BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions

Estimation of Average Marginal Effects in Multiplicative Unobserved Effects Panel Models

Robert S. Martin, U.S. Bureau of Labor Statistics

Working Paper 497
July 2017

Estimation of average marginal effects in multiplicative unobserved effects panel models*

Robert S. Martin†

July 11, 2017

Abstract

In multiplicative unobserved effects panel models for nonnegative dependent variables, estimation of average marginal effects would seem problematic with a large cross section and few time periods due to the incidental parameters problem. While fixed effects Poisson consistently estimates the slope parameters of the conditional mean function, marginal effects generally depend on the unobserved heterogeneity. However, I show that a class of fixed effects averages is consistent and asymptotically normal with only the cross section growing. This implies researchers can estimate average treatment effects in levels as opposed to settling for average proportional effects.

Keywords: Average Treatment Effects, Fixed effects Poisson model; Incidental parameters problem; Panel data; Partial effects;

JEL Codes: C13, C15, C23

*The views expressed herein are those of the author and not necessarily those of the Bureau of Labor Statistics or the U.S. Department of Labor. This is a revised version of the third chapter of my MSU Ph.D. dissertation.

†Division of Price and Index Number Research, Bureau of Labor Statistics, 2 Massachusetts Ave, NE, Washington, DC 20212, USA. Email: Martin.Robert@bls.gov

1 Introduction

The multiplicative effects panel model for nonnegative dependent variables is attractive in part because it is straightforward to handle unobserved cross sectional heterogeneity. Fixed effects Poisson (FEP) consistently estimates the slope parameters of the conditional mean function without full distributional assumptions (Wooldridge, 1999). However, it is not immediately clear how to estimate quantities like average partial effects (APE) and average treatment effects (ATE) as these depend on the unobserved heterogeneity.

I study the use of estimated individual effects from Poisson quasi maximum likelihood estimation (QMLE). There is no incidental parameters problem (IPP) with respect to QMLE slope parameter estimates, which are algebraically equivalent to FEP (Lancaster, 2000). To my knowledge, no one has formally studied estimators of average marginal effects in this model. These estimators potentially suffer from the IPP when each fixed effect is estimated using a relatively small number of observations (Arellano and Hahn, 2007). I show that for the multiplicative model, however, a class of fixed effect averages is consistent and asymptotically normal with only the cross section dimension growing.

For thorough discussions of methods for dealing with the IPP, see Lancaster (2000) and Arellano and Hahn (2007). Empirical researchers, of course, have the option to focus on quantities that do not depend on unobserved heterogeneity. For instance, with an exponential conditional mean function, the slope coefficients can be interpreted as approximate semi-elasticities, and proportional treatment effects are also identified (Lee and Kobayashi, 2001). In my view, however, using estimated fixed effects deserves more attention as average partial effects in levels may be more economically meaningful.

The rest of this paper is organized as follows. Section 2 reviews the model and derives the asymptotic properties of the proposed average marginal effects estimators. Section 3 presents a brief Monte Carlo simulation that suggests good finite sample properties of the estimators. Section 4 concludes.

2 Theory

As in Wooldridge (1999), let $\{(\boldsymbol{y}_i, \boldsymbol{x}_i, c_i),\ i = 1, \ldots\}$ be a sequence of i.i.d. random variables, where $\boldsymbol{y}_i$ is a $T \times 1$ vector of nonnegative dependent variables (not necessarily counts), $\boldsymbol{x}_i \equiv (\boldsymbol{x}'_{i1}, \ldots, \boldsymbol{x}'_{iT})'$, is a $T \times K$ matrix of explanatory variables, and c_i is unobserved scalar heterogeneity that may depend on $\boldsymbol{x}_i$. The multiplicative effects panel model assumes

$$E(y_{it}|\boldsymbol{x}_{it}, c_i) = c_i m(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0),\ t = 1, \ldots T, \tag{1}$$

where $m(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)$ is a known positive function and $\boldsymbol{\beta}_0$ is an unknown $K \times 1$ parameter vector. I also assume that the covariates are strictly exogenous conditional on the unobserved heterogeneity, written as

$$E(y_{it}|\boldsymbol{x}_i, c_i) = E(y_{it}|\boldsymbol{x}_{it}, c_i). \tag{2}$$

The most common choice in the empirical literature is $m(\boldsymbol{x}_{it}, \boldsymbol{\beta}) = \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})$, but the main results of this paper do not require this form. A more flexible option is Wooldridge's (1992) alternative to the Box-Cox transformation. For binary or fractional responses (which also require $0 < c_i < 1$), Wooldridge (1999) suggests the logistic or normal CDF as convenient choices for $m()$.

The fixed effects Poisson (FEP) estimator derives from the nominal assumption that conditional on $\boldsymbol{x}_i$ and c_i, the y_{it} are independently distributed as Poisson with mean given by (1). Conditioning on $n_i \equiv \sum_{t=1}^{T} y_{it}$ yields the multinomial conditional distribution for $\boldsymbol{y}_i$ (Hausman, Hall, and Griliches, 1984). The FEP estimator, denoted $\widehat{\boldsymbol{\beta}}$, solves $\underset{\boldsymbol{\beta} \in \mathfrak{B}}{max} \sum_{i=1}^{N} \ell_i(\boldsymbol{\beta})$, where $\ell_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} y_{it} \log \left[\frac{m(\boldsymbol{x}_{it}, \boldsymbol{\beta})}{\sum_{r=1}^{T} m(\boldsymbol{x}_{ir}, \boldsymbol{\beta})} \right]$ is the multinomial log-likelihood. Wooldridge (1999) showed that consistent estimation of $\boldsymbol{\beta}_0$ only requires (1) and (2), meaning the y_{it} need not be distributed Poisson and may have arbitrary (conditional) serial dependence.

Average marginal effects are often more salient, as $\boldsymbol{\beta}_0$ may not have any meaningful interpretation apart from the exponential case. The APE of a continuous x_j is:

$$\delta_{j,0} = E\left[\frac{\partial E(y_{it}|\boldsymbol{x}_{it}, c_i)}{\partial x_{itj}}\right] = E\left[c_i T^{-1} \sum_{t=1}^{T} \frac{\partial m(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)}{\partial x_{itj}}\right] \equiv E\left[c_i T^{-1} \sum_{t=1}^{T} M_j(\boldsymbol{x}_{it}, \boldsymbol{\beta}_0)\right],$$

where $M_j(\boldsymbol{x}_{it}, \boldsymbol{\beta}) = \partial m(\boldsymbol{x}_{it}, \boldsymbol{\beta}) / \partial x_{itj}$. The ATE for a binary x_k is:

$$\begin{aligned}\delta_{k,0} =&E\left[E(y_{it}|\boldsymbol{x}_{it(-k)}, x_{itk}=1, c_i) - E(y_{it}|\boldsymbol{x}_{it(-k)}, x_{itk}=0, c_i)\right]\\ \equiv&E\left[c_i T^{-1}\sum_{t=1}^{T}\left(m(\boldsymbol{x}_{it(-k)}, 1, \boldsymbol{\beta}_0) - m(\boldsymbol{x}_{it(-k)}, 0, \boldsymbol{\beta}_0)\right)\right]\end{aligned}$$

where the subscript $(-k)$ indicates element k has been omitted, and where $m(\boldsymbol{x}_{it(-k)}, 1, \boldsymbol{\beta})$ and $m(\boldsymbol{x}_{it(-k)}, 1, \boldsymbol{\beta})$ correspond to a 1 or 0 being inserted for x_{itk} in $m(\boldsymbol{x}_{it}, \boldsymbol{\beta})$.

The APE and ATE are examples of fixed effect averages of the form $\boldsymbol{\lambda}_0 = E\left[c_i \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\right]$, where $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta})$ is a $P \times 1$ random function of the covariates. The APE and ATE use $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}) = T^{-1}\sum_{t=1}^{T} M_j(\boldsymbol{x}_{it}, \boldsymbol{\beta})$ and $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}) = T^{-1}\sum_{t=1}^{T}\left(m(\boldsymbol{x}_{it(-k)}, 1, \boldsymbol{\beta}) - m(\boldsymbol{x}_{it(-k)}, 0, \boldsymbol{\beta})\right)$, respectively. The candidate estimator of $\boldsymbol{\lambda}_0$ is given in equation (3). It uses the Poisson QMLE for c_i, denoted $c(\boldsymbol{w}_i, \widehat{\boldsymbol{\beta}})$, when estimating the individual effects along with $\boldsymbol{\beta}_0$.

$$\widehat{\boldsymbol{\lambda}} = N^{-1}\sum_{i=1}^{N} c(\boldsymbol{w}_i, \widehat{\boldsymbol{\beta}})\boldsymbol{h}(\boldsymbol{x}_i, \widehat{\boldsymbol{\beta}}), \tag{3}$$

where $c(\boldsymbol{w}_i, \boldsymbol{\beta}) = n_i / \sum_{t=1}^{T} m(\boldsymbol{x}_{it}, \boldsymbol{\beta})$ and $\boldsymbol{w}_i \equiv \{\boldsymbol{y}_i, \boldsymbol{x}_i\}, i = 1, \ldots, N$. Poisson QMLE and FEP are algebraically equivalent for $\boldsymbol{\beta}_0$, but when N is large, it may be more computationally practical to estimate c_i following FEP estimation of $\boldsymbol{\beta}_0$ (Cameron and Trivedi, 2013).

While it is already known that there is no IPP in this model in terms of estimating $\boldsymbol{\beta}_0$, one should not generally expect averages over estimated incidental parameters to be consistent in non-linear models, even if slope parameter estimates are consistent (Arellano and Hahn, 2007). Clearly $c(\boldsymbol{w}_i, \boldsymbol{\beta}) \neq c_i$, even if evaluated at $\boldsymbol{\beta}_0$, and with T fixed, $c(\boldsymbol{w}_i, \widehat{\boldsymbol{\beta}})$ cannot be consistent for c_i. However, Theorem (1) shows for this model, there is no IPP for fixed effect averages over the cross section like in equation (3).

Theorem 1 *Assume* (1), (2), *and that each element of the* $P \times 1$ *random vector* $\boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\beta}) \equiv c(\boldsymbol{w}_i, \boldsymbol{\beta})\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta})$ *satisfies the regularity conditions on* $q(\boldsymbol{w}_i, \boldsymbol{\beta})$ *from Theorem 12.2 of Wooldridge (2010). Then as* $N \to \infty$,

$$\widehat{\boldsymbol{\lambda}} \xrightarrow{p} \boldsymbol{\lambda}_0$$

Proof. By Lemma 12.1 in Wooldridge (2010), consistency of $\widehat{\boldsymbol{\beta}}$ and the regularity conditions imply

$N^{-1}\sum_{i=1}^{N} c(\boldsymbol{w}_i,\widehat{\boldsymbol{\beta}})\boldsymbol{h}(\boldsymbol{x}_i,\widehat{\boldsymbol{\beta}}) \xrightarrow{p} E\left[c(\boldsymbol{w}_i,\boldsymbol{\beta}_0)\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\right]$. Then, by the Law of Iterated Expectations,

$$
\begin{aligned}
E\left[c(\boldsymbol{w}_i,\boldsymbol{\beta}_0)\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\right] =& E\left\{E\left[c(\boldsymbol{w}_i,\boldsymbol{\beta}_0)\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)|\boldsymbol{x}_i,c_i\right]\right\} \\
=& E\left[\frac{\sum_{t=1}^{T} E(y_{it}|\boldsymbol{x}_i,c_i)}{\sum_{t=1}^{T} m(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\right] \\
=& E\left[\frac{c_i\sum_{t=1}^{T} m(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)}{\sum_{t=1}^{T} m(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)}\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\right] \\
=& E\left[c_i\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}_0)\right]
\end{aligned}
$$

■

A priori, one might expect $\widehat{\boldsymbol{\lambda}}$ to perform well anyway for large enough T, as $c(\boldsymbol{w}_i,\widehat{\boldsymbol{\beta}})$ may approximate c_i better as T grows. The result that $\widehat{\boldsymbol{\lambda}}$ should perform well with as few as two time periods (the minimum needed for FEP), is perhaps less intuitive. Furthermore, consistency of $N^{-1}\sum_{i=1}^{N} c(\boldsymbol{w}_i,\widehat{\boldsymbol{\beta}})$ for $E(c_i)$ follows from setting $\boldsymbol{h}(\boldsymbol{x}_i,\boldsymbol{\beta}) = 1$, but one cannot use $c(\boldsymbol{w}_i,\widehat{\boldsymbol{\beta}})$ to learn about other features of the distribution of c_i except in more restrictive cases. For instance, $Var(c_i)$ is identified only under additional assumptions. A simple example is when the Poisson variance assumption, $Var(y_{it}|\boldsymbol{x}_i,c_i) = E(y_{it}|\boldsymbol{x}_i,c_i)$, and zero conditional covariance, $Cov(y_{it},y_{ir}|\boldsymbol{x}_i,c_i) = 0, t \neq r$, both hold. In this case, one can show that $Var(c_i) = Var\left[c(\boldsymbol{w}_i,\boldsymbol{\beta}_0)\right] - E\left[c_i/\sum_{t=1}^{T} m(\boldsymbol{x}_{it},\boldsymbol{\beta}_0)\right]$.

Asymptotic normality of $\widehat{\boldsymbol{\lambda}}$ follows from a standard argument similar to the delta method, but making sure to account for the randomness in $\boldsymbol{w}_i$. The asymptotic variance formula in Theorem (2) uses that $Avar\left[\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right] = \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}$, where $\boldsymbol{A}_0 = -E\left[\nabla_{\beta}^2\ell_i(\boldsymbol{\beta}_0)\right]$, $\boldsymbol{B}_0 = Var\left[\boldsymbol{s}_i(\boldsymbol{\beta}_0)\right]$, and $\boldsymbol{s}_i(\boldsymbol{\beta}_0) = \nabla_{\boldsymbol{\beta}}\ell_i(\boldsymbol{\beta}_0)'$ (Wooldridge, 1999).

Theorem 2 *Under the assumptions in Theorem* (1), *as* $N \to \infty$,

$$\sqrt{N}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0) \xrightarrow{d} N(\boldsymbol{0},\boldsymbol{D}_0),$$

where

$$\boldsymbol{D}_0 = Var\left[\boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\beta}_0) - \boldsymbol{\lambda}_0 - \boldsymbol{G}_0 \boldsymbol{A}_0^{-1} \boldsymbol{s}_i(\boldsymbol{\beta}_0)\right],$$

$$\boldsymbol{G}_0 = E\left[\nabla_{\boldsymbol{\beta}} \boldsymbol{g}(\boldsymbol{w}_i, \boldsymbol{\beta}_0)\right] = E\left[c(\boldsymbol{w}_i, \boldsymbol{\beta}_0)\nabla_{\boldsymbol{\beta}} \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) + \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\nabla_{\boldsymbol{\beta}} c(\boldsymbol{w}_i, \boldsymbol{\beta}_{\mathbf{0}})\right],$$

$$\nabla_{\boldsymbol{\beta}} c(\boldsymbol{w}_i, \boldsymbol{\beta}) = -c(\boldsymbol{w}_i, \boldsymbol{\beta}) \left(\frac{\sum_{t=1}^{T} \nabla_{\boldsymbol{\beta}} m(\boldsymbol{x}_{it}, \boldsymbol{\beta})}{\sum_{t=1}^{T} m(\boldsymbol{x}_{it}, \boldsymbol{\beta})} \right),$$

$\nabla_{\boldsymbol{\beta}} \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta})$ *is the* $P \times K$ *Jacobian of* $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta})$*, and*

$\nabla_{\boldsymbol{\beta}} m(\boldsymbol{x}_{it}, \boldsymbol{\beta})$ *is the* $1 \times K$ *gradient of* $m(\boldsymbol{x}_{it}, \boldsymbol{\beta})$*.*

The proof is similar to the solution to Problem 12.17 of Wooldridge (2010), and is available from the author upon request. Consistent estimation of $\boldsymbol{D}_0$ involves plugging in $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ and replacing the expectations and variances with their sample analogs.

2.1 Exponential Models

The exponential conditional mean case is particularly interesting given its prevalence in empirical research. For example, Lee and Kobayashi (2001) use an exponential model estimate the average proportional change in health care demand from a treatment (exercise). By construction, the average proportional effect does not depend on the heterogeneity or the coefficients on time-constant regressors. While it may be interesting in its own right, my analysis implies the APE and ATE of time-varying regressors are still identified even if the population model includes time-constant variables. To see this, suppose the following:

$$E(y_{it}|\boldsymbol{x}_{it}, \boldsymbol{z}_i, v_i) = v_i \exp(\boldsymbol{x}_{it}\boldsymbol{\beta}_0 + \boldsymbol{z}_i\boldsymbol{\gamma}_0), \tag{4}$$

where $\boldsymbol{x}_{it}$ is time-varying, $\boldsymbol{z}_i$ is time-constant, and v_i denotes the unobserved heterogeneity. Defining $c_i \equiv v_i \exp(\boldsymbol{z}_i\boldsymbol{\gamma}_0)$, equation (4) is equivalent to the exponential case of equation (1). Now c_i represents the total contribution from all time-constant variables—observed and unobserved. Even though $\boldsymbol{\gamma}_0$ is not identified, $\boldsymbol{\lambda}_0$ still is as long as $\boldsymbol{h}()$ does not depend on $\boldsymbol{\gamma}_0$, which is true for the average marginal effects of the $\boldsymbol{x}_{it}$.

Furthermore, in the exponential case with linear index, the APE of a continuous variable is given by the following:

$$\delta_{j,0} = E\left[\frac{\partial E(y_{it}|\boldsymbol{x}_{it}, c_i)}{\partial x_{itj}}\right] = E\left[T^{-1}\sum_{t=1}^{T} c_i \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})\right]\beta_{j,0} = \left[T^{-1}\sum_{t=1}^{T} E(y_{it})\right]\beta_{j,0}, \quad (5)$$

where the last equality is by the Law of Iterated Expectations. Equation (5) shows the APE does not actually depend on c_i and looks similar to the cross section case. Therefore, one can actually estimate $\boldsymbol{\beta}_0$ with weakly exogenous $\boldsymbol{x}_{it}$ by using sequential moment restrictions as in Chamberlain (1992) or Wooldridge (1997).

3 Monte Carlo

To test the theoretical results in this paper, I run a Monte Carlo simulation of a simple count model with endogenous regressors due to unobserved heterogeneity. Elements of the data generating process resemble simulations in Greene (2004) and Fernandez-Val and Weidner (2016).

3.1 Design

For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, I generate the data as:

$$y_{it}|(\boldsymbol{x}_i, \boldsymbol{d}_i, c_i) \sim \text{Poisson}\left[c_i \exp(\beta_1 x_{it} + \beta_2 d_{it})\right],$$

$$\log(c_i) \sim Normal(0, 1/2)$$

$$x_{it} = \log(c_i) + \rho x_{i,t-1} + v_{it},\ t > 1$$

$$x_{i1} = \log(c_i)/(1-\rho) + v_{i1}/\sqrt{1-\rho^2},\ \rho = 0.3,\ v_{it} \sim N(0, 1/2),$$

$$d_{it} = \mathbf{1}\left[x_{it} + \log(c_i) + h_{it} > 0\right],\ h_{it} \sim N(0, 1/2)$$

The conditional marginal distribution of y_{it} is Poisson with an exponential mean function. I set $\beta_1 = 0.5$ and $\beta_2 = -0.5$, magnitudes similar to estimates from Hausman, et. al. (1984). The continuous covariate x and the binary covariate d are both correlated with the heterogeneity. The scaling of x_{i1} is intended to keep $Var(x_t)$ constant across different values of T (Vamoş, Şoltuz, and Crăciun, 2007). I study panels of dimensions $N = 2000$ and $T \in \{2, 4, 10\}$ to mimic the large-N, small T setting typical in microeconometrics. I draw 2000 replications.

For each draw, I estimate β_1 and β_2 using FEP, calculate $c(\boldsymbol{w}_i, \widehat{\boldsymbol{\beta}})$, and estimate the APE and ATE as described in Section 2. In the results to follow, I denote APE estimate as $\widehat{\delta}_1$ and the ATE estimate as $\widehat{\delta}_2$. True values were estimated from a single draw with $N = 10,000,000$. Though not reported, I simulated the performance of pooled Poisson QMLE that ignores c_i entirely and found the APE and ATE estimates to have a positive bias of roughly 80% and 55%, respectively. The asymptotic standard errors are of the form derived in Theorem 2 and based on cluster robust standard errors for $\widehat{\boldsymbol{\beta}}$, though technically for this application one could use the nonrobust versions.

3.2 Results

Table 1 reports the results of the simulation exercise. Reported are the mean and standard deviation of the empirical distribution, the estimated bias, the ratio of the mean standard error to the empirical standard deviation, and the probability of rejecting a true null hypothesis at the five percent significance level. Finite sample bias in $\widehat{\delta}_1$ is less than 0.005 in magnitude for each value of T. Bias in the ATE estimates is also very small—0.01 or less. The finite sample standard deviations behave in a predictable way, decreasing as T increases. The variability in $\widehat{\delta}_2$ seems to be greater than that of $\widehat{\delta}_1$, which might be related to the fact that $\widehat{\delta}_1$ does not actually use $c(\boldsymbol{w}_i, \widehat{\boldsymbol{\beta}})$ in the exponential case. The asymptotic standard errors derived in this paper perform reasonably well. At worst, in the $T = 4$ case, the standard error appears to underestimate the empirical standard deviation of $\widehat{\delta}_2$ by about 3 percent. Overall, these simulations support this paper's theoretical findings. The asymptotic properties derived in Section 2 for the APE and ATE estimators that use estimated incidental parameters seem to approximate their finite sample behavior quite well.

Table 1: Fixed Effects Poisson Average Marginal Effect Estimators

	$\widehat{\delta}_1$ (APE)					$\widehat{\delta}_2$ (ATE)				
	Mean	Bias	SD	SE/SD	RP(0.05)	Mean	Bias	SD	SE/SD	RP(0.05)
$T = 2$	0.73	0.00	0.06	1.01	0.05	-0.88	0.00	0.17	1.01	0.05
$T = 4$	0.73	0.00	0.04	0.97	0.06	-0.88	-0.01	0.10	1.00	0.05
$T = 10$	0.73	0.00	0.03	0.98	0.05	-0.88	0.00	0.07	0.98	0.05

4 Conclusion

It is already well-known that in static multiplicative panel models under strict exogeneity, estimating the heterogeneity still leads to consistent estimation of the slope parameters of the conditional mean function. This paper adds the result that APE and ATE estimators that use estimated heterogeneity are consistent and $\sqrt{N}$-asymptotically normal. In fact, the results hold for estimating the mean of a wider class of random quantities where the heterogeneity is multiplicatively separable from functions of the data. I also derive asymptotic standard errors for these estimators that perform well in simulations.

References

Arellano, M. and J. Hahn (2007). *Understanding Bias in Nonlinear Panel Models: Some Recent Developments. In Advances in Economics and Econometrics, Blundell R, Newey W, Persson T (eds)*. Cambridge: Cambridge University Press.

Cameron, A.C. and P.K. Trivedi (2013). *Regression analysis of count data*. 2nd ed. Cambridge University Press.

Chamberlain, G. (1992). "Comment: Sequential moment restrictions in panel data". In: *Journal of Business & Economic Statistics* 10.1, pp. 20–26.

Fernández-Val, Iván and Martin Weidner (2016). "Individual and time effects in nonlinear panel models with large N, T". In: *Journal of Econometrics* 192.1, pp. 291–312.

Greene, W.H. (2004). "The Behavior of the Fixed Effects Estimator in Nonlinear Models". In: *The Econometrics Journal* 7, pp. 98–119. DOI: 10.1111/j.1368-423x.2004.00123.x.

Hausman, J., B. Hall, and Z. Griliches (1984). "Econometric Models for Count Data with an Application to the Patents-R&D Relationship". In: *Econometrica* 52.4, p. 909. DOI: 10.2307/1911191.

Lancaster, T. (2000). "The Incidental Parameters Problem since 1948". In: *Journal of Econometrics* 95, pp. 391–413. DOI: 10.1016/s0304-4076(99)00044-5.

Lee, M. and S. Kobayashi (2001). "Proportional treatment effects for count response panel data: effects of binary exercise on health care demand". In: *Health Economics* 10.5, pp. 411–428.

Vamoş, C., Ş. Şoltuz, and M. Crăciun (2007). "Order 1 autoregressive process of finite length". In: *Rev. Anal. Numér. Théor. Approx.* 36.2, pp. 199–214.

Wooldridge, J.M. (1997). "Multiplicative panel data models without the strict exogeneity assumption". In: *Econometric Theory* 13.05, pp. 667–678.

Wooldridge, J.M. (1999). "Distribution-free estimation of some nonlinear panel data models". In: *Journal of Econometrics* 90.1, pp. 77–97. DOI: 10.1016/s0304-4076(98)00033-5.