# BIAS IN INSTRUMENTAL-VARIABLE ESTIMATORS OF FIXED-EFFECT MODELS FOR COUNT DATA

KOEN JOCHMANS*

TOULOUSE SCHOOL OF ECONOMICS, UNIVERSITY OF TOULOUSE CAPITOLE

This version: October 29, 2021

**Abstract**

This note looks at the properties of instrumental-variable estimators of models for non-negative outcomes in the presence of individual effects. We show that fixed-effect versions of the estimators of Mullahy (1997) and Windmeijer and Santos Silva (1997) are inconsistent under conventional asymptotics, in general, and that inference based on them in long panels requires bias correction. Such corrections are derived and their effectiveness is investigated in numerical experiments. Consistent estimation in short panels is nonetheless possible in the setting underlying Mullahy's (1997) approach using a differencing strategy along the lines of Wooldridge (1997) and Windmeijer (2000).

**Keywords:** count data, bias, fixed effects, inconsistency, instrumental variable, multiplicative-error model, poisson.

**JEL classification:** C23, C26.

# Introduction

The pseudo-poisson maximum-likelihood estimator is routinely used for the purpose of analysing non-negative outcomes. It is consistent under a conditional-mean specification only (Gouriéroux, Monfort and Trognon, 1984a,b) and is well-known to possess a certain robustness against the inclusion of fixed effects (see Wooldridge 1999 and Fernández-Val

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.
Support from the European Research Council through grant nº 715787 (MiMo), and from the French Government and the ANR under the Investissements d' Avenir program, grant ANR-17-EURE-0010 is gratefully acknowledged.

and Weidner 2016). This may lead to the presumption that the instrumental-variable generalizations of the pseudo-poisson estimator, too, are unaffected by the presence of such incidental parameters; it has been used by Tenreyro (2007) and Haucap, Rasch and Stiebale (2019), for example, without any reference to the incidental-parameter problem. This is, however, not the case. This note provides details that underlie this conclusion for the (one-way) panel data case.

Fixed-effect versions of two instrumental-variable estimators are looked at. The first such estimator is the one proposed by Mullahy (1997). The second is the one proposed by Windmeijer and Santos Silva (1997). These estimators are based on different orthogonality conditions. Furthermore, these conditions are not compatible with one another, a point already made by Windmeijer and Santos Silva (1997). We show that both estimators are inconsistent under classical asymptotics that treat the length of the panel as fixed, and asymptotically biased under sequences where both dimensions of the panel grow at the same rate. This bias can be corrected for, either by relying on analytical formulae given below or by the application of a jackknife.

The moment conditions underlying the estimator of Mullahy (1997) can be modified to yield an alternative estimator that is consistent in short panels. The same would not appear to be true for the poisson type fixed-effect estimator of Windmeijer and Santos Silva (1997).

# 1 Panel model and fixed-effect estimators

We observe a scalar outcome $y_{it}$, a regressor vector $x_{it}$, and a vector of instruments $z_{it}$ for a random sample of $N$ individuals, $i = 1, \ldots, N$, that are followed over $T$ periods of time, $t = 1, \ldots, T$. Our ambition is to estimate the parameter $\theta$ in a multiplicative model of the form

$$y_{it} = \alpha_i \, \lambda(x_{it}, \theta) \, v_{it},$$

where $\lambda$ is a known function, the $\alpha_i$ are latent variables that capture any heterogeneity across individuals that does not vary over time, and $v_{it}$ is a time-varying disturbance term.

We will treat the $\alpha_i$ as fixed effects. Hence, all expectations below are to be understood as being conditional on $\alpha_1, \ldots, \alpha_N$. We will write $x_i := (x_{i1}, \ldots, x_{iT})$ and $z_i := (z_{i1}, \ldots, z_{iT})$.

**Multiplicative error** An instrumental-variable estimator based on $\mathbb{E}(v_{it}|z_i) = 1$, as in Mullahy (1997), involves the empirical moment condition

$$\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it} \left( \frac{y_{it}}{\alpha_i \, \lambda(x_{it}, \theta)} - 1 \right) = 0,$$

and the the empirical moment conditions for the individual effects. The latter follow from $\mathbb{E}(v_{it}) = 1$ and are

$$\sum_{t=1}^{T} \left( \frac{y_{it}}{\alpha_i \, \lambda(x_{it}, \theta)} - 1 \right) = 0,$$

for each individual.

**Additive error** The model can be reformulated as

$$y_{it} = \alpha_i \, \lambda(x_{it}, \theta) + u_{it},$$

with $u_{it} := \alpha_i \, \lambda(x_{it}, \theta) \, (v_{it} - 1)$. An instrumental-variable estimator based on the assumption that $\mathbb{E}(u_{it}|z_i) = 0$, as in Windmeijer and Santos Silva (1997), is constructed around the empirical moment condition

$$\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it} \, (y_{it} - \alpha_i \, \lambda(x_{it}, \theta)) = 0,$$

together with

$$\sum_{t=1}^{T} (y_{it} - \alpha_i \, \lambda(x_{it}, \theta)) = 0$$

for each individual. Clearly, this is an instrumental-variable version of the pseudo poisson estimator.

# 2 Large-sample behavior

As discussed by Mullahy (1997) and Windmeijer and Santos Silva (1997), the conditions $\mathbb{E}(v_{it}|z_i) = 1$ and $\mathbb{E}(u_{it}|z_i) = 0$ are not compatible, in general. The exception is when the regressors are exogenous. Hence, the multiplicative and additive specifications yield (different) estimators based on different conditional validity conditions that will usually not both be valid at the same time. In the context of simultaneous equations the former appears more natural (see, e.g., Mullahy 1997 for a discussion). We treat them separately, in turn.

**Multiplicative error** It is useful to work with a concentrated estimating equation for $\theta$. For a given value $\vartheta$ of the common parameter the corresponding estimator of $\alpha_i$ is available in closed form. It equals

$$\hat{\alpha}_i(\vartheta) := \frac{1}{T}\sum_{t=1}^{T} \frac{y_{it}}{\lambda(x_{it}, \vartheta)}.$$

Substituting this back into the moment condition for $\theta$ then yields the profiled estimating equation

$$\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}\left(\frac{y_{it}}{\hat{\alpha}_i(\vartheta)\,\lambda(x_{it}, \vartheta)} - 1\right) = 0.$$

This estimating equation is biased, in general. Indeed, letting $\bar{v}_i := T^{-1}\sum_{t=1}^{T} v_{it}$, its expectation at $\theta$ is equal to

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left(z_{it}\,\mathbb{E}\left(\frac{v_{it}}{\bar{v}_i} - 1 \middle| z_i\right)\right).$$

This will be non-zero, in general, unless $\mathbb{E}(v_{it}/\bar{v}_i|z_i) = 1$. So, the no-bias condition amounts to conditions on the distributions of the $v_{it}|z_i$ that guarantee that the expectation of a ratio is equal to the ratio of expectations. One situation in which this will be the case is when $v_{it}|z_i$ is i.i.d. over time.[1]

The presence of a non-vanishing bias in the estimating equation for $\theta$ as $N \to \infty$ with $T$ held fixed means that the corresponding instrumental-variable estimator is inconsistent for $\theta$ under such asymptotics.

[1] In this case $v_{it}/\bar{v}_i$ and $v_{ij}/\bar{v}_i$ have the same distribution (and hence the same mean), conditional on $z_i$. As this holds for all $j$, we can average to find that $\mathbb{E}(v_{it}/\bar{v}_i|z_i) = \frac{1}{T}\sum_{j=1}^{T}\mathbb{E}(v_{ij}/\bar{v}_i|z_i) = \mathbb{E}(\bar{v}_i/\bar{v}_i|z_i) = 1$, as claimed.

The bias does decrease as $T \to \infty$. Observe that $\hat{\alpha}_i := \hat{\alpha}_i(\theta)$ is (conditionally) unbiased for the individual effect $\alpha_i$. Furthermore, it is asymptotically linear, as $T \to \infty$, with representation

$$\hat{\alpha}_i - \alpha_i = \frac{1}{T} \sum_{t=1}^{T} \alpha_i \left(v_{it} - 1\right) + o_p(T^{-1}).$$

A second-order Taylor expansion of the profiled moment condition around the true $\alpha_i$ and evaluating at $\theta$ yields

$$\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \, \left(v_{it} - 1\right) - \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{z_{it} v_{it}}{\alpha_i} \left(\hat{\alpha}_i - \alpha_i\right) + \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{z_{it} v_{it}}{\alpha_i^2} \left(\hat{\alpha}_i - \alpha_i\right)^2 + o_p(N).$$

The first part corresponds to the infeasible estimating equation for $\theta$ in which the individual effects are known and has mean zero. The remaining two terms contribute bias to the feasible equation. If we plug-in the linear representation for $\hat{\alpha}_i - \alpha_i$ and take expectations the bias is found to be

$$-\sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \mathbb{E}(z_{it} v_{it} (v_{it} - 1))}{T} + \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \mathbb{E}(z_{it} v_{it})}{T} \, \frac{\sum_{t=1}^{T} \mathbb{E}(v_{it} (v_{it} - 1))}{T} + o(N),$$

where we have assumed (conditional) independence of the errors over time for simplicity of exposition.

The bias in the moment condition translates into a bias of order $T^{-1}$ in the estimator. Consequently, while consistent as $N, T \to \infty$, $\hat{\theta}$ will be asymptotically-biased if $N$ and $T$ grow at the same rate. This bias can be estimated and removed, either through the use of a plug-in estimator of the bias based on the analytical formula above, or through a jackknife. Doing so will recenter the limit distribution around zero and lead to asymptotically-valid inference.

**Additive errors** Here, on recycling notation,

$$\hat{\alpha}_i(\vartheta) := \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \lambda(x_{it}, \vartheta)},$$

and the profiled estimating equation for $\theta$ becomes

$$\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \left(y_{it} - \hat{\alpha}_i(\vartheta) \, \lambda(x_{it}, \vartheta)\right) = 0.$$

When the regressors are strictly exogenous, i.e., when $\mathbb{E}(u_{it}|x_i) = 0$ it is easily seen that $\mathbb{E}(\hat{\alpha}_i|x_i) = \alpha_i$. From this, then, unbiasedness of the profiled estimating equation follows, leading to consistency under classical asymptotics. More generally, however, unbiasedness would require that

$$\mathbb{E}\left( \left. \frac{\sum_{t=1}^T \mathbb{E}(u_{it}|x_i, z_i)}{\sum_{t=1}^T \lambda(x_{it}, \theta)} \right| z_i \right) = 0$$

and this will not be true, in general.

We can again characterize the leading bias in the estimating equation. Afirst-order expansion around the $\alpha_i$ now suffices as, here, the estimating equation is linear in the individual effects. On the other hand, the estimator of the individual effects is now biased, and this bias has to be accounted for. As $T \to \infty$, we have, again using (conditional) independence over time,

$$\hat{\alpha}_i - \alpha_i = -\frac{1}{T^2} \sum_{t=1}^T \frac{\mathbb{E}(\lambda_{it} u_{it})}{\mathbb{E}(\lambda_i)^2} + \frac{1}{T} \sum_{t=1}^T \frac{u_{it}}{\mathbb{E}(\lambda_i)} + o_p(T^{-1}),$$

where $\lambda_{it} := \lambda(x_{it}, \theta)$ and $\lambda_i := T^{-1} \sum_{t=1}^T \lambda_{it}$; this follows from standard higher-order asymptotics (Bao and Ullah, 2007). Together with the expansion

$$\sum_{i=1}^N \sum_{t=1}^T z_{it}\,(y_{it} - \hat{\alpha}_i\, \lambda_{it}) = \sum_{i=1}^N \sum_{t=1}^T z_{it} u_{it} - \sum_{i=1}^N \sum_{t=1}^T z_{it} \lambda_{it}\,(\hat{\alpha}_i - \alpha_i)$$

this yields the bias as

$$\sum_{i=1}^N \frac{\sum_{t=1}^T \mathbb{E}(z_{it}\lambda_{it})/\mathbb{E}(\lambda_i)}{T} \, \frac{\sum_{t=1}^T \mathbb{E}(u_{it}\lambda_{it})/\mathbb{E}(\lambda_i)}{T} - \sum_{i=1}^N \frac{\sum_{t=1}^T \mathbb{E}(u_{it} z_{it} \lambda_{it})/\mathbb{E}(\lambda_i)}{T} + o(N).$$

Proceeding as before yields the same conclusions about the large-sample behavior of the instrumental-variable estimator as in the multiplicative case. A bias correction to the estimator may again be constructed.

# 3 An alternative estimator

In the multiplicative specification $\mathbb{E}(v_{it}|z_i) = 1$ implies that $\mathbb{E}(y_{it}/\lambda_{it}|z_i) = \alpha_i$. Therefore,

$$\mathbb{E}\left( \left. \frac{y_{it}}{\lambda(x_{it}, \theta)} - \frac{y_{it-1}}{\lambda(x_{it-1}, \theta)} \right| z_i \right) = 0$$

holds for all $t > 1$. This leads to unconditional moments in the spirit of Wooldridge (1997) and Windmeijer (2000) that are free of incidental parameters, paving the way for consistent estimation from short panels. An example of such an estimator would be the solution to the estimating equation

$$\sum_{i=1}^{N} \sum_{2=1}^{T} z_{it} \left( \frac{y_{it}}{\lambda(x_{it}, \vartheta)} - \frac{y_{it-1}}{\lambda(x_{it-1}, \vartheta)} \right) = 0,$$

which is in line with the moment conditions used previously. More generally, optimal (unconditional) moment conditions can be constructed in the usual way (Chamberlain, 1987). Under regularity conditions the implied estimator will be $N^{-1/2}$-consistent and asymptotically normal as $N \to \infty$ with $T$ fixed. This approach can equally be used in a setting characterized by sequential moment conditions on the form $\mathbb{E}(v_{it}|z_{i1}, \ldots, z_{it}) = 1$. In this case

$$\mathbb{E} \left( \frac{y_{it}}{\lambda(x_{it}, \theta)} - \frac{y_{it-1}}{\lambda(x_{it-1}, \theta)} \middle| z_{i1}, \ldots, z_{it-1} \right) = 0$$

follows from iterating expectations. In both cases, our formulation allows for unrestricted serial dependence in the errors.

For the additive specification where $\mathbb{E}(u_{it}|z_i) = 0$ a differencing strategy would not appear to be available.

# 4 Simulations

We generated outcomes using an exponential link function without individual effects, i.e.,

$$y_{it} = \exp(x_{it}\theta)\, v_{it},$$

setting $\theta = 1$. The regressor and the instrument were both binary random variables with

$$\mathbb{P}(x_{it} = 1|z_{it} = 0) = .80, \qquad \mathbb{P}(x_{it} = 1|z_{it} = 1) = .30,$$

and

$$\mathbb{P}(z_{it} = 1) = .30.$$

Finding a simple data generating process for the outcome that satisfies the restrictions of the additive specification of Windmeijer and Santos Silva (1997) is not a simple task. We therefore proceeded as follows. We drew multiplicative errors $v_{it}$ from log-normal distributions depending on the values of the regressor and instrument. To satisfy the condition $\mathbb{E}(v_{it}|z_i) = 1$ we set

$$\mathbb{E}(v_{it}|x_{it}, z_{it}) = \begin{cases} 4.4817 & \text{if } x_{it} = 0 \text{ and } z_{it} = 0 \\ 0.1296 & \text{if } x_{it} = 1 \text{ and } z_{it} = 0 \\ 0.2231 & \text{if } x_{it} = 0 \text{ and } z_{it} = 1 \\ 2.8127 & \text{if } x_{it} = 1 \text{ and } z_{it} = 1 \end{cases}$$

in a first set of experiments. This is the multiplicative model. To satisfy $\mathbb{E}(u_{it}|z_i) = 0$, in turn, we set

$$\mathbb{E}(v_{it}|x_{it}, z_{it}) = \begin{cases} 4.4817 & \text{if } x_{it} = 0 \text{ and } z_{it} = 0 \\ 0.6798 & \text{if } x_{it} = 1 \text{ and } z_{it} = 0 \\ 0.2231 & \text{if } x_{it} = 0 \text{ and } z_{it} = 1 \\ 1.6669 & \text{if } x_{it} = 1 \text{ and } z_{it} = 1 \end{cases}$$

in a second set of experiments. This is the additive model. Tables 1 and 2 contain results for the respective designs for different samples sizes.

Table 1 reports the mean, the standard deviation, as well as the coverage rate of 95% confidence intervals (as obtained over 5000 Monte Carlo replications) for four estimators. The first, EXO, is the fixed-effect estimator of Mullahy (1997) with $x_{it}$ instrumenting for itself. The second, IV, use $z_{it}$ as an instrument for $x_{it}$. The third, BC, is the (analytically) bias-corrected version of IV using the formula derived above. (The split-panel jackknife of Dhaene and Jochmans (2015) gave very similar results.) Finally, the fourth, DIFF, is the differencing estimator described in the previous section.[2] In all cases coverage rates were computed using the nonparametric percentile bootstrap (based on 99 bootstrap samples), resampling the cross-sectional units. Doing so is particularly useful for the fixed-effect

[2] As the covariate is non-negative in our example, the empirical moment condition as stated approaches zero as $\theta \to \infty$. Evaluating them at $x_{it} - \overline{x}$, where $\overline{x}$ is the overall sample mean of the $x_{it}$, in stead of at $x_{it}$ resolves this problem. Also see Windmeijer (2000).

estimators, for which we observed that a plug-in estimator of the asymptotic variance tends to be inaccurate for relatively small values of $T$. This is in line with observations made elsewhere (see, e.g., Jochmans 2017).

Table 2 has the same structure as Table 1 but concerns the additive specification. Hence, here, all of EXO, IV, and BC are in reference to the Windmeijer and Santos Silva (1997) moment conditions. Results for a DIFF estimator are not reported as such an approach is not available here.

The results show that EXO is heavily biased and not useful for inference in any of the cases considered. This is, of course, expected as it is based on invalid moment conditions. The IV estimator is biased but consistent as $N, T \to \infty$. This is borne out in our simulations as the bias and standard deviation both shrink in larger samples. However, the bias is important relative to the standard deviation and so hypothesis tests will be size distorted. In our particular design, this is particularly visible here in Table 1 and less so in Table 2, although it would start becoming more visible also there as the cross-sectional dimension would increase. The corrected estimator, BC, removes most of the bias from IV, re-centering its sampling distribution. For the multiplicative specification in Table 1 we also observe that DIFF is well behaved for all configurations.

# Conclusion

This note has highlighted difficulties with instrumental-variable estimators for count data in the presence of fixed effects. The problem can be rectified in long panels (i.e., under rectangular-array asymptotics) and we have shown how to do so. A differencing strategy for short panel data has also been proposed.

Table 1: Simulations for multiplicative-error model

| | | MEAN | | | | STD | | | | COVERAGE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $T$ | EXO | IV | BC | DIFF | EXO | IV | BC | DIFF | EXO | IV | BC | DIFF |
| 200 | 20 | -0.5242 | 1.3827 | 1.0683 | 1.0063 | 0.0817 | 0.1893 | 0.1407 | 0.1823 | 0 | 0.4520 | 0.9150 | 0.9440 |
| 200 | 40 | -0.4591 | 1.1484 | 1.0249 | 0.9998 | 0.0598 | 0.1028 | 0.0946 | 0.1323 | 0 | 0.6650 | 0.9400 | 0.9330 |
| 200 | 60 | -0.4240 | 1.0985 | 1.0165 | 1.0032 | 0.0507 | 0.0791 | 0.0763 | 0.1106 | 0 | 0.7260 | 0.9330 | 0.9340 |
| 200 | 80 | -0.4092 | 1.0735 | 1.0107 | 1.0023 | 0.0436 | 0.0655 | 0.0651 | 0.0933 | 0 | 0.7670 | 0.9180 | 0.9230 |
| 200 | 100 | -0.3995 | 1.0558 | 1.0051 | 1.0009 | 0.0386 | 0.0586 | 0.0580 | 0.0846 | 0 | 0.8130 | 0.9320 | 0.9400 |

EXO: instrumental-variable estimator of Mullahy (1997) instrumenting $x_{it}$ by itself. IV: instrumental-variable estimator of Mullahy (1997) instrumenting $x_{it}$ by $z_{it}$. BC: bias-corrected version of IV. DIFF: differencing estimator from Section 3. True parameter value: $\theta = 1$. Coverage rates computed through the bootstrap with 99 replications. Results based on 5000 simulations.

Table 2: Simulations for additive-error model

| | | MEAN | | | STD | | | COV | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $T$ | EXO | IV | BC | EXO | IV | BC | EXO | IV | BC |
| 200 | 20 | 0.1403 | 1.0803 | 1.0049 | 0.0705 | 0.2187 | 0.2038 | 0 | 0.9040 | 0.9250 |
| 200 | 40 | 0.1399 | 1.0456 | 1.0090 | 0.0467 | 0.1367 | 0.1322 | 0 | 0.9150 | 0.9320 |
| 200 | 60 | 0.1421 | 1.0267 | 1.0027 | 0.0379 | 0.1201 | 0.1176 | 0 | 0.9300 | 0.9360 |
| 200 | 80 | 0.1427 | 1.0215 | 1.0037 | 0.0330 | 0.1007 | 0.0991 | 0 | 0.9320 | 0.9390 |
| 200 | 100 | 0.1402 | 1.0018 | 0.9987 | 0.0295 | 0.0890 | 0.0895 | 0 | 0.9320 | 0.9400 |

EXO: instrumental-variable estimator of Windmeijer and Santos Silva (1997) instrumenting $x_{it}$ by itself. IV: instrumental-variable estimator of Windmeijer and Santos Silva (1997) instrumenting $x_{it}$ by $z_{it}$. BC: bias-corrected version of IV. True parameter value: $\theta = 1$. Coverage rates computed through the bootstrap with 99 replications. Results based on 5000 simulations.

# References


Bao, Y. and A. Ullah (2007). The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics 140*, 650–669.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Econometrica 34*, 305–334.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*, 991–1030.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel data models with large $N$, $T$. *Journal of Econometrics 192*, 291–312.

Gouriéroux, C., A. Monfort, and A. Trognon (1984a). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica 52*, 701–720.

Gouriéroux, C., A. Monfort, and A. Trognon (1984b). Pseudo maximum likelihood methods: Theory. *Econometrica 52*, 681–700.

Haucap, J., A. Rasch, and J. Stiebale (2019). How mergers affect innovation: Theory and evidence. *International Journal of Industrial Organization 63*, 283–325.

Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics 99*, 478–485.

Mullahy, J. (1997). Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics 79*, 586–593.

Tenreyro, S. (2007). On the trade impact of nominal exchange rate volatility. *Journal of Development Economics 82*, 485–508.

Windmeijer, F. (2000). Moment conditions for fixed effects count data models with endogenous regressors. *Economics Letters 68*, 21–24.

Windmeijer, F. A. G. and J. M. C. Santos Silva (1997). Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics 12*, 281–294.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory 13*, 667–678.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics 90*, 77–97.