Multiple Fixed Effects in Binary Response Panel Data Models

by Karyne B. Charbonneau

www.bank-banque-canada.ca

Multiple Fixed Effects in Binary Response Panel Data Models

by

Karyne B. Charbonneau

International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
kcharbonneau@bankofcanada.ca

Acknowledgements

Abstract

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 such as Helpman, Melitz and Rubinstein (2008) 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 of Rasch (1960, 1961) to include two fixed effects for the logit. Surprisingly, despite many similarities with the logit, Manski's (1987) maximum score estimator for binary response models cannot be adapted to the presence of two fixed effects. 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.

JEL classification: C23, C25, F14
Bank classification: Econometric and statistical methods

Résumé

L'auteure évalue l'adaptabilité des méthodes d'estimation utilisées pour les modèles de données de panel à réponse binaire à l'introduction d'effets fixes multiples. Ce travail s'inspire de l'usage des modèles de gravité dans l'analyse du commerce international, domaine dans lequel des travaux importants, comme ceux de Helpman, Melitz et Rubinstein (2008), s'appuient sur des modèles binaires à effets fixes pour les pays importateurs et exportateurs. En économétrie, les théoriciens se sont surtout attachés à estimer des modèles en présence d'un seul effet fixe. L'auteure se demande s'il est possible de modifier les méthodes d'estimation existantes afin d'éliminer les effets fixes multiples sur deux modèles spécifiques où une solution a été apportée au problème des paramètres incidents dans un cadre avec un seul effet fixe. Il s'avère qu'il est possible de généraliser l'approche du maximum de vraisemblance conditionnelle proposée par Rasch (1960 et 1961) de manière à inclure deux effets fixes dans le modèle logit considéré. Fait étonnant, malgré ses nombreuses similitudes avec le modèle logit, l'estimateur du score maximum de Manski (1987), qui est appliqué aux modèles binaires, ne peut être adapté si deux effets fixes sont introduits. Les simulations de Monte-Carlo montrent que l'estimateur conditionnel logit présenté par l'auteure permet d'obtenir des estimations moins biaisées que celles d'autres estimateurs logit sans pour autant sacrifier la précision. Cette supériorité est davantage accusée dans le cas des petits échantillons. Appliqué aux

données relatives aux échanges commerciaux, l'estimateur logit étudié montre encore plus qu'il convient de bien tenir compte des doubles effets fixes.

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 as well as Manski's (1987) more general maximum score estimator. 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, Melitz and Rubinstein (2008) (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 not only relevant for data consisting of country pairs, but also in a number of other areas of empirical microeconomics. For example, in an influential paper, Abowd, Kramarz and Margolis (1999) used 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 similar fashion, Aaronson, Barrow and Sander (2007) and Rivkin, Hanushek and Kain (2005) used matched data between students and teachers to study academic achievement. In both of those papers, multiple fixed effects were also used. In a more recent paper, Kirabo Jackson (2013) combines these two ideas to study 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.

Fixed effects do not generally cause any problem in "static" linear models, since 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 analytic 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

---

[1]Other authors contributing to that literature, for example, Postel-Vinay and Robin (2002), have raised questions concerning the validity of the fixed effects estimation with these types of data and 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.

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 first restriction would exclude logit models in general and the example studied in section 4 of the present paper in particular, for obvious reasons. Moreover, the second restriction would be problematic for the gravity literature since it generally does not have a fixed T.

Very little work has been done in econometric theory for nonlinear panel data models involving multiple fixed effects. One exception is Fernandez-Val and Weidner (2011), who 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 not completely eliminate the asymptotic bias and is limited to large-T panels.

The paper will proceed as follows: for the two models mentioned above, we describe the estimation approach developed in the literature for one fixed effect and then 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. However, despite many similarities with the logit, we find that the maximum score estimator cannot be adapted for multiple fixed effects. 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 first carry out Monte Carlo simulations, after which 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 Binary Response Panel Data Models

2.1 Fixed effects in a logit model

The first binary response model we shall 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:

$$y_{it} = 1\{x'_{it}\beta + \alpha_i + \varepsilon_{it} \geq 0\} \quad i = 1, ..., n,$$

where for all i and t the ε_{it} are independent and have a logistic distribution conditional on the x's and the individual fixed effect α. This implies that we can express the following probability:

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

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

$$Pr(y_{i1} = 1 \mid y_{i1} + y_{i2} = 1, x_{i1}, x_{i2}, \alpha_i)$$
$$= \frac{Pr(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i)Pr(y_{i2} = 0 \mid x_{i1}, x_{i2}, \alpha_i)}{Pr(y_{i1} = 1, y_{i2} = 0 \mid x_{i1}, x_{i2}, \alpha_i) + Pr(y_{i1} = 0, y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i)}$$
$$= \frac{\exp(x_{i1} - x_{i2})'\beta}{1 + \exp(x_{i1} - x_{i2})'\beta}. \tag{2}$$

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 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 Manski's (1987) maximum score estimator, which we will analyze in the next subsection. 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:

$$P\left(y_{i1}, \ldots, y_{it} \mid \sum_{t=1}^{T} y_{it}, x_{i1}, \ldots, x_{it}, \alpha_i\right) = \frac{\exp(\sum_{t=1}^{T} y_{it} x'_{it} \beta)}{\sum_{(d_1, \ldots, d_t) \in B} \exp(\sum_{t=1}^{T} d_t x'_{it} \beta)}, \tag{3}$$

with B being the set of all sequences of zeros and ones that have $\sum_{t=1}^{T} d_{it} = \sum_{t=1}^{T} y_{it}$ (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 provide an analogous result. Suppose that the observations are now given by

$$y_{ij} = 1\{x'_{ij} \beta + \mu_i + \alpha_j + \varepsilon_{ij} \geq 0\} \quad i = 1, \ldots, n, j = 1, \ldots, n, \tag{4}$$

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]

$$Pr(y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, y_{lj} + y_{lk} = 1) = \frac{\exp\left[(x_{lj} - x_{lk})' \beta + \alpha_j - \alpha_k\right]}{1 + \exp\left[(x_{lj} - x_{lk})' \beta + \alpha_j - \alpha_k\right]} \tag{5}$$

and

$$Pr(y_{ij} = 1 \mid \mathbf{x}, \mu, \alpha, y_{ij} + y_{ik} = 1) = \frac{\exp\left[(x_{ij} - x_{ik})' \beta + \alpha_j - \alpha_k\right]}{1 + \exp\left[(x_{ij} - x_{ik})' \beta + \alpha_j - \alpha_k\right]}. \tag{6}$$

As can be seen, the two previous equations no longer depend on the μ fixed effects. However, they are still expressed in terms of the α's. We now try to find a conditional probability that does not depend on the latter. First, we notice that equations (5) and (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

---

[3]To remain consistent with the gravity model that motivated this paper and is used in the section 4 application, we illustrate our approach for the case where $n = T$. Note, however, that the method does not rely on this equality nor does it require a large-T panel.

[4]Throughout the paper, $\mathbf{x}$ will refer to the vector of all x's.

apply the trick a second time; hence we compare it to another pair of observations with the same "fixed effect." Using both equations (5) and (6), and defining

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

we can now write the following conditional probability:

$$
\begin{aligned}
Pr(y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, y_{lj} + y_{lk} = 1, y_{ij} + y_{ik} = 1, y_{ij} + y_{lj} = 1) \\
= \frac{Pr(y_{lj} = 1, y_{ij} + y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})}{Pr(y_{ij} + y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})} \\
= \frac{Pr(y_{lj} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c}) Pr(y_{ij} = 0 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})}{Pr(y_{lj} = 1, y_{ij} = 0 \mid \mathbf{x}, \mu, \alpha, \mathbf{c}) + Pr(y_{lj} = 0, y_{ij} = 1 \mid \mathbf{x}, \mu, \alpha, \mathbf{c})} \\
= \frac{\exp\left[(x_{lj} - x_{lk})'\beta + \alpha_j - \alpha_k\right]}{\exp\left[(x_{lj} - x_{lk})'\beta + \alpha_j - \alpha_k\right] + \exp\left[(x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k\right]} \\
= \frac{\exp\left[((x_{lj} - x_{lk}) - (x_{ij} - x_{ik}))'\beta\right]}{1 + \exp\left[((x_{lj} - x_{lk}) - (x_{ij} - x_{ik})'\beta\right]}.
\end{aligned}
\tag{7}
$$

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. Since the latter is easier to implement, the function to maximize is given by

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l,k \in Z_{ij}} \log\left(\frac{\exp\left[((x_{lj} - x_{lk}) - (x_{ij} - x_{ik}))'\beta\right]}{1 + \exp\left[((x_{lj} - x_{lk}) - (x_{ij} - x_{ik}))'\beta\right]}\right),
\tag{8}
$$

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.

In the context of epidemiological studies, Hirji, Mehta, and Patel (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

6

(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 (for example, the constant).

To assess the accuracy of this two fixed effects logit estimator and compare it with other logit estimators, we ran Monte Carlo simulations. The results are presented in section 3.

We now move on to a related model where we achieve a different outcome, thereby showing that solving the incidental parameter problem for one fixed effect does not guarantee that it can be done for two or more.

2.2   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 will be detailed later, this is due to 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. Manski's (1987) first assumption 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:

Lemma M 1.

$$x'_{i2}\beta > x'_{i1}\beta \quad \Longleftrightarrow \quad P(y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i) > P(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i)$$

$$x'_{i2}\beta = x'_{i1}\beta \quad \Longleftrightarrow \quad P(y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i) = P(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i) \qquad (9)$$

$$x'_{i2}\beta < x'_{i1}\beta \quad \Longleftrightarrow \quad P(y_{i2} = 1 \mid x_{i1}, x_{i2}, \alpha_i) < P(y_{i1} = 1 \mid x_{i1}, x_{i2}, \alpha_i).$$

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

$$P(y_{i2} = 1 \mid y_{i1} + y_{i2} = 1, x_{i1}, x_{i2}, \alpha_i) \begin{cases} > 1/2 & \text{if } (x_{i2} - x_{i1})'\beta > 0 \\ = 1/2 & \text{if } (x_{i2} - x_{i1})'\beta = 0 \\ < 1/2 & \text{if } (x_{i2} - x_{i1})'\beta < 0. \end{cases} \qquad (10)$$

The probability in (10) 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 analog of the following equation:

$$H(b) \equiv E[\text{sgn}((x_{i2} - x_{i1})'b)(y_{i2} - y_{i1})] \qquad (11)$$

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 M 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 will 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 equation (10) to eliminate μ_i and 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., equations (5) and (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 $F\big((x_{ij} - x_{ik})'\beta + \alpha_j - \alpha_k\big)$ where F is a CDF. Yet, this does not hold: we can't attest that this probability is always increasing. Therefore, we cannot apply Manski 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.

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. We will now look at the performance of the conditional logit with two fixed effects using Monte Carlo simulations.

3   Monte Carlo Simulations

In this section, we present Monte Carlo evidence to support the multiple fixed effects estimator developed in this paper: the logit estimator given by the maximization of equation

(8). For convenience, we will refer to this estimator as Logit 2FE. The simulations will compare that estimator to a regular logit (simply called Logit in what follows), ignoring the fixed effects, and a logit estimating all the fixed effects (putting in dummies, hereafter called Logit FE). Recall that this last estimator is subject to the incidental parameter problem.

To account for different possible features of the data, this comparison will be made for four different designs. All of these designs are applied to the estimation of the following model:

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

where x_{ij} is a vector of five explanatory variables[5] drawn from a standard normal distribution and the error term ε_{ij} is drawn from a logistic distribution. The first design has no fixed effects ($\mu_i = \alpha_j = 0 \quad \forall i, j$). The second design has fixed effects drawn from a standard normal distribution uncorrelated with the explanatory variables. In both of these first two designs, $\beta_1 = 1$ and $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$. The third design has fixed effects correlated with the first explanatory variables. More specifically, $x_1 = rndn + \alpha + \eta$, where $rndn$ is a standard normal and α and η are the same fixed effects as in the second design. To illustrate how this affects the coefficient on other explanatory variables uncorrelated with the fixed effects, we now also have $\beta_2 = 1$. The first three designs do not have resampling of the x's or the fixed effects in each replication. However, since it is more common in Monte Carlo studies to have resampling, the fourth and last design replicates the second design, but with resampling of the data. Each design is estimated both for $n = 136$ and $n = 68$ (i.e., with 18,496 and 4,624 observations, respectively). These are both standard size ranges for trade studies.[6] Whenever fixed effects are estimated, the coefficients are truncated in order to ensure convergence.[7] The results from 1,000 replications are given in Tables 1 through 4. For each estimator considered, we report the median bias, the median absolute deviation (MAD), the mean bias and the root mean squared error (RMSE) for all five coefficient estimates.

---

[5]The simulation design is chosen to match as closely as possible the empirical application.

[6]The sample size in Santos Silva and Tenreyro (2006) is 136 countries.

[7]This is done in order to avoid the fixed effects taking on extremely large values (in absolute terms) to accommodate individuals with only zeros or only ones.

In addition, in order to get a more precise picture of how the relative performance of these estimators varies with sample size, we simulate the design with correlated fixed effects (i.e., design 3) for $n = [50, 75, 100, \ldots, 200]$. These results, for the coefficient β_1, are reported in Figures 1 and 2.

Using the median absolute deviation as the measure of the precision and looking at the median bias[8] we see that for all designs, the Logit 2FE presented in this paper has smaller bias without sacrificing precision. As expected, sample size matters, both for the bias and the precision. Indeed, the median absolute deviation is approximately twice as large in the small sample for all three estimators in all designs. Since the sample size when $n = 68$ is four times smaller than when $n = 136$, a doubling of the MAD is what we would expect. Moreover, while for the Logit FE, cutting the sample by half almost doubles the median bias on the positive coefficient in all designs, it has no significant impact on the median bias for the Logit 2FE. Except in the first design, which has no fixed effects, the regular logit is severely biased, especially when the fixed effects are correlated with the explanatory variable. This is not surprising since except for the case without fixed effects, that estimator is misspecified. Design 1 (see Table 1) shows that wrongly assuming that there are fixed effects when there are none can lead to biased estimates when using the Logit FE, but not when using the Logit 2FE presented in this paper. Comparing the second and third designs (uncorrelated and correlated fixed effects, respectively), we observe that when the fixed effects are correlated with one of the explanatory variables, not only does it increase the bias on the coefficient of that variable for the Logit FE, but it also causes bias on the other positive coefficients. Indeed, we see that β_1 has a median bias of 0.0266, compared to 0.0201 in design 2, and β_2 a median bias of 0.0241 (in design 3). This is amplified in the small sample. Still comparing designs with uncorrelated and correlated fixed effects, we further observe that the median bias remains similarly small for the Logit 2FE, showing that this estimator is equally capable of dealing with correlated and uncorrelated fixed effects. Comparing designs 2 and 4, we see that, as anticipated, resampling the x's in each replication has little to no effect on the results.

---

[8]We mostly discuss results for the positive coefficient(s).

Table 1: Design 1 - no fixed effects

		Median bias	MAD	Mean bias	RMSE
Full sample: $n = 136$					
Logit	β_1	0.0001	0.0107	0.0002	0.0155
	β_2	−0.0004	0.0068	−0.0003	0.0103
	β_3	−0.0002	0.0077	−0.0002	0.0108
	β_4	−0.0002	0.0074	−0.0001	0.0109
	β_5	0.0003	0.0072	0.0003	0.0109
Logit FE	β_1	0.0187	0.0209	0.0186	0.0290
	β_2	−0.0012	0.0126	−0.0003	0.0186
	β_3	0.0007	0.0126	0.0005	0.0194
	β_4	0.0003	0.0122	−0.0001	0.0190
	β_5	0.0007	0.0130	0.0001	0.0185
Logit 2FE	β_1	0.0007	0.0159	0.0011	0.0226
	β_2	−0.0008	0.0128	−0.0002	0.0188
	β_3	0.0008	0.0129	0.0005	0.0193
	β_4	0.0002	0.0118	−0.0002	0.0189
	β_5	0.0006	0.0129	0.0001	0.0184
Small sample: $n = 68$					
Logit	β_1	0.0027	0.0208	0.0025	0.0306
	β_2	−0.0006	0.0134	0.0000	0.0212
	β_3	−0.0007	0.0136	−0.0006	0.0207
	β_4	−0.0005	0.0148	0.0003	0.0218
	β_5	−0.0005	0.0148	−0.0003	0.0220
Logit FE	β_1	0.0392	0.0420	0.0392	0.0604
	β_2	0.0020	0.0268	0.0006	0.0393
	β_3	0.0005	0.0254	0.0002	0.0377
	β_4	0.0004	0.0253	0.0005	0.0392
	β_5	−0.0021	0.0256	−0.0017	0.0380
Logit 2FE	β_1	0.0031	0.303	0.0038	0.0459
	β_2	0.0027	0.0272	0.0005	0.0386
	β_3	−0.0002	0.0276	0.0005	0.0376
	β_4	0.0010	0.0251	0.0013	0.0388
	β_5	−0.0024	0.0252	−0.0016	0.0373

The true coefficients are as follows: $\beta_1 = 1$ and $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

Table 2: Design 2 - uncorrelated fixed effects

		Median bias	MAD	Mean bias	RMSE
Full sample: n = 136					
Logit	β_1	-0.2448	0.2448	-0.2455	0.2490
	β_2	0.0032	0.0222	0.0014	0.0332
	β_3	0.0018	0.0222	0.0003	0.0321
	β_4	0.0006	0.0218	0.0003	0.0324
	β_5	0.0005	0.0217	0.0002	0.0327
Logit FE	β_1	0.0201	0.0225	0.0207	0.0320
	β_2	-0.0005	0.0134	-0.0004	0.0204
	β_3	-0.0009	0.0138	-0.0010	0.0204
	β_4	0.0003	0.0140	-0.0005	0.0200
	β_5	0.0003	0.0137	0.0007	0.0208
Logit 2FE	β_1	0.0019	0.0174	0.0021	0.0252
	β_2	-0.0001	0.0139	-0.0002	0.0209
	β_3	-0.0012	0.0139	-0.0011	0.0207
	β_4	0.0000	0.0138	-0.0006	0.0199
	β_5	0.0001	0.0142	0.0005	0.0210
Small sample: n = 68					
Logit	β_1	-0.2408	0.2408	-0.2414	0.2492
	β_2	-0.0007	0.0337	0.0001	0.0495
	β_3	-0.0024	0.0319	-0.0018	0.0466
	β_4	0.0005	0.0336	0.0004	0.0492
	β_5	-0.0024	0.0341	-0.0018	0.0434
Logit FE	β_1	0.0403	0.0439	0.0402	0.0635
	β_2	0.0006	0.0287	0.0008	0.0427
	β_3	-0.0002	0.0284	0.0008	0.0410
	β_4	0.0003	0.0297	-0.0007	0.0429
	β_5	-0.0034	0.0283	-0.0018	0.0434
Logit 2FE	β_1	0.0005	0.0325	0.0019	0.0493
	β_2	0.0000	0.0302	0.0013	0.0426
	β_3	0.0006	0.0276	0.0013	0.0401
	β_4	-0.0022	0.0298	-0.0008	0.0437
	β_5	-0.0029	0.0279	-0.0020	0.0432

Design 2 has fixed effects drawn from a random normal uncorrelated with the x's and the true coefficients are as follows: $\beta_1 = 1$ and $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

Table 3: Design 3 - correlated fixed effects

		Median bias	MAD	Mean bias	RMSE
Full sample: n = 136					
Logit	β_1	0.4986	0.4986	0.4987	0.4992
	β_2	−0.0917	0.0917	−0.0915	0.0929
	β_3	−0.0112	0.0121	−0.0114	0.0166
	β_4	0.0016	0.0087	0.0016	0.0127
	β_5	0.0083	0.0105	0.0083	0.0155
Logit FE	β_1	0.0266	0.0283	0.0256	0.0384
	β_2	0.0241	0.0257	0.0244	0.0366
	β_3	−0.0013	0.0168	−0.0016	0.0243
	β_4	−0.0007	0.0152	−0.0006	0.0236
	β_5	−0.0002	0.0159	−0.0001	0.0242
Logit 2FE	β_1	0.0025	0.0202	0.0014	0.0306
	β_2	0.0002	0.0196	0.0006	0.0296
	β_3	−0.0007	0.0172	−0.0012	0.0250
	β_4	−0.0003	0.0163	0.0000	0.0249
	β_5	−0.0005	0.0164	0.0005	0.0251
Small sample: n = 68					
Logit	β_1	0.4918	0.4918	0.4951	0.4974
	β_2	−0.0902	0.0902	−0.0899	0.0972
	β_3	−0.0102	0.0203	−0.0119	0.0315
	β_4	0.0008	0.0211	0.0011	0.0294
	β_5	0.0102	0.0218	0.0108	0.0316
Logit FE	β_1	0.0491	0.0521	0.0504	0.0769
	β_2	0.0552	0.0585	0.0538	0.0807
	β_3	−0.0027	0.0349	−0.0010	0.0512
	β_4	0.0014	0.0342	0.0005	0.0503
	β_5	−0.0023	0.0346	−0.0007	0.0513
Logit 2FE	β_1	−0.0011	0.0416	0.0011	0.0613
	β_2	0.0042	0.0434	0.0045	0.0632
	β_3	−0.0024	0.0358	0.0000	0.0523
	β_4	0.0020	0.0367	0.0004	0.0522
	β_5	−0.0041	0.0334	−0.0017	0.0517

Design 3 has fixed effects correlated with x_1 but not with the other explanatory variables and the true coefficients are as follows: $\beta_1 = \beta_2 = 1$ and $\beta_3 = \beta_4 = \beta_5 = 0$.

Table 4: Design 4 - uncorrelated fixed effects and resampling

		Median bias	MAD	Mean bias	RMSE
Full sample: n = 136					
Logit	β_1	−0.2455	0.2455	−0.2447	0.2479
	β_2	−0.0009	0.0221	−0.0008	0.0323
	β_3	0.0010	0.0218	0.0003	0.0328
	β_4	−0.0003	0.0229	−0.0003	0.0323
	β_5	0.0011	0.0221	0.0009	0.0326
Logit FE	β_1	0.0200	0.0221	0.0199	0.0309
	β_2	−0.0003	0.0128	−0.0003	0.0198
	β_3	0.0006	0.0137	0.0000	0.0208
	β_4	0.0021	0.0143	0.0016	0.0207
	β_5	0.0005	0.0126	0.0008	0.0204
Logit 2FE	β_1	0.0015	0.0163	0.0008	0.0242
	β_2	0.0008	0.0134	−0.0001	0.0202
	β_3	0.0004	0.0144	−0.0002	0.0212
	β_4	0.0016	0.0142	0.0014	0.0207
	β_5	0.0005	0.0138	0.0006	0.0206
Small sample: n = 68					
Logit	β_1	−0.2447	0.2447	−0.2430	0.2493
	β_2	−0.0013	0.0327	0.0002	0.0487
	β_3	0.0011	0.0350	0.0010	0.0501
	β_4	−0.0015	0.0341	−0.0012	0.0501
	β_5	0.0045	0.0330	0.0019	0.0469
Logit FE	β_1	0.0403	0.0449	0.0410	0.0637
	β_2	0.0021	0.0306	0.0021	0.0438
	β_3	−0.0012	0.0275	0.0003	0.0404
	β_4	0.0017	0.0274	0.0005	0.0409
	β_5	0.0026	0.0288	0.0020	0.0431
Logit 2FE	β_1	−0.0011	0.0340	0.0016	0.0490
	β_2	0.0012	0.0304	0.0016	0.0436
	β_3	0.0009	0.0275	0.0010	0.0402
	β_4	0.0004	0.0260	0.0007	0.0403
	β_5	0.0033	0.0283	0.0023	0.0429

Design 4 has fixed effects drawn from a random normal uncorrelated with the x's, there is resampling of the x's and the fixed effects, and the true coefficients are as follows: $\beta_1 = 1$ and $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

Finally, we look more closely at the effect that sample size has on the relative performance of the Logit FE and the Logit 2FE in the presence of correlated fixed effects (design 3). We chose this design because it is likely to be the most relevant, and because looking at the previous tables, it is, unsurprisingly, where we observe the largest differences in performance. We do not report results for the regular logit since it is severely biased[9] and less commonly used when the importance of fixed effects is suspected. Figure 1 shows how the median bias on the coefficient of the correlated explanatory variable (on β_1) of both estimators evolves as we increase sample size. The top line represents the Logit FE and the bottom line the Logit 2FE. For small samples, the median bias of the Logit 2FE is much smaller than that of the Logit FE. This difference decreases as sample size increases, but does not completely disappear for values of n as large as 200. This is of particular interest to the international trade literature, where the number of countries is close to 200. We therefore see that appropriately dealing with the two fixed effects will always be important for those concerned with unbiased estimation of this type of model with data sets involving countries. Further evidence with international trade data will be provided in the next section. Moreover, Figure 2 shows that the MAD, our measure of precision, is very similar for both estimators in all sample sizes. In short, these Monte Carlo simulations confirm that the Logit 2FE presented in this paper is less biased than, as precise as, and more robust to different fixed effects than other logit estimators. We now move on to applying this estimator to trade data.

---

[9]Results are available on demand.

Figure 1: Median bias of the Logit 2FE and Logit FE for different sample sizes
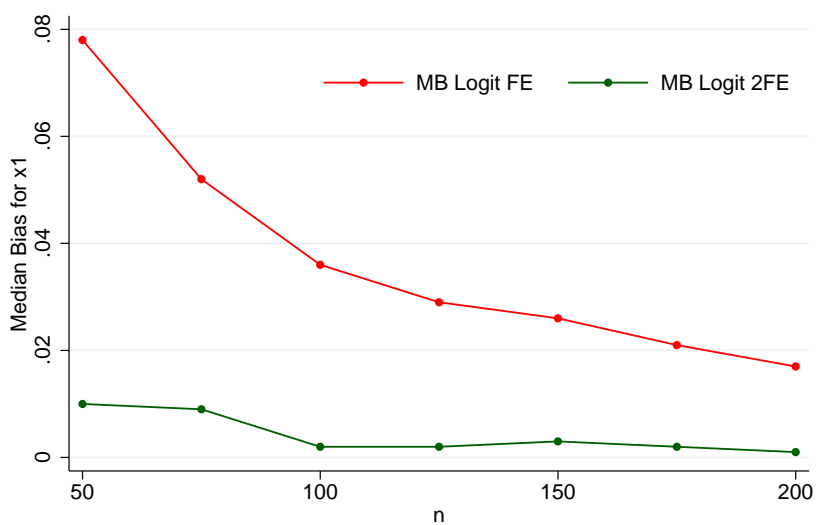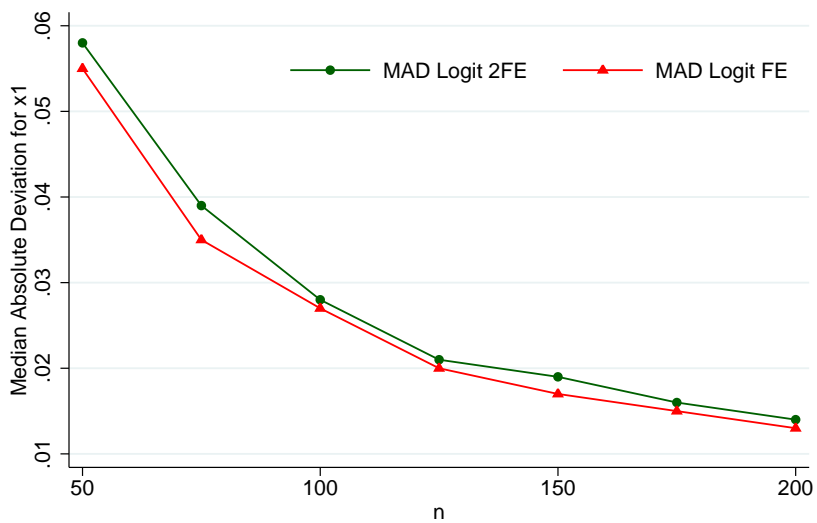

Figure 2: MAD of the Logit 2FE and Logit FE for different sample sizes

17

4 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[10] and has derived gravity equations for these extensive margin adjustments (see, for example, Bernard, Redding and Schott 2011, and Mayer, Melitz and Ottaviano 2012). For the purpose of this application, we will refer to the model where the dependent variable is binary (trade or not) as the binary gravity equation.

4.1 Data

The application of the Logit 2FE estimator on the binary gravity equation is done on two different data sets. First, we use the CEPII data (both the BACI and gravity data sets) for 2005. This trade database is widely used in the literature, as in Head and Mayer (2013), because it is not only from a reputable source, but also contains information on trade between most countries of the world. We illustrate the estimation of two-way fixed effects using data on bilateral trade and including importer and exporter fixed effects for a single year. The year 2005 was chosen based on data availability, but results are similar for other years.

---

[10]Trade frictions have an impact on aggregate trade flows both through 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.

After merging the bilateral trade data with data on country characteristics such as distance, common border or colonial status, we obtain a balanced panel of 211 countries that account for the majority of world trade.

Second, to allow direct comparison with results found in the literature, we use the data from Helpman, Melitz and Rubinstein (2008). Similar to the CEPII data, 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 Logit 2FE with those produced by other commonly used estimators.

Table 5: Summary statistics

Panel A				
	Trade	Trade1	Trade3	Trade5
% zeros	44.1729	51.2999	58.0749	61.7513
Panel B				
	Range	Mean	Median	Std. Dev.
pn if $Trade = 1$	[1,4941]	273.738	20	647.315
dni	[6,208]	117.237	120	50.5608
dnj	[8,208]	117.237	114	55.9579

pn is the number of products (defined at the HS 6-digit level)
traded.
$Trade = 1$ if $pn > 0$, otherwise 0.
$Trade1 = 1$ if $pn > 1$, otherwise 0.
$Trade3 = 1$ if $pn > 3$, otherwise 0.
$Trade5 = 1$ if $pn > 5$, otherwise 0.
$dni(j)$: number of countries the importer (exporter) imports
from (exports to).

Table 5 gives summary statistics for the CEPII data. More detailed information about the data used by Helpman, Melitz and Rubinstein (2008) can be found in their paper. Among the 211 countries of the CEPII database, about 44 per cent of unilateral trade flows are zeros. Note that this is in line with what Helpman, Melitz and Rubinstein find for a sample of 158 countries. If we restrict the definition of positive trade to flows of more than one product, we

see that the number of zeros exceeds 50 per cent (and more for three or five products). This restriction is used as a robustness check in the logit estimation. The prominence of zeros in international trade further motivates the use of econometric methodologies that account for those zero trade flows.

Among countries that do trade, we see that the product number ranges from 1 to 4,941,[11] with a mean of 274 products, but a median of 20. Approximately 12 per cent of positive trade flows are only composed of 1 product, while only 26 per cent of trade flows have more than 150 products. Finally, countries import from and export to 117 other countries on average.

4.2   Logit estimation

There are many "zeros and one" 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 an influential paper, Helpman, Melitz and Rubinstein (2008) 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 in 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)[12] model with French firm-level data, Crozet and Koenig (2010) also use the probability of exporting as a first stage in their empirical strategy. 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.

The following applications show that papers like Helpman, Melitz and Rubinstein (2008) and Crozet and Koenig (2010), as well as any other that uses a similar framework, would find

---

[11]There are 5,017 HS 6-digit products. Note also that, not surprisingly, the two countries that trade the largest number of products (4,941) are the United States and Canada.

[12]This paper 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.

significantly different estimates if they applied the conditional logit to properly account for the multiple fixed effects. Indeed, estimates can differ significantly. As mentioned before, we will demonstrate this using two different models and data sets: first, comparing estimates for a somewhat traditional gravity model applied to the popular CEPII data, then by performing the estimation using the same model and data as Helpman, Melitz and Rubinstein (2008).

For our first step, using the CEPII data, we write the probability of country j exporting to country i as:

$$\text{Prob}[Trade_{ij}] = \beta_0 + \beta_1 \ln(D_{ij}) + \beta_2 Border_{ij} + \beta_3 Legal_{ij} + \beta_4 Language_{ij} \tag{12}$$
$$+ \beta_5 Colony_{ij} + \beta_6 Currency_{ij} + \beta_7 RTA_{ij} + \mu_i + \alpha_j + \varepsilon_{ij},$$

where D_{ij} is the simple distance between country i's and country j's most populated cities, $Border_{ij}$ is a dummy that takes the value 1 if i and j share a border, $Legal_{ij}$ is a dummy that takes the value 1 if the two countries have the same legal system, $Language_{ij}$ 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, $Currency_{ij}$ is a dummy that takes the value 1 if the two countries use the same currency, RTA_{ij} is a dummy that takes the value 1 if i and j are in a regional trade agreement, and, finally, μ_i and α_j are respectively importer and exporter fixed effects. The results are presented in Table 6.

The results for β_1[13] differ greatly between estimators, more than what is expected when looking at the Monte Carlo simulations. Both estimators being relatively precise, this difference is puzzling and could indicate model misspecification. We will return to this issue at the end of the section. However, it also suggests that distance might have a smaller impact on the probability of exporting than what traditional estimates indicate. Of course, caution should be exercised when comparing the results from this application with those of the Monte Carlo simulations since it is not clear how the data relate to the distributional assumptions made in the latter. Note that the Logit 2FE estimated effect of distance on the

---

[13]In what follows we mostly discuss results for the coefficient on distance, because it is generally the most talked-about trade barrier. *Border* is also commonly discussed, but as it is not significant in our estimations, it is of lesser interest.

Table 6: Logit results (benchmark)

Variables	OLS FE	Logit	Logit FE	Logit 2FE
Distance	−0.1116	−0.6491	−1.2146	−0.2282
	(0.0028)	(0.0169)	(0.0329)	(0.0461)
Border	−0.0765	0.0620	−0.4194	1.4751
	(0.0159)	(0.1606)	(0.2424)	(0.2743)
Legal	0.0314	0.4098	0.2662	0.2989
	(0.0044)	(0.0247)	(0.0422)	(0.0650)
Language	0.0756	−0.4588	0.7870	0.9452
	(0.0056)	(0.0283)	(0.0597)	(0.0824)
Colonial ties	0.0377	3.5425	0.8699	1.9481
	(0.0149)	(0.3251)	(0.4086)	(0.5377)
Currency	0.0540	−0.5732	0.5458	0.0331
	(0.0198)	(0.1386)	(0.2055)	(0.0066)
RTA	−0.0959	2.2365	1.3113	1.4140
	(0.0082)	(0.1120)	(0.1529)	(0.1788)

These results are for $\text{Prob}[Trade = 1]$ where $Trade = 1$ when the number of products traded is greater than 0.

OLS FE refers to a simple linear probability model with fixed effects. The results in the second column are for the regular logit ignoring the fixed effects, the results in the third column are for the logit estimating all the fixed effects and finally, the results in the fourth column are for the estimator presented in this paper.

Standard errors clustered at ij level (allowing for importer and exporter correlation).

probability of exporting is closer to that produced by the linear probability model than that of the logit estimating the fixed effects.

Generally, the estimated effects of the other variables on the probability of exporting differ for the Logit 2FE and the Logit FE, but this difference, unlike that for distance, is much closer to what the Monte Carlo simulations suggested. One notable exception is the effect of a common border. Indeed, like the probit in Helpman, Melitz and Rubinstein (2008), the Logit FE produces a surprising negative effect of sharing a border on the probability of trading. However, the Logit 2FE's positive estimate suggests that this result might be due to the inconsistency of the estimator. Note that the Logit 2FE is the only one of the four estimators where all estimated coefficients match their expected sign.

As a robustness check and to make sure that none of these estimators were strongly influenced by one outlier country, they were each calculated dropping each country in succession. The results are all very similar. Defining the dependent variable as "0" or "1" can give a heavy weight to relatively small trade flows: a country pair trading one product gets the same dependent variable value as a country pair trading 1,000 products. To test the robustness of the estimators to that issue, we recalculate all of them using three different definitions of positive trade (more than one product, more than three, and more than five). The results are presented in Table 7.

As predicted, considering trade as positive only when the number of products exported is greater than one increases the impact of distance on the probability of exporting. Accounting for a rough measure of trade flow size suggests that to trade many products, countries do have to be closer. The effect of distance becomes more and more negative as we increase the number of zeros by changing the $Trade$ variable, which is true for all estimators except the regular logit. This could be because, as the Monte Carlo simulations illustrate, the bigger the coefficient (in absolute value), the bigger the bias. Though respective coefficients for both the Logit 2FE and the Logit FE move in the same direction, they are moving farther apart, thus emphasizing our original concern about their large difference. As a final robustness check, we have done the same calculations with different measures of distance (i.e., distance between capital cities, distance weighted by population). It does not affect the results significantly.

We now turn to the Helpman, Melitz and Rubinstein data and estimate their gravity model:

$$
\begin{aligned}
\text{Prob}[Trade_{ij}] = {} & \beta_0 + \beta_1 \ln(D_{ij}) + \beta_2 Border_{ij} + \beta_3 Island_{ij} \\
& + \beta_4 Landlock_{ij} + \beta_5 Legal_{ij} + \beta_6 Language_{ij} + \beta_7 Colony_{ij} \\
& + \beta_8 Currency_{ij} + \beta_9 RTA_{ij} + \beta_{10} Religion_{ij} + \mu_i + \alpha_j + \varepsilon_{ij},
\end{aligned}
\tag{13}
$$

where common variables are defined as in equation (12), $Island_{ij}$ is a dummy that takes the value 1 if either one or both countries are islands, and $Landlock_{ij}$ is a dummy that takes the value 1 if either one or both countries do not have access to an ocean. The religion variable

Table 7: Logit results (robustness check)

	Variables	OLS FE	Logit	Logit FE	Logit 2FE
Panel A: Trade1	Distance	−0.1185	−0.6389	−1.4553	−0.2502
		(0.0028)	(0.0163)	(0.0371)	(0.0492)
	Border	−0.0517	0.3069	−0.1719	1.6626
		(0.0170)	(0.1542)	(0.2532)	(0.3545)
	Legal	0.0225	0.3283	0.2721	0.3055
		(0.0043)	(0.0242)	(0.0440)	(0.0731)
	Language	0.0945	−0.4194	0.9819	1.1447
		(0.0054)	(0.0284)	(0.0638)	(0.0942)
	Colonial ties	0.0491	3.5133	1.1530	2.2784
		(0.0155)	(0.2759)	(0.3703)	(0.5831)
	Currency	0.0334	−0.6519	0.6564	0.0372
		(0.0205)	(0.1315)	(0.2062)	(0.0086)
	RTA	−0.0524	2.1052	1.2354	1.6153
		(0.0084)	(0.0931)	(0.1402)	(0.1675)
Panel B: Trade3	Distance	−0.1207	−0.6184	−1.7427	−0.2751
		(0.0028)	(0.0159)	(0.0428)	(0.0505)
	Border	−0.0142	0.5755	0.2205	1.7963
		(0.0178)	(0.1512)	(0.2772)	(0.4537)
	Legal	0.0196	0.2644	0.3879	0.3772
		(0.0042)	(0.0243)	(0.0472)	(0.0833)
	Language	0.1027	−0.3710	1.0679	1.2772
		(0.0052)	(0.0290)	(0.0683)	(0.0989)
	Colonial ties	0.0517	3.1438	1.1043	2.1905
		(0.0160)	(0.3251)	(0.3315)	(0.5520)
	Currency	0.0235	−0.5948	1.0211	0.0407
		(0.0208)	(0.1271)	(0.2249)	(0.0085)
	RTA	0.0065	2.0604	1.2449	1.8341
		(0.0086)	(0.0812)	(0.1389)	(0.1673)
Panel C: Trade5	Distance	−0.1165	−0.5965	−1.8686	−0.2788
		(0.0028)	(0.0158)	(0.0463)	(0.0535)
	Border	0.0100	0.6822	0.4508	1.9881
		(0.0179)	(0.1460)	(0.2817)	(0.4265)
	Legal	0.0151	0.2328	0.4026	0.3838
		(0.0041)	(0.0246)	(0.0495)	(0.0928)
	Language	0.1104	1.2253	0.7870	1.4356
		(0.0051)	(0.0296)	(0.0728)	(0.0898)
	Colonial ties	0.0538	3.0461	1.0601	1.9623
		(0.0165)	(0.1843)	(0.3205)	(0.7479)
	Currency	0.0166	−0.5604	1.2330	0.0372
		(0.0208)	(0.1260)	(0.2449)	(0.0092)
	RTA	0.0448	2.0147	1.2473	1.9366
		(0.0086)	(0.0747)	(0.1405)	(0.1840)

is defined as follows:

$$Religion_{ij} = (\%\text{Protestants in country } i \cdot \%\text{Protestants in country } j)$$
$$+ (\%\text{Catholics in country } i \cdot \%\text{Catholics in country } j)$$
$$+ (\%\text{Muslims in country } i \cdot \%\text{Muslims in country } j).$$

The results are reported in Table 8. 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 the original paper (Helpman, Melitz and Rubinstein 2008). This is done to allow a more direct comparison of our results with theirs.

Table 8: 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	−0.5184
	(0.0239)	(0.0041)	(0.0192)	(0.0424)	(0.0788)
Border	−0.3825	−0.0810	0.2834	−0.7624	0.2503
	(0.0993)	(0.0218)	(0.1184)	(0.1761)	(0.2156)
Island	−0.3447	−0.0648	−0.3355	−0.6030	0.0298
	(0.0743)	(0.0135)	(0.0287)	(0.1359)	(0.2221)
Landlock	−0.1806	−0.0316	−0.6225	−0.3607	0.1395
	(0.0973)	(0.0156)	(0.0308)	(0.1808)	(0.1667)
Legal	0.0964	−0.0039	−0.4939	0.1867	0.1494
	(0.0298)	(0.0053)	(0.0289)	(0.0539)	(0.0718)
Language	0.2838	0.0783	0.1208	0.5037	0.5938
	(0.0378)	(0.0068)	(0.0313)	(0.0696)	(0.1219)
Colonial ties	0.3252	0.0408	4.0701	0.5392	2.4803
	(0.2444)	(0.0196)	(0.4165)	(0.4647)	(0.8262)
Currency	0.4917	0.0772	−0.7229	0.8973	1.3338
	(0.1227)	(0.0243)	(0.1497)	(0.2249)	(0.4543)
RTA	1.9851	−0.0526	2.9708	3.4708	4.5319
	(0.2651)	(0.0413)	(0.4193)	(0.4928)	(0.4460)
Religion	0.2605	0.0739	0.3289	0.4155	0.9521
	(0.0583)	(0.0108)	(0.0478)	(0.1056)	(0.2045)

Standard errors clustered at ij level (allowing for importer and exporter correlation).

The results for the Logit FE closely match those of the Probit FE, which in turn are a successful replication of the estimates presented in Helpman, Melitz and Rubinstein (2008). As an illustration of that, Figure 3 shows how the probability of trading, estimated at the sample averages, changes with the measure of distance: the two bottom lines represent the Logit FE and Probit FE and we observe that they trace each other almost exactly. Therefore, what is implied by the comparison of the Logit FE with the Logit 2FE also applies to the results of Table 1 in Helpman, Melitz and Rubinstein (2008). When comparing Tables 6 and 8, the first thing we notice is that the difference between the distance coefficients of the Logit FE and 2FE is much smaller in the latter. Indeed, when using the HMR model and data, the estimate for the Logit 2FE has almost doubled, while that of the Logit FE is not statistically significantly different. Moreover, the effect of sharing a border still has its puzzling negative (and statistically significant) effect when looking at the Logit FE, but is now no longer significant for the Logit 2FE. In general, the change in model and data seems to have had a slightly larger effect on the estimates produced by the Logit 2FE and some of those coefficients are now much closer to those of the Logit FE. Of the three new variables, only religion is significant for the logit 2FE, while *Island* and *Landlock* are negative and statistically significant for the other estimators.

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 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 3 for the Probit FE, Logit FE and Logit 2FE (top line). As expected, the Logit 2FE predicts a much higher probability of trading for all values of distance, particularly in the middle ranges. Note that Figure 3 represents all distance values in the sample, but that most observations have a log of distance in the range [3,5], with a mean of 4.1.

Figure 3: Probability of trading relative to distance
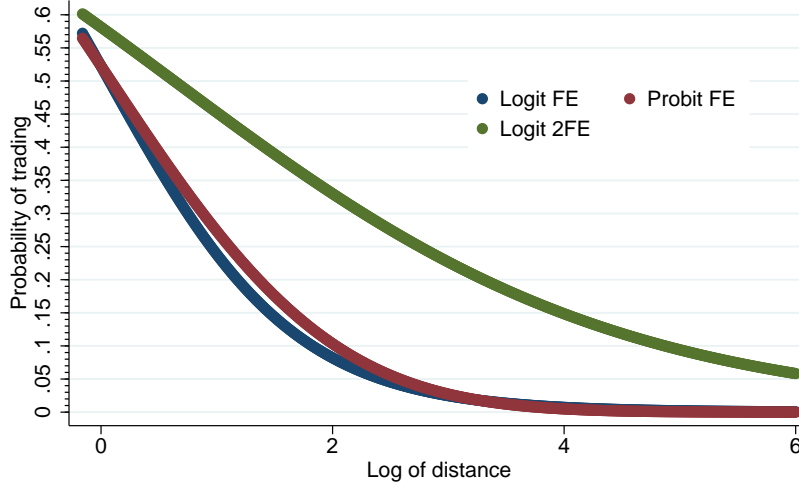
Finally, in order to identify the source (i.e., data or model) of the difference in estimates between Table 6 and Table 8, we estimated our benchmark model given by equation (12) on the HMR data. Results are reported in Table 9. It appears that, as expected, the data, rather than the addition of the three variables, are responsible for the change in coefficients. Interestingly, both the Logit FE and the Logit 2FE suggest that distance might have had a larger impact on the probability of trading in 1986 than in 2005. This difference is particularly obvious, and statistically significant, for the Logit 2FE, which could imply that, as many have conjectured, the effect of distance is decreasing over the years.

Overall, the large difference between estimators, especially between the Logit 2FE and the logit estimating the fixed effects, shows the importance of properly accounting for multiple fixed effects. On the one hand, all the results suggest that the impact of distance on the probability of exporting could be smaller than what we thought, especially in more recent years. On the other hand, however, this large difference can be cause for concern. Indeed, its magnitude is not in line with the Monte Carlo simulations. Therefore, it might indicate that the model is misspecified. If, for example, different countries had different β_1 or if distance

Table 9: Logit results (HMR data, benchmark model)

Variables	OLS FE	Logit	Logit FE	Logit 2FE
Distance	-0.1141	-0.3111	-1.2924	-0.5601
	(0.0041)	(0.0184)	(0.0412)	(0.0792)
Border	-0.0744	0.3312	-0.7355	0.3477
	(0.0219)	(0.1165)	(0.1759)	(0.2046)
Legal	-0.0025	-0.4311	0.1994	0.1692
	(0.0053)	(0.0283)	(0.0539)	(0.0690)
Language	0.0875	0.1386	0.5695	0.7088
	(0.0068)	(0.0304)	(0.0679)	(0.1187)
Colonial ties	4.0476	3.5425	0.5609	2.6637
	(0.0196)	(0.4174)	(0.4670)	(0.8239)
Currency	0.0676	-0.8877	0.8141	1.2023
	(0.0242)	(0.1512)	(0.2243)	(0.4560)
RTA	-0.0500	3.0358	3.4869	4.5281
	(0.0415)	(0.4177)	(0.5224)	(0.4182)

Standard errors clustered at ij level (allowing for importer and exporter correlation).

was in fact interacted with something else, then both estimators would give an average β_1. Since each gives different weights to the same observations, this could explain why the estimates differ so much. It could then imply that the true β_1 is quite different from all the estimates. As a preliminary check, we relaxed the functional form assumptions on distance by using non-parametric dummies for quartiles of the distance distribution. The results did not suggest any problem and were very much in line with the original results presented in Tables 6 and 8. Be that as it may, and whether or not the model is misspecified, this application illustrates the pertinence of computing the conditional logit estimator for two fixed effects.

5 Conclusion

This paper 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 developed a method to appropriately deal with two fixed effects for the logit model, and showed that Manski's maximum score estimator cannot be adapted for multiple fixed effects.

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), 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. However, we found that if the same conditioning can be used on Manski's maximum score estimator with a single fixed effect, the method fails when there are two. That estimator cannot be generalized to the case of two fixed effects. If, in principle, the recursive structure of the logit model could be used to apply the conditioning exercise more than twice, hence dealing with more than two fixed effects, it is the lack of recursive structure in Manski's maximum score estimator that explains its inability to adapt to multiple fixed effects.

We showed that the conditioning method that is the core of this paper performs well in recovering the true parameters in Monte Carlo studies. Indeed, we 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, for example, the size that one can get when studying trade between countries. We also showed that this same procedure yields quite different estimated coefficients from methods subjected to the incidental parameters problem in applications with actual trade data. To demonstrate this, we applied the logit to a gravity model with importer and exporter fixed effects.The coefficients estimated with our procedure may explain some of the "puzzles" this literature has encountered. Indeed,

we find that distance has a smaller impact on the probability of trading, especially in recent years, and that border has a positive effect rather than the suspicious negative effect found with usual estimators.

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.

This project opens two areas for future work. First, there are many other widely used nonlinear panel data models suffering from the incidental parameter problem. For example, estimators such as the Poisson and Negative Binomial maximum likelihood of Hausman, Hall and Griliches (1984) are commonly applied with multiple fixed effects. It would therefore be important to determine whether the conditional maximum likelihood used in the present paper for the logit can be similarly adapted for those estimators. We are currently working on this project. Finally, the asymptotic properties of the estimator developed in this paper need to be explored.

References

AARONSON, D., L. BARROW, AND W. SANDER (2007): "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 25(1), 95–135.

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

ANDERSON, J. E., AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *The American Economic Review*, 93(1), 170–192.

ARELLANO, M., AND S. BONHOMME (2009): "Robust Priors in Nonlinear Panel Data Models," *Econometrica*, 77(2), 489–536.

ARELLANO, M., AND B. E. HONORÉ (2001): "Panel Data Models: Some Recent Developments," *Handbook of Econometrics*, 5(53), 3229–3296.

BERNARD, A., J. B. JENSEN, S. REDDING, AND P. SCHOTT (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105–130.

BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2011): "Multiproduct Firms and Trade Liberalization," *The Quarterly Journal of Economics*, 126(3), 1271–1318.

BONHOMME, S. (2012): "Functional Differencing," *Econometrica*, 80(4), 1337–1385.

CHANEY, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *The American Economic Review*, 98(4), 1707–1721.

CROZET, M., AND P. KOENIG (2010): "Structural Gravity Equations with Intensive and Extensive Margins," *Canadian Journal of Economics*, 43(1).

FERNANDEZ-VAL, I., AND M. WEIDNER (2011): "Individual and Time Effects in Nonlinear Panel Data Models with Large N, T," *Unpublished manuscript*.

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72(4), 1295–1319.

HAUSMAN, J., B. H. HALL, AND Z. GRILICHES (1984): "Econometric Models for Count Data with an Application to the Patents-RD Relationship," *Econometrica*, 52(4), 909–938.

HEAD, K., AND T. MAYER (2013): "Gravity Equations: Workhorse, Toolkit and Cookbook," *CEPR Discussion Paper*.

HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, CXXIII(2), 441–487.

HIRJI, K. F., C. R. MEHTA, AND N. R. PATEL (1987): "Computing Distributions for Exact Logistic Regression," *Journal of the American Statistical Association*, 82(400), 1110–1117.

KIRABO JACKSON, C. (2013): "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers," *The Review of Economics and Statistics*, 95(4), 1096–1116.

MANSKI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3(3), 205–228.

——— (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357–362.

MAYER, T., M. MELITZ, AND G. I. P. OTTAVIANO (2012): "Market Size, Competition, and the Product Mix of Exporters," *Working Paper*.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocation and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

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(6), 2295–2350.

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

——— (1961): "On the General Laws and the Meaning of Measurement in Psychology," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*.

RIVKIN, S. G., E. A. HANUSHEK, AND J. F. KAIN (2005): "Teachers, Schools, and Academic Achievement," *Econometrica*, 73(2), 417–458.

SANTOS SILVA, J. M. C., AND S. TENREYRO (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.