Economics Letters 160 (2017) 16–19

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Estimation of average marginal effects in multiplicative unobserved effects panel models

Robert S. Martin

Division of Price and Index Number Research, Bureau of Labor Statistics, 2 Massachusetts Ave, NE, Washington, DC 20212, USA

HIGHLIGHTS

- I show consistency of some fixed effects averages in the F.E. Poisson setting.
- This implies average marginal effects are estimable in levels, not just proportions.
- I derive the asymptotic variance for this class of estimators.

ARTICLE INFO

Article history:
Received 11 July 2017
Received in revised form 4 August 2017
Accepted 18 August 2017
Available online 23 August 2017

JEL classification:
C13
C23

Keywords:
Average treatment effects
Fixed effects Poisson model
Incidental parameters problem
Panel data
Partial effects

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.

Published by Elsevier B.V.

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 the 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 also 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 some observations about exponential models. Section 4 applies the estimators to data on patents, while Section 5 concludes. An online

E-mail address: Martin.Robert@bls.gov.

http://dx.doi.org/10.1016/j.econlet.2017.08.020
0165-1765/Published by Elsevier B.V.

appendix (see Appendix A) contains a brief Monte Carlo experiment showing good performance of the proposed estimators.

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 strict exogeneity 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 the Wooldridge (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 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 $\sum_{t=1}^T y_{it}$ yields the multinomial conditional distribution for $\boldsymbol{y}_i$ (Hausman et al., 1984). The FEP estimator, denoted $\hat{\boldsymbol{\beta}}$, solves $\max_{\boldsymbol{\beta}} \sum_{i=1}^N \ell_i(\boldsymbol{\beta})$, where $\ell_i(\boldsymbol{\beta}) = \sum_{t=1}^T y_{it} \ln\left[m(\boldsymbol{x}_{it}, \boldsymbol{\beta})/\sum_{r=1}^T m(\boldsymbol{x}_{ir}, \boldsymbol{\beta})\right]$ 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 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:

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

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. For example, the APE uses $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}) = T^{-1}\sum_{t=1}^T M_j(\boldsymbol{x}_{it}, \boldsymbol{\beta})$. The candidate estimator of $\boldsymbol{\lambda}_0$, given in Eq. (3), uses the Poisson QMLE for c_i, denoted $c(\boldsymbol{w}_i, \hat{\boldsymbol{\beta}})$, when these are estimated along with $\boldsymbol{\beta}_0$.

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

where $c(\boldsymbol{w}_i, \boldsymbol{\beta}) = \sum_{t=1}^T y_{it}/\sum_{t=1}^T m(\boldsymbol{x}_{it}, \boldsymbol{\beta})$ 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, 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, $c(\boldsymbol{w}_i, \hat{\boldsymbol{\beta}})$ 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 Eq. (3).

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

$$\hat{\boldsymbol{\lambda}} \xrightarrow{p} \boldsymbol{\lambda}_0.$$

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

$$N^{-1}\sum_{i=1}^N c(\boldsymbol{w}_i, \hat{\boldsymbol{\beta}})\boldsymbol{h}(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}) \xrightarrow{p} E\left[c(\boldsymbol{w}_i, \boldsymbol{\beta}_0)\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\right].$$

Then, by the Law of Iterated Expectations,

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

A priori, one might expect $c(\boldsymbol{w}_i, \hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\lambda}}$ to perform well anyway for larger T. The result that $\hat{\boldsymbol{\lambda}}$ should perform well with as few as two time periods (the minimum needed for FEP), is perhaps less intuitive. Furthermore, consistency of $N^{-1}\sum_{i=1}^N c(\boldsymbol{w}_i, \hat{\boldsymbol{\beta}})$ for $E(c_i)$ follows from setting $\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}) = 1$, but using $c(\boldsymbol{w}_i, \hat{\boldsymbol{\beta}})$ to learn about $Var(c_i)$ or other features of the distribution of c_i requires more assumptions.

Asymptotic normality of $\hat{\boldsymbol{\lambda}}$ 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 $Avar\left[\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] = \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}$, where $\boldsymbol{A}_0 = -E\left[\nabla_{\boldsymbol{\beta}}^2 \ell_i(\boldsymbol{\beta}_0)\right]$, $\boldsymbol{B}_0 = Var\left[\boldsymbol{s}_i(\boldsymbol{\beta}_0)\right]$, 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* $N \to \infty$,

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

where

$$\begin{aligned}
\boldsymbol{D}_0 &= Var\left[\mathbf{g}(\boldsymbol{w}_i, \boldsymbol{\beta}_0) - \boldsymbol{\lambda}_0 - \boldsymbol{G}_0\boldsymbol{A}_0^{-1}\boldsymbol{s}_i(\boldsymbol{\beta}_0)\right],\\
\boldsymbol{G}_0 &= E\left[\nabla_{\boldsymbol{\beta}}\mathbf{g}(\boldsymbol{w}_i, \boldsymbol{\beta}_0)\right]\\
&= E\left[c(\boldsymbol{w}_i, \boldsymbol{\beta}_0)\nabla_{\boldsymbol{\beta}}\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}_0) + \boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\nabla_{\boldsymbol{\beta}}c(\boldsymbol{w}_i, \boldsymbol{\beta}_0)\right],
\end{aligned}$$

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

$\nabla_{\boldsymbol{\beta}}\boldsymbol{h}(\boldsymbol{x}_i, \boldsymbol{\beta})$ *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* $m(\boldsymbol{x}_{it}, \boldsymbol{\beta})$.