau} - \tau)$$

where

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} \end{bmatrix}$$

$$\hat{\mathbf{V}} = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1} = \begin{bmatrix} \hat{\mathbf{V}}_{11} & \hat{\mathbf{V}}_{12} \\ \hat{\mathbf{V}}_{21} & \hat{\mathbf{V}}_{22} \end{bmatrix},$$

$$\hat{\mathbf{B}} = N^{-1}\sum_{i=1}^{N}(\tilde{\mathbf{S}}_{i,\tau}\tilde{\mathbf{S}}_{i,\tau}^{\top}),$$

$$\hat{\mathbf{A}}^{-1} \xrightarrow{p} \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

$$\hat{p}_{itl} = \hat{\xi}_{tl} + \mathbf{x}_{it}\hat{\beta}_l + t_{it}\hat{\alpha}_l + \hat{\lambda}_l\mathbf{h}_i\hat{\pi} + \hat{\tau}_l\hat{g}r_{it}$$

Then the Wald statistic is,

$$W = (\sum_{i=1}^{N} S_{i,\tau})^{\top}\mathbf{A}_{22}\hat{\mathbf{V}}_{22}^{-1}\mathbf{A}_{22}(\sum_{i=1}^{N} S_{i,\tau})/N \tag{33}$$

Under the null hypothesis that $\tau = 0, \zeta = 0$, $(\hat{\tau} - \tau) \xrightarrow{p} 0$, $\sqrt{N}(\hat{\theta} - \theta)$ and $\sqrt{N}(\tilde{\theta} - \theta)$ converge in distribution. Then $LM - W \xrightarrow{p} 0$, which implies that the tests are asymptotically equivalent (see section 12.6.2 and section 12.6.3 in Wooldridge (2010)).

This result is almost identical to the result from Lin and Wooldridge (2017), but the form of the test statistic is considerably more complex. The multiple response setting leads to complicated forms of the score and Hessian matrices.

6 Simulations

I performed simulations to examine, when equation (9) does not hold, how the estimates of the APEs differ from the true values. The main difficulty in carrying out simulations for the estimator described in this paper is how to approximate the integrals contained within the moment conditions that are a part of each score function. There are two broad classes of methods that were considered to approximate the integrals. Monte Carlo integration could be done, but the sample size used to approximate the integral may have to be so large that computation time may be found to be unacceptably high. Gaussian quadrature was also considered. The specific method used to simulate the integrals was Gauss-Laguerre quadrature.

6.1 Data Generating Process

For each simulation, N=1,000, $L = 3$, and T=4. I used 500 replications.

6.1.1 Regressors

Within each replication, 1,000 observations at each time period t are generated of x_{it}, z_{it}, u_{it}, and v_{it}, where

$$\begin{aligned} x_{it} &\sim Normal(0,4) \\ z_{it} &\sim Uniform(0,1) \\ u_{it} &\sim Normal(0,1) \end{aligned}$$

The variance of x_{it} was chosen to be 4 in order to induce more variation in the data so that when the minimization problem is performed, a local minima will not be chosen as the solution over the global minimum. v_{it} is generated from either a $N(0,1)$, Logistic(0,1), or a χ^2_1 distribution.

Furthermore,

$$\begin{aligned} t_{it} &= 1[\mathbf{z}_{it}\delta_z + \mathbf{x}_{it}\delta_x + u_{it} \geq 0] \\ \delta_z &\in \{0.1, 0.5, 1\} \\ \delta_x &= 0 \end{aligned}$$

and

$$r_{it1} = \zeta_1 u_{it} + v_{it}$$

$$r_{it2} = \zeta_2 u_{it} + v_{it}$$

where

$$\zeta_1 \in \{0.5, 1\}$$

$$\zeta_2 \in \{0.1, 1\}$$

6.1.2 The Fractional Responses

The structural mean function G will be chosen to be the multinomial logit function such that,

$$E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, z_{it}, e_{it}] = \frac{e^{\xi_{tl}+\mathbf{x}_{it}\beta_l+t_{it}\alpha_l\lambda_l+\lambda_l\mathbf{h}_i\pi+\zeta_l u_{it}}}{1+\sum_{k=1}^{L-1} e^{\xi_{tk}+\mathbf{x}_{it}\beta_k+t_{it}\alpha_k\lambda_k+\lambda_k\mathbf{h}_i\pi+\zeta_k u_{it}}} \tag{34}$$

$$E[y_{itL}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, z_{it}, u_{it}] = \frac{1}{1+\sum_{k=1}^{L-1} e^{\xi_{tk}+\mathbf{x}_{it}\beta_k+t_{it}\alpha_k\lambda_k+\lambda_k\mathbf{h}_i\pi+\zeta_k u_{it}}} \tag{35}$$

where L =3, $\beta_1 = 1, \beta_2 = 2, \alpha_1 = 1, \alpha_2 = 2, \lambda_1 = 1, \lambda_2 = 2, \pi = (1, 1, 1)$, and $\xi_{tl} = 0$ for all t and l. In order to generate the fractional response variables, the following procedure from Nam (2013) is used,

1. Calculate the response probabilities using (24), (25), and the aforementioned regressors.

2. Draw 100 multinomial outcomes at each i and t among choices 1,2, and 3 based upon the response probabilities.

3. Count the frequencies at each i and t and obtain the proportion for each outcome.

When $v_{it} \sim Normal(0, 1)$, then this is a special case in which equation (9) does not hold yet the estimates of the APEs will be consistent. This is because the choice of G to be the multinomial logit function and $u_{it} \sim Normal(0, 1)$ causes the QMLE problem to be set up as if the researcher is working directly with the structural mean function. In this situation, the estimates of the parameters are the estimates of the structural parameters. In this case, the simulations will serve as a check upon the consistency of the estimates.

6.2 Simulation Results

Each replication examines the data at the 25th, 50th, 75th, and 90th percentile of the data, whereby the first time period will be used to examine the APEs at the 25th percentile, the second time period will be used for the 50th percentile, and so on. The tables include the results from the 25th, 50th, and the 75th percentiles. Although there are no time varying intercepts in these simulations, presumably a researcher would want to know the APEs at each time period in order to account for the time-varying intercepts, and they might want to see how the percentile APEs change given these intercepts.

The true APEs are constructed from using within each replication the data at the aforementioned percentiles and the true values of the parameters. Then, across these replications, the mean value is obtained. The standard errors of the estimates are obtained by calculating the square root of the sample variance across the 500 replications. All simulation tables can be found in the appendix. Table 1 displays the result that at the lower percentiles and when $v_{it} \sim Normal(0, 1)$, the point estimates sometimes underestimate the true APEs; however, at the 50th and 75th percentiles, the point estimates are accurate and precise. While the standard errors are somewhat large on the partial effect corresponding to the binary EEV, they are not so large as to be greater than the estimates. Furthermore, all of the APEs have the correct sign. Differences in the estimates are minor when adjusting the values of δ_z, ζ_1, and ζ_2.

Now, consider when $v_{it} \sim Logistic(0, 1)$. In this situation, equation (9) does not hold, yet we have a random variable with a probability density function that has heavier tails than the standard normal pdf, and logistic random variables have been used to approximate normal random variables. The results from this simulation are in Table 2. The results are similar to the normal case, though the Monte Carlo standard deviations are noticeably larger. Once again, note that the estimates in the estimates are minor as the parameter values noted in the tables change.

When $v_{it} \sim \chi^2_1$, then the model misspecification is substantial. The distribution of v_{it} cannot be used to approximate the normal distribution, nor is the distribution symmetric. The results of this simulation are in Table 3. In some sense, this is an improvement over the logistic case. The standard errors are often smaller on the APEs. It is important to note that the estimates do not give the incorrect sign at any of the percentiles. This is in contrast to Nam (2013), in which she points out that when a linear control function approach is taken, the estimates of the APEs can take the wrong sign.

The estimates of the APEs taken with respect to the joint expectation of c_i and e_{it} are also provided. In this case, $v_{it} \sim N(0, 1)$. The results are provided in Table 4. The estimated APE of t_1 at the 25th percentile gives a poor estimate of the estimate true APE. It is suspicious that the estimate equals its standard error. There is no reason to think that the estimator would perform poorly at the 25th percentile, especially in light of the full results presented above. It is worth noting that the distribution of u_{it} used to construct these estimates was not made while simulating the correct distribution of $u_{it}|x_{it}, z_{it}, t_{it}$. This is done for two reasons. First, the researcher would not know the distribution of $u_{itr}|\mathbf{h}_i$. Second, this would add another layer of complexity to the simulation, and it would only change the values of the percentile mean and estimates of the average partial effects. The results would still demonstrate that the estimator is consistent.

7 Application

In order to demonstrate the practical usefulness estimator, I used data from the NLSY97 project. The project, undertaken by the U.S. Bureau of Labor Statistics, gathered information on individuals born between 1980 and 1984. The survey data available is based upon eighteen rounds of questioning, from 1997 to 2018. Respondents were asked questions in the areas of employment, education, geography and household information, parental and childhood information, dating and marriage, health, income and assets, attitudes, expectations, activities, crime, and substance use.

I constructed this dataset by using a balanced panel of respondents from the years 2007, 2009, 2010, and 2011. These were years for which there was data on the question of how many hours on average each night each respondent slept. I multiplied by the number of days each year and divided by the total number of hours in that year in order to obtain the fraction of time devoted to sleep in that year. Respondent were also asked each week of each year how many hours they had worked that week. I then multiplied by the number of weeks in that year and divided by the total number of hours in the year in order to obtain the fraction of time in that year devoted to work. The sum of these two fractions are then subtracted from 1, which gives the fraction of time in that year devoted to leisure.

The binary endogenous variable is the marital status of the respondent. In the survey, marital status has six categories. These are whether the respondent is neither married nor cohabitating, not married but cohabitating, married and cohabitating, legally separated, divorced, and widowed. This is collapsed into a binary variable indicating whether the respondent is legally married. The respondent would be counted as married if they are married and cohabitating or legally separated. The marital status of each respondent is updated each month throughout the survey. Since the unit of time for the panel is a year, the martial status is recorded to be the status at the end of the year.

The instrument that I chose for marital status gives insight into working with fractional time variables in the context of time-use data. I created a variable for the number of times that a respondent engaged in sexual intercourse within each year. This was done by using data across three survey questions for each year. In one question, respondents are asked if they engaged in sexual intercourse. Respondent are then asked how often they engaged in intercourse. If they respond that they do not know, respondents are then asked to give an estimate of the amount of time that they engaged in intercourse. I combined the answers for each of these survey questions into a single variable which equaled the number of times that a respondent engaged in sexual intercourse. The results from the first-stage probit are given in Table 5. Additional covariates include the level of education of each respondent, whether they live in an urban area, the number of biological children at their residence, and their household income in thousands of US Dollars. Health controls are added such as the number of times that a respondent was treated by a doctor or a nurse during the year, the number of times that a respondent was sick but did not seek treatment from a doctor or a nurse, whether their health affected the amount of work that they engaged in, and if they have a chronic condition. The z-score is approximately 6.00 on the instrument for the entire sample, and approximately 4 for the male and female subsamples. Table 6 gives the estimates from assuming a linear probability model for the binary marriage variable. This model was not used to obtain estimates of δ in order to estimate the APEs, but instead to display the strength of the instrument. The conventional wisdom, based upon Staiger and Stock (1997), is that a sufficiently strong instrument yields a first stage F-statistic of at least 10. In the full sample the F-statistics of marriage is approximately 36, while in each subsample the F-statistic is approximately 16. The reasons for this are partly due to the issue of finding a new partner. Levinger and Moles (1979) notes that cohabitating couples consider the frequency of sex as well as the ease of finding other partners before dissolving the relationship. Before marriage, the costs of finding a new partner and dissolving the relationship are low. As Oppenheimer (1988) notes, there are considerable search costs to finding a new partner in order to replace lost sexual activity from a relationship. In the absence of children and the legal barriers that must be overcome to end a marriage, these search costs should be lower for unmarried couples. Couples then enter into marriage in order to secure the benefits of the current relationship and facilitate continued emotional investment into their relationships (see Yabiku and Gager (2009)).

For married couples, in addition to higher costs of ending a marriage, the issue of infertility can play a role in ending the relationship. Sexual intercourse can act as the channel through which infertility acts.As Andrews, Abbey, and Halman (1991) note, infertility can lead to lower sexual self-esteem, which in turn leads to a lower frequency of sexual intercourse among married couples. Furthermore, just as for unmarried couples, the frequency of sexual intercourse plays a role in determining whether to end the relationship. In fact, a past national survey revealed that it was the second greatest issue of concern for young married couples (see Yabiku and Gager (2009)). The count of the number of times that each respondent had sex in each year also satisfies the excludability condition, and the reasons why are as follows. First, the information given by the variable itself does not communicate the number of hours or fraction of time in a year that a respondent spent engaging in intercourse. Second, even if this information was known, there is nothing to suggest that the time used on intercourse would not still be counted amongst the broadly defined leisure fraction. In other words, if respondents engaged in less intercourse, there is no reason to think that they would not devote that time instead to other activities that are neither work nor sleep, particularly for the population under study over the time period under study given the controls. Liu (2000) provides some evidence for this among married couples, in the sense that declines in sexual intercourse amongst this group is due to substitutions away from intercourse towards other goods, services, and activities. Contrast this variable with the marriage variable or the number of children at the respondent's residence. These variables change how much an individual allocates of their time to labor, leisure, and work, as opposed to merely the activities within those categories. There seems to be a specific relationship between the frequency of sexual intercourse and the raising of young children. Based upon a panel of German households, Schröder and Schmiedeberg (2015) note that sexual frequency declines until a child reaches approximately six years old, and then sexual frequency tends to increase. The inclusion of the number of biological children in the respondent's household is meant to control for some relationship between sexual intercourse and the raising for children. The health variables are included in order to control for possible correlations between sex and health outcomes, which in turn could have an effect on sleeping patterns or hours worked in a year. After all, the frequency of sexual intercourse is correlated with the health of individuals (see, for example,Walfisch, Maoz, and Antonovsky (1984)). It is important to control for these specific effects, since these effects are noted in health and marriage literature to be correlated with sex and would seem to have an effect on how individuals allocate their time.

Tables 7 and 8 display the APE estimates of marriage over the distribution of the unobserved heterogeneity across respondents. The APE at the 25th percentile evaluates the APE at the 25th percentile of the data in the first time period, the APE at the 50th percentile evaluates the APE at the 50th percentile, and so on, as was done in the simulations. The standard errors are generated using 50 bootstrap replications with a sample size of 100. I chose the value of R to be 300. The intercepts are not time varying, though they do vary with each fractional response. Within each table I have included estimates using only the male respondents and female respondents, in order to determine the effect of marriage upon the use of time of men and women separately. Across the entire sample and the subsamples, the APEs are significant for marriage at α= .05, though the effect upon sleep is not significant at every percentile level. These effects differ depending upon the subsample. Using the entire sample, it would seem that marriage is not associated with a significant effect upon the fraction of time devoted to sleep, but it is associated with a significant negative effect at each percentile across all time periods in the fraction of time devoted to work. For the subsample of men, the APE's are significant and negative for sleep, but they are positive and significant for work. For women, marriage suggests declines in the fraction of time devoted to both work and sleep, which leads to an increase in what is broadly considered by myself to be leisure, though activities which might have more time devoted to their completion may not be considered "leisurely."The estimated average partial effects of Theorems 1 and 2 are not included in this paper, though they are included in the code accompanying this paper.

Table 7 displays the results from applying the estimator of the parameters that I have introduced and the technique of obtaining the APEs that is necessary when a binary covariate is endogenous. Table 8 displays the results from applying the Mundlack-Chamberlain device, but the additional source of endogeneity has not been integrated out. Both the magnitudes and signs of the average partial effects differ across the tables. The estimated effect of marriage on the fraction of time devoted to work and the fraction of time devoted to sleep both seem negligible. This statement also applies to the estimates that are based upon the subsample of men. The effects are larger for women, though it would seem that marriage does not lead to declines in fraction of time devoted to sleep. This suggests that there is a significant source of individual endogeneity that arises from the individual and time specific errors.

8 Conclusion

For the number of observations N and the number of draws R ,I have provided a $\sqrt{NR}$ consistent estimator of the APEs for the multiple fractional response setting while allowing for panel data, unobserved heterogeneity, and a binary EEV. An advantage of this approach is that the constraint upon the conditional mean $\sum_{l=1}^{L} E(y_{itl}|\mathbf{h}_i, \mathbf{x}_{it}, \mathbf{z}_{it}, t_{it}, u_{it}) = 1$ is satisfied, which will not always hold for every choice of the conditional mean function. For example, a probit conditional mean specification is not guaranteed to satisfy the constraint that $\sum_{l=1}^{L} E(y_{itl}|\mathbf{h}_i, \mathbf{x}_{it}, \mathbf{z}_{it}, t_{it}, u_{it}) = 1$,and to assume a linear model is not any better. At best, a linear model could allow for a system regression to provide linear approximations of the average partial effects, and the estimator would be expensive if differences are taken to eliminate the unobserved heterogeneity. At worst, the APE's could be such poor estimates that they not only fail to reflect the relationship between the dependent variable and the relevant covariate, but they fail to appropriately consider the relationship amongst the dependent variables.

If the multinomial logit conditional mean specification (or any specification that satisfies the aforementioned constraint) is chosen, then there are still some choices estimators. Additional moment conditions could be added to the GMM estimator in order to increase efficiency. A working correlation matrix could be used to obtain a GEE estimator. Both estimators would bring efficiency gains over the estimator I have proposed, but they would be more computationally burdensome. Even in a setting where $L = 3$ and $T = 2$, such estimators would increase efficiency but may significantly increase computation time.

A separate issue is whether to rely upon this method or some QMLE estimator derived from some multinomial likelihood problem while not integrating out the unobserved heterogeneity. A failure to integrate out a source of endogeneity that is persistent across subjects and time periods can lead to insignificant average partial effects. In order to determine whether the full method and estimator in this paper are necessary, the variable addition test for endogeneity should be applied.

References

Andrews, Frank M, Antonia Abbey, and L Jill Halman (1991). “Stress from infertility, marriage factors, and subjective well-being of wives and husbands”. In: *Journal of health and social behavior*, pp. 238–253.

Blundell, Richard and James L Powell (2001). “Endogeneity in nonparametric and semiparametric regression models”. In:

Chamberlain, Gary (1980). “Analysis of Covariance with Qualitative Data”. In: *The Review of Economic Studies* 47.1, pp. 225–238.

Gourieroux, Christian et al. (1987). “Generalised residuals”. In: *Journal of econometrics* 34.1-2, pp. 5–32.

Hajivassiliou, Vassilis A and Paul A Ruud (1994). “Classical estimation methods for LDV models using simulation”. In: *Handbook of econometrics* 4, pp. 2383–2441.

Heckman, James J (1979). “Sample selection bias as a specification error”. In: *Econometrica: Journal of the econometric society*, pp. 153–161.

Levinger, George George Klaus, Oliver C Moles, et al. (1979). *Divorce and separation*. Basic Books.

Lin, Wei and Jeffrey M Wooldridge (2017). *Binary and fractional response models with continuous and binary endogenous explanatory variables*. Tech. rep. Working paper, available at http://www. weilinme trics. com/uploads/5/1/4/0 ?

Liu, Chien (2000). “A theory of marital sexual life”. In: *Journal of Marriage and Family* 62.2, pp. 363–374.

Mullahy, John (2015). “Multivariate fractional regression estimation of econometric share models”. In: *Journal of Econometric Methods* 4.1, pp. 71–100.

Nam, Suhyeon. (2013). “Multiple Fractional Response Variables with Continous Endogenous Explanatory Variables”. Job Market Paper.

Newey, KW and Daniel McFadden (1994). “Large sample estimation and hypothesis”. In: *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, pp. 2112–2245.

Oppenheimer, Valerie Kincade (1988). “A theory of marriage timing”. In: *American journal of sociology* 94.3, pp. 563–591.

Papke, Leslie E and Jeffrey M Wooldridge (2008). “Panel data methods for fractional response variables with an application to test pass rates”. In: *Journal of econometrics* 145.1-2, pp. 121–133.

Petrin, Amil and Kenneth Train (2010). “A control function approach to endogeneity in consumer choice models”. In: *Journal of marketing research* 47.1, pp. 3–13.

Rivers, Douglas and Quang Vuong (2002). “Model selection tests for nonlinear dynamic models”. In: *The Econometrics Journal* 5.1, pp. 1–39.

Schröder, Jette and Claudia Schmiedeberg (2015). “Effects of relationship duration, cohabitation, and marriage on the frequency of intercourse in couples: Findings from German panel data”. In: *Social Science Research* 52, pp. 72–82.

Staiger, Douglas and James H Stock (1997). “Instrumental Variables Regression with Weak Instruments”. In: *Econometrica* 65.3, pp. 557–586.

Walfisch, S, B Maoz, and H Antonovsky (1984). “Sexual satisfaction among middle-aged couples: correlation with frequency of intercourse and health status”. In: *Maturitas* 6.3, pp. 285–296.

Woodland, Alan D (1979). “Stochastic specification and the estimation of share equations”. In: *Journal of Econometrics* 10.3, pp. 361–383.

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.

— (2014). “Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables”. In: *Journal of Econometrics* 182.1, pp. 226–234.

Yabiku, Scott T and Constance T Gager (2009). “Sexual frequency and the stability of marital and cohabiting unions”. In: *Journal of marriage and family* 71.4, pp. 983–1000.

Zeger, Scott L, Kung-Yee Liang, and Paul S Albert (1988). “Models for longitudinal data: a generalized estimating equation approach”. In: *Biometrics*, pp. 1049–1060.

Appendices

A Deriving the Average Partial Effects

Note that,

$$E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, \mathbf{z}_{it}, r_{itl}] = G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l \mathbf{h}_i \pi + r_{itl}) \tag{36}$$

$$E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, \mathbf{z}_{it}, u_{it}] = E[E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, u_{it}, r_{itl}]|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, u_{it}] \tag{37}$$

$$= \int_{\mathbb{R}} G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \mathbf{h}_i \pi + r_{itl}) f(r_{itl}|u_{it}) dr_{itl} \tag{38}$$

where the second equality follows by the independence of v_{it} and the explanatory variables and instruments. If we take the partial derivative of the above with respect to $v^j_{\cdot}$, then

$$\partial E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, \mathbf{z}_{it}, u_{it}]/\partial v^j_{(.)} = \int_{\mathbb{R}} (\partial G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \mathbf{h}_i \pi + r_{itl})/\partial v^j_{(.)}) f(r_{itl}|u_{it}) dr_{itl} \tag{39}$$

Then for values $\mathbf{x}_{it}^0, t_{it}^0, \mathbf{h_i^0}$, take the expectation of both sides of equation (39) with respect to u_{it},

$$\int_{\mathbb{R}} \partial\Lambda(d_{tl}^{v0} + \zeta_l^v u_{it})/\partial v_{(.)}^j f(u_{it}) du_{it} = E[\partial E[y_{itl}|\mathbf{x}^0, t^0, \mathbf{h}^0, u_{it}]/\partial v_{(.)}^j] \tag{40}$$

$$= E[E[\partial G(\xi_{tl} + \mathbf{x}^0\beta_l + t^0\alpha_l + \mathbf{h}^0\pi + r_{itl})/\partial v_{(.)}^j)|u_{it}]] \tag{41}$$

$$= E[\partial G(\xi_{tl} + \mathbf{x}^0\beta_l + t^0\alpha_l + \mathbf{h}^0\pi + r_{itl})/\partial v_{(.)}^j)] \tag{42}$$

Hence, $\int_{\mathbb{R}} \partial\Lambda(d_{tl}^{v0} + \zeta_l^v u_{it})/\partial v_{(.)}^j f(u_{it}) du_{it} = E[\partial G(\xi_{tl} + \mathbf{x}^0\beta_l + t_{it}^0\alpha_l + \mathbf{h}_i^0\pi + r_{itl})/\partial v_{(.)}^j)]$.

Now, note that

$$E[y_{itl}|\mathbf{x}_{it}, t_{it}, c_i, \mathbf{z}_{it}, e_{it}] = G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it}) \tag{43}$$

$$E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, \mathbf{z}_{it}, u_{it}] = E[E[y_{itl}|\mathbf{x}_{it}, t_{it}, c_i, e_{it}]|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, u_{it}] \tag{44}$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it}) f(c_i, e_{it}|h_i, u_{it}) dh_i\, du_{it} \tag{45}$$

where the second equality follows from $r_{itl} = \zeta_l + v_{it}$, and v_{it} is independent of all of the explanatory variables, the instruments, and u_{it}. If we take the partial derivative of the above with respect to $\mathbf{x}_{it}^j$, then

$$\partial E[y_{itl}|\mathbf{x}_{it}, t_{it}, \mathbf{h}_i, \mathbf{z}_{it}, u_{it}]/\partial \mathbf{x}_{it}^j = \int_{\mathbb{R}}\int_{\mathbb{R}} (\partial G(\xi_{tl} + \mathbf{x}_{it}\beta_l + t_{it}\alpha_l + \lambda_l c_i + e_{it})/\partial \mathbf{v}_{it}^j) f(c_i, e_{it}|h_i, u_{it}) dh_i\, du_{it}$$

Then for values $\mathbf{x}_{it}^0, t_{it}^0$,

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} \partial\Lambda(\xi_{tl} + \mathbf{x}_{it}^0\beta_l + t_{it}^0\alpha_l + \lambda_l h_i\pi + \zeta_l u_{it})/\partial \mathbf{v}_{it}^j f(u_{it}|\mathbf{h}_i) f(\mathbf{h}_i) dh_i\, du_{it} \\
&= E[\partial E[y_{itl}|\mathbf{x}_{it}^0, t_{it}^0, \mathbf{h}_i, u_{it}]/\partial \mathbf{v}_{it}] \\
&= E[E[\partial G(\xi_{tl} + \mathbf{x}_{it}^0\beta_l + t_{it}^0\alpha_l + \lambda_l c_i + e_{it})/\partial \mathbf{v}_{it}^j|\mathbf{h}_i, u_{it}]] \\
&= E[\partial G(\xi_{tl} + \mathbf{x}_{it}^0\beta_l + t_{it}^0\alpha_l + \lambda_l c_i + e_{it})/\partial \mathbf{v}_{it}^j]
\end{aligned}$$

B Proof of Theorem 1

$$\sqrt{N}(\hat{\eta}-\eta) = \sqrt{N}\begin{bmatrix}\hat{\delta}-\delta \\ \hat{\theta}-\theta\end{bmatrix} = \mathbf{K}_N + o_p(1) \tag{46}$$

Where $\mathbf{K}_N = -\mathbf{B}_0^{-1}\mathbf{G}_0^{\top}N^{-1/2}\sum_{i=1}^{N}\psi_i(\theta_0,\delta_0)$. Now, apply a mean value expansion to $\hat{\Xi}$ around η. Then,

$$\sqrt{N}\hat{\Xi}_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0) = \frac{1}{\sqrt{N}}[\sum_{i=1}^{N}[\partial\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j} + \nabla_{\delta,\theta}\partial(\Lambda(d_{tl}^{\widetilde{v}}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j})](\hat{\eta}-\eta)] \tag{47}$$

where $d_{tl}^{\widetilde{v}}$ is d_{tl} evaluated at some $\widetilde{\theta}$ between $\hat{\theta}$ and θ_0. By (46) and the Weak Law of Large Numbers,

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nabla_{\delta,\theta}\partial(\Lambda(d_{itl}^{\widetilde{v}}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j})](\hat{\eta}-\eta)\\ =&\frac{1}{N}\sum_{i=1}^{N}\nabla_{\delta,\theta}\partial(\Lambda(d_{tl}^{\widetilde{v}}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j})]\sqrt{N}(\hat{\eta}-\eta)\\ =&E[\nabla_{\delta,\theta}\partial(\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j})])\mathbf{K}_N] + o_p(1)\end{aligned}$$

Then after subtracting $\sqrt{N}\Xi_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0)$ from $\sqrt{N}\hat{\Xi}_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0)$ the following is obtained,

$$\sqrt{N}(\hat{\Xi}_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0)-\Xi_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0)) = \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}[\partial\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j} - \Xi_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0) + E[\nabla_{\delta,\theta}\partial(\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j})])\mathbf{K}_i] + o_p(1)$$

where $\mathbf{K}_i = -\mathbf{B}_0^{-1}\mathbf{G}_0^{\top}\psi_i(\theta_0,\delta_0)$. Since

$$E[\partial\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j} - \Xi_{tl}(\mathbf{x}^0,\mathbf{h}^0,t^0) + E[\nabla_{\delta,\theta}\partial(\Lambda(d_{tl}^{v0}+\zeta^v u_{it})/\partial\upsilon_{(.)}^{j})])\mathbf{K}_i]] = 0$$

Then by the Central Limit Theorem,

$$\sqrt{N}(\hat{\Xi}_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0) - \Xi_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0)) \to N(0, V_{tl}) \tag{48}$$

where,

$V_{tl} = E[V_{tli}^{\top} V_{tli}]$

$V_{tli} = \partial\Lambda(d_{tl}^{v0} + \zeta^v u_{it})/\partial\upsilon_{(.)}^{j}) - \Xi_{tl}(\mathbf{x}^0, \mathbf{h}^0, t^0) + E([\nabla_{\delta,\theta}(\partial\Lambda(d_{tl}^{v0} + \zeta^v u_{it})/\partial\upsilon_{(.)}^{j})])\mathbf{K}_i$

The proof of Theorem 2 is similar.

C Simulation Tables

Table 1: APE: $v_{it} \sim Normal(0, 1)$

	Percentile Mean			Percentile Estimates		
$\delta_z = 1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0952	-0.0933	-0.0023	0.0685	-0.0815	-0.0020
				(.0204)	(.0323)	(.0011)
t_2	0.0352	0.1549	0.0023	0.0281	0.1316	0.0021
				(.0113)	(.0522)	(.0011)
x_1	0.1365	0.0695	0.0013	0.1207	0.0624	0.0012
				(.0105)	(.0110)	(.0003)
x_2	0.0820	0.1629	0.0027	0.0694	0.1638	0.0026
				(.0214)	(.0320)	(.0006)
$\delta_z = 0.5, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0952	-0.0933	-0.0023	0.0734	-0.0815	-0.0024
				(.0149)	(.0238)	(.0008)
t_2	0.0352	0.1549	0.0023	0.0306	0.1504	0.0024
				(.0110)	(.0409)	(.0009)
x_1	0.1365	0.0695	0.0013	0.1198	0.0599	0.0012
				(.0104)	(.0102)	(.0003)
x_2	0.0820	0.1629	0.0027	0.0698	0.1573	0.0026
				(.0219)	(.0298)	(.0006)
$\delta_z = 0.1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0952	-0.0933	-0.0023	0.0708	-0.0934	-0.0024
				(.0159)	(.0258))	(.0009)
t_2	0.0352	0.1549	0.0302	0.1489	0.1504	0.0024
				(.0114)	(.0430)	(.0009)
x_1	0.1365	0.0695	0.0013	0.1190	0.0601	0.0012
				(.0105)	(.0108)	(.0003)
x_2	0.0820	0.1629	0.0027	0.0698	0.148	0.0025
				(.0220)	(.0317)	(.0006)

Standard errors in parentheses

Table 2: APE: $v_{it} \sim Logistic(0,1)$

	Percentile Mean			Percentile Estimates		
$\delta_z = 1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0881	-0.0886	-0.0023	0.0299	-0.0721	-0.0031
				(.0465)	(.0932)	(.0104)
t_2	0.0385	0.1666	0.0024	0.0065	0.1278	0.0036
				(.0727)	(.1819)	(.0168)
x_1	0.1333	0.0684	0.0013	0.0985	0.0534	0.0014
				(.0104)	(.0185)	(.0026)
x_2	0.0915	0.1758	0.0027	0.0694	0.1752	0.0037
				(.0252)	(.0404)	(.0057)
$\delta_z = 0.5, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0881	-0.0886	-0.0023	0.0175	-0.0790	-0.0024
				(.0262)	(.0437)	(.0025)
t_2	0.0385	0.1666	0.0024	0.0252	0.1289	0.0025
				(.0175)	(.0768)	(.0027)
x_1	0.1333	0.0684	0.0013	0.0943	0.0479	0.0010
				(.0069)	(.0096)	(.0002)
x_2	0.0915	0.1758	0.0027	0.0666	0.1711	0.0027
				(.0214)	(.0365)	(.0007)
$\delta_z = 0.1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0881	-0.0886	-0.0023	0.0062	-0.0688	-0.0019
				(.0230)	(.0441)	(.0015)
t_2	0.0385	0.1666	0.0024	0.0205	0.1091	0.0020
				(.0151)	(.0759)	(.0016)
x_1	0.1333	0.0684	0.0013	0.0921	0.0498	0.0010
				(.0069)	(.0109)	(.0003)
x_2	0.0915	0.1758	0.0027	0.0633	0.1802	0.0028
				(.0188)	(.0397)	(.0007)

Standard errors in parentheses

Table 3: APE: $v_{it} \sim \chi^2_1$

	Percentile Mean			Percentile Estimates		
$\delta_z = 1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.1116	-0.1004	-0.0023	0.1836	-0.0265	-0.0006
				(.0475)	(.0533)	(.0021)
t_2	0.0504	0.1300	0.0023	0.0334	0.0547	0.0006
				(.0473)	(.0768)	(.0023)
x_1	0.1712	0.0701	0.0013	0.1625	0.0576	0.0010
				(.0104)	(.0185)	(.0026)
x_2	0.1200	0.1503	0.0027	0.1086	0.1508	0.0025
				(.0308)	(.0329)	(.0027)
$\delta_z = 0.5, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.1116	-0.1004	-0.0023	0.2763	-0.0895	-0.0020
				(.0268)	(.0246)	(.0007)
t_2	0.0504	0.1300	0.0023	0.0724	0.1414	0.0021
				(.0231)	(.0415)	(.0008)
x_1	0.1712	0.0701	0.0013	0.1730	0.0489	0.0009
				(.0068)	(.0089)	(.0002)
x_2	0.1200	0.1503	0.0027	0.1311	0.1250	0.0021
				(.0359)	(.0236)	(.0005)
$\delta_z = 0.1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.1116	-0.1004	-0.0023	0.3166	-0.0919	-0.0021
				(.0290)	(.0254)	(.0008)
t_2	0.0504	0.1300	0.0023	0.0779	0.1463	0.0021
				(.0255)	(.0422)	(.0008)
x_1	0.1712	0.0701	0.0013	0.1773	0.0497	0.0009
				(.0062)	(.0096)	(.0002)
x_2	0.1200	0.1503	0.0027	0.1382	0.1259	0.0021
				(.0383)	(.0251)	(.0005)

Standard errors in parentheses

Table 4: APE: $v_{it} \sim Normal(0,1)$

	Percentile Mean			Percentile Estimates		
$\delta_z = 1, \zeta_1 = 1, \zeta_2 = 1$						
Covariate	25	50	75	25	50	75
t_1	0.0484	-0.0412	-0.0284	0.0077	-0.0466	-0.0285
				(.0077)	(.0082)	(.0058)
t_2	0.1474	0.1470	0.0477	0.1256	0.1422	0.0496
				(.0202)	(.0232)	(.0104)
x_1	0.1338	0.0950	0.0265	0.1173	0.0817	0.0226
				(.0035)	(.0057)	(.0033)
x_2	0.2112	0.2017	0.0604	0.1979	0.1897	0.0569
				(.0101)	(.0099)	(.0081)

Standard errors in parentheses

D First Stage Marriage Coefficient Estimates

Table 5: Probit marriage coefficient estimates

	(all)	(male only)	(female only)
education	0.0474	0.0561	0.0418
	(.0127)	(.0176)	(.0196)
household income	0.0085	0.0074	0.0069
	(.0012)	(.0016)	(.0021)
children at residence	0.5223	0.8566	0.3462
	(.0368)	(.0703)	(.0470)
urban	-0.2565	-0.1025	-0.3888
	(.0800)	(.1163)	(.1138)
sexual intercourse	0.0010	0.0008	0.0011
	(.00017)	(.0002)	(.0003)
ill w/ treatment	0.0008	0.008	-0.0110
	(.0227)	(.0392)	(.0275)
ill w/o treatment	0.0106	0.0292	0.0152
	(.0189)	(.0292)	(.0248)
chronic	0.0936	0.1234	0.1008
	(.0718)	(.1045)	(.0984)
work limited by illness	-0.3571	-0.3608	-0.2585
	(.1568)	(.2407)	(.1914)
_const	-1.487	-1.811	-1.118
	(.2012)	(.2725)	(.3211)
N	1388	734	654
NT	5552	2936	2616

Standard errors in parentheses

Table 6: Linear marriage coefficient estimates

	(all)	(male only)	(female only)
education	0.0155	0.0149	0.0152
	(.0041)	(.0050)	(.0070)
household income	0.0029	0.0023	0.0026
	(.0004)	(.0005)	(.0008)
children at residence	0.1831	0.2591	0.1280
	(.0109)	(.0140)	(.0162)
urban	-0.0857	-0.0362	-0.1396
	(.0270)	(.0350)	(.0411)
sexual intercourse	0.0003	0.0002	0.0004
	(.000055)	(.00006)	(.0001)
ill w/ treatment	0.00002	0.0043	-0.0042
	(.0078)	(.0122)	(.0100)
ill w/o treatment	0.0031	0.0003	0.0055
	(.0063)	(.0086)	(.0090)
chronic	0.0299	0.0280	0.0356
	(.0242)	(.0319)	(.0356)
work limited by illness	-0.1086	-0.0881	-0.0922
	(.0458)	(.0536)	(.0650)
_const	-.0.0061	-0.0407	0.0884
	(.0638)	(.0747)	(.1146)
N	1391	736	655
NT	5564	3944	2620

Standard errors in parentheses

E APE estimates

Table 7: APE Estimates accounting for correlated random effects and a binary EEV

	All data							
	Work				Sleep			
percentile	25	50	75	90	25	50	75	90
marital status	-0.0121	-0.0173	-0.0260	-0.0390	-0.1244	-0.1311	-0.1346	-0.1421
	(.0579)	(.0138)	(.0198)	(.0496)	(.0301)	(.0087)	(.0102)	(.0272)
	Men only							
	Work				Sleep			
marital status	-0.2511	-0.2121	-0.1799	-0.1374	-5.38×10^{-7}	-0.0015	-0.0058	-0.0230
	(.0645)	(.0182)	(.0225)	(.0578)	(.0339)	(.0109)	(.0122)	(.0307)
	Women only							
	Work				Sleep			
marital status	-0.3194	-0.0952	-0.0214	-0.0014	-0.0159	-0.0388	-0.0344	-0.0146
	(.0422)	(.0111)	(.0149)	(.0381)	(.0498)	(.0145)	(.0179)	(.0454)

Standard errors in parentheses

Table 8: APE Estimates w/o integrating out Endogenous Error

	All data							
	Work				Sleep			
percentile	25	50	75	90	25	50	75	90
marital status	-2.99×10^{-4}	-1.95×10^{-4}	-1.27×10^{-4}	-1.03×10^{-4}	-7.79×10^{-4}	-8.92×10^{-4}	-2.51×10^{-4}	-3.62×10^{-4}
	(.0195)	(.0039)	(.0115)	(.0118)	(.0067)	(.0013)	(.0028)	(.0034)
	Men only							
	Work				Sleep			
marital status	-0.0011	-7.33×10^{-4}	-5.00×10^{-4}	-4.01×10^{-4}	2.55×10^{-17}	1.31×10^{-35}	1.65×10^{-52}	1.93×10^{-77}
	(.0126)	(.0067)	(.0078)	(.0102)	(.0045)	(.0011)	(.0017)	(.0017)
	Women only							
	Work				Sleep			
marital status	-2.35×10^{-4}	-8.81×10^{-5}	-4.54×10^{-5}	-4.29×10^{-5}	0.0032	3.11×10^{-4}	1.86×10^{-5}	1.75×10^{-6}
	(.0047)	(.0023)	(.0022)	(.0044)	(.0322)	(.0095)	(.0134)	(.0134)

Standard errors in parentheses