The Econometrics Journal

Econometrics Journal (2017), volume **20**, pp. S1–S13.
doi: 10.1111/ectj.12093

Multiple fixed effects in binary response panel data models

KARYNE B. CHARBONNEAU[†]

[†]*Bank of Canada, 234 Wellington St, Ottawa, Ontario, K1A 0G9, Canada.*
E-mail: `kcharbonneau@bankofcanada.ca`

First version received: March 2017; final version accepted: May 2017

Summary This paper considers the adaptability of estimation methods for binary response panel data models to multiple fixed effects. It is motivated by the gravity equation used in international trade, where important papers use binary response models with fixed effects for both importing and exporting countries. Econometric theory has mostly focused on the estimation of single fixed effects models. This paper investigates whether existing methods can be modified to eliminate multiple fixed effects for two specific models in which the incidental parameter problem has already been solved in the presence of a single fixed effect. We find that it is possible to generalize the conditional maximum likelihood approach to include two fixed effects for the logit. Monte Carlo simulations show that the conditional logit estimator presented in this paper is less biased than other logit estimators without sacrificing on precision. This superiority is emphasized in small samples. An application to trade data using the logit estimator further highlights the importance of properly accounting for two fixed effects.

Keywords: *Fixed effects, Logit, Maximum score, Panel data.*

1. INTRODUCTION

Fixed effects have long been recognized as a key element of econometric modelling of panel data, and a significant literature now exists in econometric theory on the inclusion of fixed effects in both linear and nonlinear panel data models. Although econometric theory has largely focused on single fixed effects estimators, empirical studies often include multiple fixed effects. The present paper attempts to bridge part of this gap by looking at the very popular logit model. The empirical relevance is demonstrated using Monte Carlo simulations and an application to international trade data.

This paper was motivated by the fixed effects gravity equation models used in international trade. This area of economics is concerned with the estimation of the factors conducive to trade between countries. The importance of using fixed effects to control for country-specific characteristics has been emphasized in an influential paper by Anderson and Van Wincoop (2003). They called these characteristics multilateral resistance factors, and they were meant to capture the fact that some countries simply trade more, or less. Many subsequent papers contributing to the gravity equation literature have included fixed effects in their estimation strategies. For example, Helpman et al. (2008) – hereafter referred to as HMR – and Santos Silva and Tenreyro (2006) estimate nonlinear panel data models with fixed effects for both importing and exporting countries. The first paper is a prominent study in the particular strand of the gravity equation literature that uses binary response panel data models to estimate the probability of positive trade.

This paper investigates whether existing methods for eliminating a single fixed effect can be modified to eliminate multiple fixed effects. This is relevant not only for data consisting of country pairs, but also for a number of other areas of empirical microeconomics. For example, in labour economics, influential papers, such as Abowd et al. (1999) use matched firm–employee data to study wage determinants of French workers. For such data sets, one might want to allow for both firm and worker fixed effects.[1] In a more recent paper, Kirabo Jackson (2013) studies the effect of match quality between employee and employer on productivity by using matched student–teacher–school data. Part of his methodology relies on a logit with both teacher and school fixed effects. These types of models are also used in the network literature. For example, Graham (2016) uses a fixed effect approach to develop similar estimators to the one presented here, in order to leave the joint distribution of the unobserved degree heterogeneity and observed agent attributes unrestricted.

Fixed effects do not generally cause any problem in static linear models, as they can easily be differenced out to allow consistent estimation of the relevant parameters. However, when considering nonlinear panel data models, we encounter the well-known incidental parameter problem identified by Neyman and Scott (1948).[2] This has motivated a rich literature on the estimation of single fixed effects nonlinear panel data models. The first model considered in the literature is the logit model studied in Rasch (1960, 1961). Manski (1987) generalized this to develop a conditional maximum score estimator for binary response models that remains consistent under weak assumptions on the distribution of the errors. These solutions to the incidental parameter problem, like those introduced in this paper, are model-specific.

With a more general approach to the problem, Hahn and Newey (2004) show that when n and T grow at the same rate, the fixed effects estimator is asymptotically biased and the asymptotic confidence intervals are wrong. They suggest two bias correction methods – the panel jackknife and the analytical bias correction – for the case of a single fixed effect. Also working on a general method, Arellano and Bonhomme (2009) suggest bias-reducing weighting schemes that can produce asymptotically valid confidence intervals when N and T grow at the same rate. In addition, Bonhomme (2012) proposes a systematic approach encompassing all nonlinear panel data models. He constructs moment restrictions on the parameters of interest that are free of the individual effects (once again, only one effect). This method applies to models with continuous dependent variables and is consistent for fixed T. The continuity requirement would exclude logit models in general and the example studied in Section 3 of the present paper in particular, for obvious reasons. Moreover, the second restriction would be problematic for the gravity literature as it generally has the T and N dimensions grow simultaneously.

Although there is still limited work in econometric theory for nonlinear panel data models involving multiple fixed effects, this paper is part of a growing literature on this topic. Fernandez-Val and Weidner (2016) adapt the analytical and jackknife bias correction methods introduced in Hahn and Newey (2004) to nonlinear models with additive or interactive individual and time effects. Their approach allows them to cover a broad class of popular models but does

[1] Other authors contributing to that literature, such as Postel-Vinay and Robin (2002), have raised questions concerning the validity of the fixed effects estimation with these types of data and have found alternative ways to allow for worker and employer heterogeneity.

[2] The incidental parameter problem refers to the fact that in nonlinear models with a fixed number of observations for each individual, the bias in the estimation of the fixed effects contaminates the estimates of the parameters of interest.

not completely eliminate the asymptotic bias and is limited to large-T panels. In a recent paper, Jochmans (2016) uses some of the results presented in an earlier version of this paper – Charbonneau (2013) – and develops the asymptotic properties of GMM estimators for models with two-way multiplicative fixed effects.

The paper proceeds as follows. First, we describe the estimation approach developed in the literature for one fixed effect in a logit model and then we try to generalize it to two. We find that we can adapt the conditional maximum likelihood to estimate a logit with two fixed effects. In general, this conditioning method is analogous to the difference-in-differences estimator used in linear models. We then proceed to show the relevance of appropriately dealing with two fixed effects in binary response models using the logit estimator presented in this paper. To accomplish that, we carry out Monte Carlo simulations (presented in the online Appendix), and we use data on trade flows between countries to test the logit estimator on the gravity equation.

Given the large number of empirical applications using multiple fixed effects and the general popularity of binary response models, this method has broad applicability. Furthermore, we find that appropriately controlling for multiple fixed effects has a substantial effect on the estimated parameters of interest relative to models without fixed effects or models inappropriately controlling for fixed effects.

2. MULTIPLE FIXED EFFECTS IN A LOGIT MODEL

The binary response model we consider is the simple and well-documented logit model. There is a well-known application of the conditional maximum likelihood 'trick' that allows us to solve the incidental parameter problem in a logit in the presence of one fixed effect. As we will see, it is possible to generalize this method to include two fixed effects. We begin by presenting the original solution, following somewhat closely the exposition of Arellano and Honoré (2001), before moving on to two fixed effects.

For $T = 2$, suppose that we have observations generated by

$$y_{it} = 1\{x_{it}'\beta + \alpha_i + \varepsilon_{it} \geq 0\} \quad i = 1, \ldots, n,$$

where for all i and t the ε_{it} are independent and have a logistic distribution conditional on x and the individual fixed effect α. This implies that we can express the following probability:

$$Pr(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i) = \frac{\exp(x_{i1}'\beta + \alpha_i)}{1 + \exp(x_{i1}'\beta + \alpha_i)}. \tag{2.1}$$

It is then easy to show that the conditional likelihood will eliminate the fixed effect such that

$$\begin{aligned} &Pr(y_{i1} = 1 \mid y_{i1} + y_{i2} = 1, x_{i1}, x_{i2}, \alpha_i) \\ &\quad = \frac{Pr(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i) Pr(y_{i2} = 0 \mid x_{i1}, x_{i2}, \alpha_i)}{Pr(y_{i1} = 1, y_{i2} = 0 \mid x_{i1}, x_{i2}, \alpha_i) + Pr(y_{i1} = 0, y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i)} \\ &\quad = \frac{\exp(x_{i1} - x_{i2})'\beta}{1 + \exp(x_{i1} - x_{i2})'\beta}. \end{aligned} \tag{2.2}$$

We can then find an estimator for the parameter β by applying this function to all pairs of observations for a given individual, and for all individuals. This can be generalized to the case where $T > 2$, and this is easy enough to calculate. Note that we are conditioning on

$y_{i1} + y_{i2} = 1$, which means that we are using the information contained in pairs of observations where the binary indicator changed. This approach to eliminate the fixed effects is also the one used in the maximum score estimator of Manski (1987), which we analyse in the Appendix. It is possible to obtain a likelihood function when $T > 2$, by conditioning on $\sum_{t=1}^{T} y_{it}$ to obtain the conditional distribution

$$P\Big(y_{i1}, \dots, y_{it} \mid \sum_{t=1}^{T} y_{it}, x_{i1}, \dots, x_{it}, \alpha_i\Big) = \frac{\exp(\sum_{t=1}^{T} y_{it} x'_{it}\beta)}{\sum_{(d_1,\dots,d_t)\in B} \exp(\sum_{t=1}^{T} d_t x'_{it}\beta)}, \tag{2.3}$$

where B is the set of all sequences of zeros and ones that have $\sum_{t=1}^{T} d_{it} = \sum_{t=1}^{T} y_{it}$; see Arellano and Honoré (2001). Note that this implies that $\sum_{t=1}^{T} y_{it}$ is a sufficient statistic for α_i.

We now show that a similar approach can be used in the case of two fixed effects in a logit model and we provide an analogous result. Suppose that the observations are now given by

$$y_{ij} = 1\{x'_{ij}\beta + \mu_i + \alpha_j + \varepsilon_{ij} \geq 0\} \quad i = 1, \dots, n,\ j = 1, \dots, n, \tag{2.4}$$

where μ_i and α_j are the fixed effects and ε_{ij} follows a logistic distribution.[3] Then, by applying the method used above to eliminate one fixed effect, we can write the following probabilities:[4]

$$Pr(y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, y_{lj} + y_{lk} = 1) = \frac{\exp((x_{lj} - x_{lk})'\beta + \alpha_j - \alpha_k)}{1 + \exp((x_{lj} - x_{lk})'\beta + \alpha_j - \alpha_k)} \tag{2.5}$$

and

$$Pr(y_{ij} = 1 \mid \mathbf{x}, \mu, \alpha, y_{ij} + y_{ik} = 1) = \frac{\exp((x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k)}{1 + \exp((x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k)}. \tag{2.6}$$

As can be seen, the two previous equations no longer depend on the μ fixed effects. However, they are still expressed in terms of α. We now try to find a conditional probability that does not depend on the latter. First, we notice that (2.5) and (2.6) are like a logit with $(x_{ij} - x_{ik})$ as an explanatory variable and $(\alpha_j - \alpha_k)$ as a fixed effect. We can therefore apply the trick a second time; hence, we compare it to another pair of observations with the same 'fixed effect'. Using both (2.5) and (2.6), and defining

$$\mathbf{c} \equiv \{y_{lj} + y_{lk} = 1,\ y_{ij} + y_{ik} = 1\},$$

we can now write the following conditional probability:

$$\begin{aligned} &Pr(y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, y_{lj} + y_{lk} = 1, y_{ij} + y_{ik} = 1, y_{ij} + y_{lj} = 1) \\ &\quad = \frac{Pr(y_{lj} = 1, y_{ij} + y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})}{Pr(y_{ij} + y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})} \\ &\quad = \frac{Pr(y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c}) Pr(y_{ij} = 0 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})}{Pr(y_{lj} = 1, y_{ij} = 0 \mid \mathbf{x}, \mu, \alpha, \mathbf{c}) + Pr(y_{lj} = 0, y_{ij} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})} \end{aligned}$$

[3] To remain consistent with the gravity model that motivated this paper and that is used in the application in Section 3, we illustrate our approach for the case where both dimensions of the panel are equal, or $T = n$. Note, however, that the method does not rely on this equality nor does it require a large-T panel.

[4] Throughout the paper, $\mathbf{x}$ refers to the vector of all x.

$$= \frac{\exp((x_{lj} - x_{lk})'\beta + \alpha_j - \alpha_k)}{\exp((x_{lj} - x_{lk})'\beta + \alpha_j - \alpha_k) + \exp((x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k)}$$

$$= \frac{\exp(((x_{lj} - x_{lk}) - (x_{ij} - x_{ik}))'\beta)}{1 + \exp(((x_{lj} - x_{lk}) - (x_{ij} - x_{ik})'\beta)}. \tag{2.7}$$

The probability no longer depends on the fixed effects, hence allowing us to solve the incidental parameter problem in the presence of two fixed effects. Indeed, we could now write a conditional maximum likelihood function or apply the last expression to all quadruples of observations, as with one fixed effect. Because the latter is easier to implement, the function to maximize is given by

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l,k \in Z_{ij}} \log\left(\frac{\exp(((x_{lj} - x_{lk}) - (x_{ij} - x_{ik}))'\beta)}{1 + \exp(((x_{lj} - x_{lk}) - (x_{ij} - x_{ik}))'\beta)}\right), \tag{2.8}$$

where Z_{ij} is the set of all the potential k and l that satisfy $y_{lj} + y_{lk} = 1$, $y_{ij} + y_{ik} = 1$, $y_{ij} + y_{lj} = 1$ for the pair ij.[5]

In the context of epidemiological studies, Hirji et al. (1987) show that a similar recursive conditioning can be used to eliminate what they call nuisance parameters and speed-up computations. The nuisance parameters that they consider are not fixed effects and do not relate to the incidental parameters problem; they are simply normal covariates (like the x variables in our model) that one needs to control for but for which the effect on the dependent variable is not of interest (e.g. the constant).

Note that the possibility of solving the incidental parameter problem for one fixed effect does not guarantee that it can be done for two or more. For instance, despite its similarities with the logit, Manski's maximum score estimator cannot be adapted to the case of two fixed effects (see the Appendix). Fundamentally, the maximum score estimator fails in the presence of multiple fixed effects because it does not have a recursive structure. The logit can accommodate two fixed effects because using the known method once to deal with the first fixed effect gives us another logit, therefore allowing a second application of that method. This does not hold for Manski's maximum score estimator.

To assess the accuracy of this two fixed effects logit estimator and to compare it with other logit estimators, we ran Monte Carlo simulations. The results are presented in the online Appendix. In short, these Monte Carlo simulations confirm that the logit estimator presented in this paper is less biased than, as precise as and more robust to different fixed effects than other logit estimators. Moreover, standard errors based on this estimator have the correct size. We now move on to apply this estimator to trade data.

[5] A conditional maximum likelihood function would be analogous to (2.3). The sufficient statistics for μ_i and α_j would then be $\sum_{j=1}^{n} y_{ij}$ and $\sum_{i=1}^{n} y_{ij}$. As emphasized in Arellano and Honoré (2001), it would be computationally burdensome to calculate the conditional maximum likelihood because of the large number of terms in the denominator, which is even larger in the case of two fixed effects. Therefore, similarly to the one fixed effect case, we apply (2.8).

3. APPLICATION: THE GRAVITY EQUATION AND THE EXTENSIVE MARGINS OF TRADE

Understanding how different trade barriers influence trade flows is key when one wants to study the impact of distance, trade agreements and other trade frictions. To do that, economists have been using the gravity equation for over 50 years. As Bernard et al. (2007) state, 'the gravity equation for bilateral trade flows is one of the most successful empirical relationships in international economics'. The gravity equation was first applied to aggregate trade. As its name suggests, it was initially motivated by the Newtonian theory of gravitation (bilateral trade should be positively related to the size of countries, as measured by their GDP, and negatively related to their distance from one another). It now has a plethora of microeconomic foundations. More recent work has emphasized the role of extensive margin adjustments in understanding the variations of aggregate trade flows and has derived gravity equations for these extensive margin adjustments; see, e.g. Bernard et al. (2011) and Mayer et al. (2014).[6] For the purpose of this application, we refer to the model where the dependent variable is binary (trade or not) as the binary gravity equation.

There are many 'zeros and ones' relationships in trade and the logit is very widely used. In the context of the gravity equation literature, the logit is most commonly used to study the extensive margins of trade in heterogeneous firm models. In their influential paper, HMR try to improve on traditional estimates of the gravity equation by accounting for both firm heterogeneity – in a Melitz (2003) framework – and the frequently forgotten zero trade flows. To do this, they use a two-stage procedure, where the first stage consists of estimating the probability that a country trades with another. Although they use a probit with importer and exporter fixed effects, one could also similarly use a logit. In a paper estimating the Chaney (2008) model with French firm-level data, Crozet and Koenig (2010) also use the probability of exporting as a first stage in their empirical strategy.[7] More specifically, they run a logit with firm and import country–year fixed effects to disentangle the elasticity of trade barriers on the intensive and extensive margins.

In order to allow direct comparison with results found in the literature, the application of the logit estimator on the binary gravity equation is done using the data from HMR. This data set consists of information on trade flows and country characteristics for 158 countries in 1986. Applying their specific gravity model to their data gives added weight to the comparison of the estimates produced by the estimator presented in this paper with those produced by other commonly used estimators. However, note that all the results discussed below hold when using different trade data.[8] More detailed information about the data used by HMR can be found in their paper. We find that applying the conditional logit to properly account for the multiple fixed effects gives significantly different estimates of the probability of trading conditional on a set of explanatory variables. For convenience, we hereafter refer to the logit estimator given by the maximization of (2.8) as Logit 2FE. In what follows, we also call the regular logit that ignores fixed effects, Logit, and the logit that estimates all the fixed effects (putting in dummies), Logit FE.

[6] Trade frictions have an impact on aggregate trade flows through both the amount that each firm or country exports (the intensive margins) and the number of firms or countries exporting (the extensive margin). Note that the extensive margin can also refer to the number of products exported.

[7] Chaney (2008) introduces firm heterogeneity in a model of international trade to look at the effect the elasticity of substitution between goods has on the intensive and extensive margins of trade. It is essentially a Melitz model with a Pareto productivity distribution.

[8] Results available on demand.

The HMR gravity model to be estimated is given by

$$\begin{aligned}
&\text{Prob}[\mathit{Trade}_{ij} = 1|\text{observed variables}] \\
&\quad = F(\beta_0 + \beta_1 \ln(D_{ij}) + \beta_2 \mathit{Border}_{ij} + \beta_3 \mathit{Island}_{ij} \\
&\qquad + \beta_4 \mathit{Landlock}_{ij} + \beta_5 \mathit{Legal}_{ij} + \beta_6 \mathit{Language}_{ij} + \beta_7 \mathit{Colony}_{ij} \\
&\qquad + \beta_8 \mathit{Currency}_{ij} + \beta_9 \mathit{RTA}_{ij} + \beta_{10} \mathit{Religion}_{ij} + \mu_i + \alpha_j),
\end{aligned} \tag{3.1}$$

where Trade_{ij} is an indicator variable that equals 1 when country i exports to country j and 0 otherwise, and $F(\cdot)$ is the cumulative distribution function (cdf) of the assumed distribution.[9] Other variables are defined as follows: D_{ij} is the simple distance between the most populated cities of country i and country j, Border_{ij} is a dummy that takes the value 1 if i and j share a border, Island_{ij} is a dummy that takes the value 1 if either one or both countries are islands, $\mathit{Landlock}_{ij}$ is a dummy that takes the value 1 if either one or both countries do not have access to an ocean, Legal_{ij} is a dummy that takes the value 1 if the two countries have the same legal system, $\mathit{Language}_{ij}$ is a dummy that takes the value 1 if i and j have the same official language, Colony_{ij} is a dummy that takes the value 1 if i and j were ever in a colonial relationship, $\mathit{Currency}_{ij}$ is a dummy that takes the value 1 if the two countries use the same currency and RTA_{ij} is a dummy that takes the value 1 if i and j are in a regional trade agreement. Finally, the religion variable is defined as

$$\begin{aligned}
\mathit{Religion}_{ij} = {} & (\%\text{Protestants in country } i \cdot \%\text{Protestants in country } j) \\
& + (\%\text{Catholics in country } i \cdot \%\text{Catholics in country } j) \\
& + (\%\text{Muslims in country } i \cdot \%\text{Muslims in country } j).
\end{aligned}$$

Finally, μ_i and α_j are importer and exporter fixed effects, respectively. The results are presented in Table 1.

In addition to the OLS with fixed effects, the Logit, the Logit FE and the Logit 2FE, we present results for the Probit with fixed effects used in HMR. Note that the latter are a successful replication of the estimates presented in the original paper. Overall, the estimated coefficients on distance differ somewhat more between the Logit FE and Logit 2FE than would have been expected based on the Monte Carlo simulations. Indeed, the Logit 2FE suggests that distance might have a smaller impact on the probability of exporting than traditional binary models estimates indicate, and the difference with the Logit FE is statistically significant. The estimated coefficient on border is also different between the two models, but that difference is not statistically significant. In general, although the estimated coefficients for the other variables are different for the Logit 2FE and the Logit FE, this difference, unlike that for distance or border, is much closer to what the Monte Carlo simulations, especially design 4, suggested. Indeed, Table A.4 of the online Appendix shows that the difference between the two models is at least twice as large for the coefficients on the distance and border variables. In the case of the border dummy, this could be due to its sparsity (only 17% are non-zero) and its high correlation with distance.

Since the Logit 2FE does not provide estimates of the fixed effects, it is not possible to compute the marginal effect of each variable. However, we can still compute the effect implied by the difference in the distance coefficients of the various fixed effects estimators on the

[9] In the original paper, the cdf is a unit-normal distribution, but here we assume it is a logistic distribution.

Table 1. Logit results: HMR model and data.

Variables	Probit FE	OLS FE	Logit	Logit FE	Logit 2FE
Distance	−0.6597	−0.1086	−0.2939	−1.2526	−0.9948
	(0.0239)	(0.0041)	(0.0192)	(0.0424)	(0.0547)
Border	−0.3825	−0.0810	0.2834	−0.7624	−0.5171
	(0.0993)	(0.0218)	(0.1184)	(0.1761)	(0.2285)
Island	−0.3447	−0.0648	−0.3355	−0.6030	−0.3924
	(0.0743)	(0.0135)	(0.0287)	(0.1359)	(0.1315)
Landlock	−0.1806	−0.0316	−0.6225	−0.3607	−0.1984
	(0.0973)	(0.0156)	(0.0308)	(0.1808)	(0.1860)
Legal	0.0964	−0.0039	−0.4939	0.1867	0.1832
	(0.0298)	(0.0053)	(0.0289)	(0.0539)	(0.0608)
Language	0.2838	0.0783	0.1208	0.5037	0.4156
	(0.0378)	(0.0068)	(0.0313)	(0.0696)	(0.0726)
Colonial ties	0.3252	0.0408	4.0701	0.5392	1.1366
	(0.2444)	(0.0196)	(0.4165)	(0.4647)	(0.9944)
Currency	0.4917	0.0772	−0.7229	0.8973	1.0565
	(0.1227)	(0.0243)	(0.1497)	(0.2249)	(0.2499)
RTA	1.9851	−0.0526	2.9708	3.4708	3.5227
	(0.2651)	(0.0413)	(0.4193)	(0.4928)	(0.1696)
Religion	0.2605	0.0739	0.3289	0.4155	0.4850
	(0.0583)	(0.0108)	(0.0478)	(0.1056)	(0.1145)

Note: Standard errors clustered at *ij* level (allowing for importer and exporter correlation). These are the estimated coefficient values, and therefore not directly comparable across models (except for the three logit models).

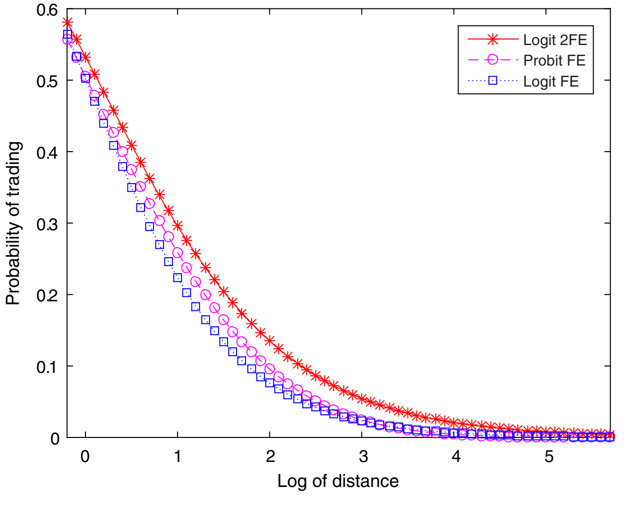

Figure 1. Probability of trading relative to distance. [Colour figure can be viewed at wileyonlinelibrary.com]

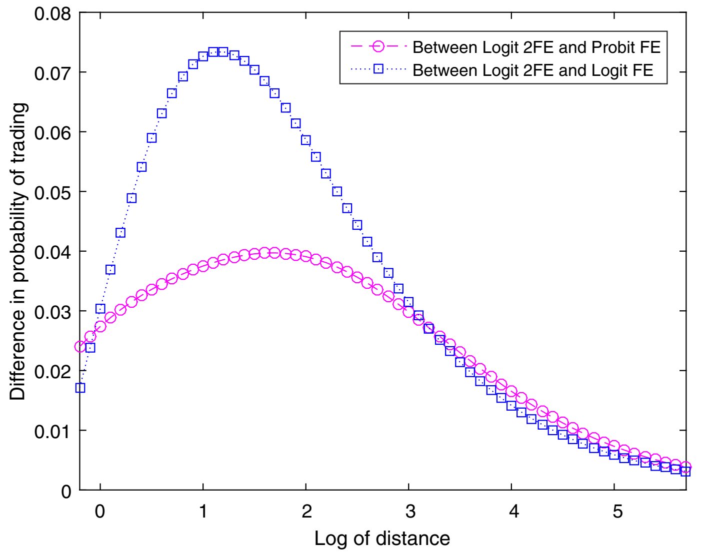

Figure 2. Difference in probability of trading relative to distance between different estimators. [Colour figure can be viewed at wileyonlinelibrary.com]

probability of trading by looking at how this probability, estimated at the sample averages (and therefore setting all fixed effects to zero), varies with the measure of distance. This is illustrated in Figure 1 for the Probit FE, Logit FE and Logit 2FE. As expected, the Logit 2FE predicts a higher probability of trading for all values of distance, particularly in the middle ranges. Note that Figure 1 represents all distance values in the sample. Figure 2 plots the difference between the probability of trading predicted by the Logit 2FE and the other two main models. The difference is larger with the Logit FE, especially for countries that are relatively close, and peaks at 7.3 percentage points. This application highlights the importance of properly accounting for multiple fixed effects.

4. CONCLUSION

This paper has examined estimators of binary response panel data models with multiple fixed effects. There are an abundance of empirical methods applying two fixed effects in binary response models in general, and in the logit in particular. However, current estimators are subject to the incidental parameters problem. Although many methods have been developed to address this problem in models with a single fixed effect, very little has been done for the cases with two or more fixed effects. Attempting to fill this important gap, we have developed a method to appropriately deal with two fixed effects for the logit model.

Our method is based on the conditional maximum likelihood of Rasch (1960, 1961). If, with one fixed effect, it suffices to condition on the sum of the observations in one dimension (typically, for one individual, the sum of y_{it} over time), then with two fixed effects we condition on the sums in both dimensions (for one importer i, the sum of y_{ij} for all exporters j; for one exporter j, the sum of y_{ij} over all importers i). This approach allows us to consistently estimate the parameters of interest.

We have shown that the conditioning method, which is the core of this paper, performs well in recovering the true parameters in Monte Carlo studies. Indeed, we have found that the conditional logit presented in this paper is less biased than, as precise as and more robust to different fixed effects than other logit estimators. Importantly, if this superiority is emphasized in small samples, it does not disappear for large samples (e.g. the size that one can obtain when studying trade between countries). Moreover, standard errors produce correctly sized tests. We have also shown that this same procedure yields quite different estimated coefficients from methods subjected to the incidental parameters problem in applications with actual trade data.

The method developed in this paper has broad applicability, and our Monte Carlo studies and applications highlight the importance of appropriately controlling for multiple fixed effects in binary response panel data models for recovering the parameters of the underlying relationships of interest.

ACKNOWLEDGEMENTS

The author is grateful to Bo Honoré for his valuable help and support since the start of this project. The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Bank of Canada.

REFERENCES

Abowd, J. M., F. Kramarz and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*, 251–333.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society 32*, 283–301.

Anderson, J. E. and E. Van Wincoop (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review 93*(1), 170–92.

Arellano, M. and S. Bonhomme (2009). Robust priors in nonlinear panel data models. *Econometrica 77*, 489–536.

Arellano, M. and B. E. Honoré (2001). Panel data models: some recent developments. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, *Volume 5*, Chapter 53, 3229–96. Amsterdam: North-Holland.

Bernard, A., J. B. Jensen, S. Redding and P. Schott (2007). Firms in international trade. *Journal of Economic Perspectives 21*, 105–30.

Bernard, A. B., S. J. Redding and P. K. Schott (2011). Multiproduct firms and trade liberalization. *Quarterly Journal of Economics 126*, 1271–318.

Bonhomme, S. (2012). Functional differencing. *Econometrica 80*, 1337–85.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Charbonneau, K. B. (2013). Multiple fixed effects in theoretical and applied econometrics. PhD Thesis, Princeton University.

Crozet, M. and P. Koenig (2010). Structural gravity equations with intensive and extensive margins. *Canadian Journal of Economics 43*, 41–62.

Fernandez-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.

Graham, B. S. (2016). An econometric model of network formation with degree heterogeneity. Forthcoming in *Econometrica*.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*, 1295–319.

Helpman, E., M. Melitz and Y. Rubinstein (2008). Estimating trade flows: trading partners and trading volumes. *Quarterly Journal of Economics 123*, 441–87.

Hirji, K. F., C. R. Mehta and N. R. Patel (1987). Computing distributions for exact logistic regression. *Journal of the American Statistical Association 82*, 1110–17.

Jochmans, K. (2016). Two-way models for gravity. Forthcoming in *Review of Economics and Statistics*.

Kirabo Jackson, C. (2013). Match quality, worker productivity, and worker mobility: direct evidence from teachers. *Review of Economics and Statistics 95*, 1096–116.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–28.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica 55*, 357–62.

Mayer, T., M. Melitz and G. I. P. Ottaviano (2014). Market size, competition, and the product mix of exporters. *American Economic Review 104*(2), 495–536.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocation and aggregate industry productivity. *Econometrica 71*, 1695–725.

Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*, 2295–350.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institut.

Rasch, G. (1961). On the general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, *Volume IV: Contributions to biology and problems of medicine*, 321–33. Berkeley, CA: University of California Press.

Santos Silva, J. M. C. and S. Tenreyro (2006). The log of gravity. *Review of Economics and Statistics 88*, 641–58.

APPENDIX: MANSKI'S MAXIMUM SCORE ESTIMATOR

Manski (1987) developed a consistent maximum score estimator for binary response models allowing for individual fixed or random effects in panel data. This estimator, unlike its predecessors – e.g. Andersen (1970) – remains consistent under very weak assumptions on the disturbances. This characteristic could make a multiple fixed effects maximum score estimator very useful. Therefore, we want to investigate the possibility of generalizing this estimator to the case where there are two fixed effects. The conditional maximum score estimator is similar to the estimator of the logit model. Indeed, it is also applied to a binary response model and uses pairs of observations for the same individual where the value of the indicator variable differs. However, unlike the logit conditional maximum likelihood, this estimator does not generalize to the case with two fixed effects, even under a stronger set of assumptions. As detailed later, this is because of the lack of recursive structure in this particular model.

In Manski's original paper, the model has the form

$$P(y_{it} = 1 \mid x_{i1}, x_{i2}, \alpha_i) = F_i(x_{it}'\beta + \alpha_i) \quad t = 1, 2,$$

where α_i once again represents the individual effect. The first assumption of Manski (1987) is that the distribution F depends on i. It requires the disturbance to be stationary conditional on the identity of the panel member but does not restrict it to be the same across individuals.

Manski's key result resides in his first lemma, as follows.

LEMMA A.1.

$$\begin{aligned}
x_{i2}'\beta > x_{i1}'\beta &\iff P(y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i) > P(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i)\\
x_{i2}'\beta = x_{i1}'\beta &\iff P(y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i) = P(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i)\\
x_{i2}'\beta < x_{i1}'\beta &\iff P(y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i) < P(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i).
\end{aligned} \tag{A.1}$$

If we condition on $y_{i1} + y_{i2} = 1$, we obtain

$$P(y_{i2} = 1 \mid y_{i1} + y_{i2} = 1, x_{i1}, x_{i2}, \alpha_i) \begin{cases} >1/2 & \text{if } (x_{i2} - x_{i1})'\beta > 0\\ = 1/2 & \text{if } (x_{i2} - x_{i1})'\beta = 0\\ < 1/2 & \text{if } (x_{i2} - x_{i1})'\beta < 0. \end{cases} \tag{A.2}$$

The probability in (A.2) takes the same form as in Manski (1975), so it is possible to use the maximum score estimator. This first lemma allows him to develop, under some identification conditions, a consistent estimator by maximizing for b the sample analogue of the following equation

$$H(b) \equiv E[\mathrm{sgn}((x_{i2} - x_{i1})'b)(y_{i2} - y_{i1})], \tag{A.3}$$

for the observations where $y_{i1} \neq y_{i2}$.

Unfortunately, this approach cannot be generalized in such a way as to generate an equivalent to this necessary lemma for the case of multiple fixed effects panel data models. Indeed, following a similar line of thought as for the logit case presented earlier, we would hope to adapt Lemma A.1 by applying the same type of conditioning twice.

Introducing a second fixed effect in the model, we now have

$$P(y_{ij} = 1 \mid \mathbf{x}, \mu, \alpha) = F(x_{ij}'\beta + \mu_i + \alpha_j) \quad i, j = 1, \ldots, n.$$

Here we restrict F to be the same for all observations. In other words, all the disturbances are drawn from the same distribution. This is more restrictive than Manski's assumption, but still allows for an interesting range of models. We show that even under this stricter set of assumptions, we cannot generalize this estimator to the case of two fixed effects. To do so, we first apply an analogous conditioning to that of (A.2) to eliminate μ_i and we obtain

$$P(y_{ij} = 1 \mid y_{ij} + y_{ik} = 1, \mathbf{x}, \mu, \alpha) \begin{cases} >1/2 & \text{if } (x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k > 0\\ = 1/2 & \text{if } (x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k = 0\\ < 1/2 & \text{if } (x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k < 0. \end{cases}$$

This is similar to the first-step equations of the logit model (i.e. (2.5) and (2.6)): explanatory variable $(x_{ij} - x_{ik})$ and fixed effect $\alpha_j - \alpha_k$. However, to apply this conditioning again, we would need $P(y_{ij} = 1 \mid y_{ij} + y_{ik} = 1, \mathbf{x}, \mu, \alpha)$ to have the form $F((x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k)$, where F is a cdf. Yet, this does not hold: we cannot attest that this probability is always increasing. Therefore, we cannot apply Manski's conditioning a second time: Manski's maximum score estimator cannot be adapted to the presence of two fixed effects, even under a stronger set of assumptions.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Online Appendix
Replication Files