Two-way models for gravity

Koen Jochmans*

Draft: February 2, 2015

Empirical models for panel data frequently feature fixed effects in both directions of the panel. Settings where this is prevalent include student-teacher interaction, the allocation of workers to firms, and the import-export flows between countries. Estimation of such fixed-effect models is difficult. We derive moment conditions for models with multiplicative unobservables and fixed effects and use them to set up generalized method of moments estimators that have good statistical properties. We estimate a gravity equation with multilateral resistance terms as an application of our methods.

Introduction

A common type of data provides information on dyadic relations between two sets of agents. For example, in the student-achievement literature (see, e.g., Aaronson, Barrow and Sander 2007; Rivkin, Hanushek and Kain 2005), the sets of agents could be students on the one hand, and teachers or classrooms on the other hand. Alternatively, the data of Abowd, Kramarz and Margolis (1999) matches workers with firms, while data on bilateral trade flows has been used to study import/export behavior of firms or countries at least since the work of Tinbergen (1962).

Empirical models for such linked data almost invariably contain agent-specific fixed effects, that is, teacher and student effects, worker and firm effects, or importer and exporter effects in the applications just mentioned. The inclusion of such effects can capture unobserved characteristics that are heterogeneous across agents. In our examples these could be student quality and teacher ability or

* Address: Sciences Po, Département d'économie, 28 rue des Saints Pères, 75007 Paris, France; koen.jochmans@sciencespo.fr.

dedication, worker and firm productivity, and openness toward international trade or lenience toward importing and exporting goods, respectively. Furthermore, including such effects can be seen as a parsimonious approach to capturing latent network effects (see, e.g., Graham 2013, 2014).

Linked data of this form can be seen as panel data where each dimension of the panel corresponds to a type of agent. A specification that includes fixed effects in each dimension gives rise to a two-way model for panel data. While the inclusion of such effects is intuitively attractive and widespread, there is little theoretical work on the statistical properties of the corresponding estimators. Moreover, pooled estimators and also fixed-effect estimators that are popular for one-way models may lose their attractive properties. Indeed, in most linked data sets, both dimensions of the panel tend to be reasonably large. Under such a sampling scheme, estimators that treat the fixed effects as parameters to be estimated will, in general, be heavily biased (see, e.g., Hahn and Newey 2004). This implies that the usual confidence intervals and hypothesis tests have poor properties, which may lead to erroneous policy conclusions.

In this paper we consider estimation and inference for a class of nonlinear models with two-way fixed effects and disturbances that enter in a multiplicative manner. A notable member of this class is the exponential-regression model. In addition to its usefulness for analyzing count data, the exponential-regression model is popular in the estimation of constant-elasticity models ever since Santos Silva and Tenreyro (2006) pointed to the potentially-large bias in least-squares estimates of log-linearized models. We derive moment conditions that difference-out the fixed effects and use them to construct GMM estimators that have standard asymptotic properties.

There is recent work by Fernández-Val and Weidner (2014) on likelihood-based estimation of a class of two-way models. This class of models is different from the one under study here, and they are not nested. An important difference between their approach and ours is that they consider methods for reducing the order of the incidental-parameter bias in the maximum-likelihood estimator while we are able to eliminate it completely. Charbonneau (2013) investigates the potential of conditional-likelihood estimation for the binary-choice logit model as well as for Poisson and Gamma models. The latter two models are special cases of our model. Furthermore, the conditional-likelihood approach requires specifying the full distribution of the data (conditional on the fixed effects) while our approach is moment based.

A leading two-way model is a gravity model for international trade featuring both importer and exporter fixed effects. Such a specification is arguably the benchmark gravity equation since Anderson and van Wincoop (2003), but the literature has not reached an agreement on how to estimate it (Anderson 2011; Head and Mayer 2014).[1] As an empirical application we estimate a version of the Anderson and van Wincoop (2003) model for a panel of 136 countries using the estimator developed here. This makes our estimates of the gravity equation the first ones to be based on a sound statistical theory adapted to the presence of multilateral resistance terms.

In the first section we describe the model of interest and gradually build our way from a model featuring no fixed effects to a two-way model. This is useful as it helps to illustrate the difficulties that arise with traditional techniques when applied to two-way models. In doing so, we derive the moment conditions that will be the basis for our estimator. The second section is dedicated to the estimator. We first present its large-sample distribution, which allows for estimation and inference. We next evaluate its small-sample performance in a set of Monte Carlo experiments. The third section deals with our empirical application. We conclude by showing that our approach readily generalizes to datasets with more than two dimensions, such as data on repeated interactions between the agents or observing agents actions on multiple markets.

[1] The estimators that have been put forth range from naive log-linearized least squares over nonlinear least-squares to Gamma and Poisson pseudo maximum-likelihood estimators. However, their use is based on cross-sectional arguments (Santos Silva and Tenreyro 2006; Head and Mayer 2014), but these do not always carry over to the panel setting with fixed effects.

I. Multiplicative models for panel data

Suppose our aim is to estimate the parameter vector ψ_0 from a panel data set $\{(y_{11}, x_{11}), \ldots, (y_{nm}, x_{nm})\}$ in a model of the form

$$y_{ij} = \varphi(x_{ij}; \psi_0)\, u_{ij}, \tag{1.1}$$

where u_{ij} is an unobserved error term and φ is a function known up to ψ_0.[2] For example, one important special case is an exponential-regression model, where $\varphi(x_{ij}; \psi) = \exp(x_{ij}'\psi_0)$. Whether or not this can be achieved depends on the restrictions we are willing to place on the joint distribution of the regressors $\{x_{ij}\}$ and the errors $\{u_{ij}\}$.

A. Pooled models

If we assume that the errors u_{ij} are mean-independent of all regressors x_{ij}, that is, that the regressors are strictly exogenous, and that $E[u_{ij}] = 1$, we have $E\left[y_{ij} - \varphi(x_{ij}; \psi_0) \,|\, x_{11}, \ldots, x_{nm}\right] = 0$ and a GMM estimator is easily constructed. One obvious and simple example would be the pooled estimator that solves the empirical moment condition

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}(y_{ij} - \varphi(x_{ij}; \psi)) = 0 \tag{1.2}$$

for ψ. In the exponential-regression context, this particular estimator is known as the Poisson pseudo maximum-likelihood estimator (Gouriéroux, Monfort and Trognon 1984), popularized by Santos Silva and Tenreyro (2006). Alternative unconditional moment conditions that lead to more efficient estimators can also be constructed (Chamberlain 1987; Newey 1990; Donald, Imbens and Newey 2003).

[2] An extended model that can be dealt with in the same way has $y_{ij} = \mu(x_{ij}; \theta_0) + \varphi(x_{ij}; \psi_0)\, u_{ij}$ for known function μ and Euclidean parameter θ_0.

B. One-way models

With panel data, the validity of pooled estimators can be hard to justify. Indeed, since the seminal work of Hausman, Hall and Griliches (1984), the importance of controlling for fixed effects is well established. The one-way version of (1.1) equals

$$y_{ij} = \varphi(x_{ij}; \psi_0)\, \alpha_i\, \varepsilon_{ij}, \tag{1.3}$$

that is, u_{ij} is decomposed as $u_{ij} = \alpha_i\, \varepsilon_{ij}$. Not only does the presence of α_i imply serial correlation in the series $u_{i1}, \ldots, u_{im}$ even if the ε_{ij} are jointly independent and independent of α_i, it also implies that $E[u_{ij}|x_{11}, \ldots, x_{nm}] = 1$ no longer holds, in general. This renders pooled estimators inconsistent. For example, if $E[\varepsilon_{ij}|\alpha_1, \ldots, \alpha_n; x_{11}, \ldots, x_{nm}] = 1$, then

$$E[u_{ij}|x_{11}, \ldots, x_{nm}] = E[\alpha_i|x_{11}, \ldots, x_{nm}]$$

and, unless α_i is independent of the covariates and has mean one, in which case $E[y_{ij}|x_{11}, \ldots, x_{nm}] = \varphi(x_{ij}; \psi_0)$, any pooled estimator of ψ_0, such as (1.2), yields inconsistent estimates.

Estimators of ψ_0 that do not restrict the relation between $\alpha_1, \ldots, \alpha_n$ and $x_{11}, \ldots, x_{nm}$ have received some attention. $E[\varepsilon_{ij}|\alpha_1, \ldots, \alpha_n; x_{11}, \ldots, x_{nm}] = 1$ implies that

$$E\left[u_{ij} \middle| \alpha_1, \ldots, \alpha_n; x_{11}, \ldots, x_{nm}\right] = \alpha_i.$$

So in the spirit of Chamberlain (1992) and Wooldridge (1997), we obtain the moment restrictions

$$E\left[u_{ij} - u_{ij'} \middle| x_{11}, \ldots, x_{nm}\right] = 0 \tag{1.4}$$

for all $j, j' \in \{1, \ldots, m\}$. GMM estimators may then again be constructed. In analogy to (1.2) we could, for example, base estimation on

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j<j'}(x_{ij}-x_{ij'})\left(u_{ij}(\psi)-u_{ij'}(\psi)\right)=0, \qquad u_{ij}(\psi)=\frac{y_{ij}}{\varphi(x_{ij};\psi)},$$

but, again, other unconditional moments derived from (1.4) can equally be used. Under regularity conditions, such estimators are consistent and asymptotically normal whether m is treated as fixed or grows with n.

A popular and seemingly different approach to estimate one-way models of the form in (1.3) when the covariates are strictly exogenous is via the Poisson pseudo maximum-likelihood estimator. Under regularity conditions, this estimator is known to be consistent even when m remains fixed while $n\to\infty$ (Wooldridge 1999). Notably, when $\varphi(x;\psi)=\exp(x'\psi)$, that is, for an exponential-regression model with fixed effects, the score equation of the pseudo maximum-likelihood estimator is

$$\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij}\left(y_{ij}-\varphi(x_{ij};\psi)\,\widehat{\alpha}_i(\psi)\right)=0, \qquad \widehat{\alpha}_i(\psi)=\frac{\sum_{j=1}^{m}y_{ij}}{\sum_{j=1}^{m}\varphi(x_{ij};\psi)}. \tag{1.5}$$

Consistency of this fixed-effect estimator follows from the remarkable feature that $E[\widehat{\alpha}_i(\psi_0)|x_{11},\ldots,x_{nm}]=\alpha_i$, and so $\varphi(x_{ij};\psi_0)\,\widehat{\alpha}_i(\psi_0)$ is an unbiased (but inconsistent, as $n\to\infty$ and m remains fixed) estimator of the conditional mean $E[y_{ij}|\alpha_1,\ldots,\alpha_n,;x_{11},\ldots,x_{nm}]=\varphi(x_{ij};\psi_0)\,\alpha_i$. A calculation shows that the score equation can equivalently be written as

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j<j'}(x_{ij}-x_{ij'})\frac{\varphi(x_{ij};\psi)\varphi(x_{ij'};\psi)}{\sum_{j'=1}^{m}\varphi(x_{ij'};\psi)}\left(u_{ij}(\psi)-u_{ij'}(\psi)\right)=0,$$

which is again the sample counterpart of a particular unconditional version of (1.4). Thus, the pseudo maximum-likelihood approach is again nested within the GMM framework. The existence of this formulation provides an alternative explanation for why (pseudo) maximum likelihood is consistent in the presence of incidental parameters in this particular setting.

Note that, even if the full distribution of y_{ij} given x_{ij} and α_i would be specified, the fixed-effect model remains semiparametric because the distribution of the α_i is not specified (contrary to in a random-effect approach). In a semiparametric framework, GMM is arguably the natural estimation paradigm. In the current context, adopting the GMM framework has several advantages that we briefly discuss next.

First, moment conditions can still be obtained when some components of x_{ij} are endogenous but instrumental variables z_{ij} are available. Indeed, in such a case, it is immediate that

$$E\left[u_{ij} - u_{ij'} \,\middle|\, z_{11}, \ldots, z_{nm}\right] = 0,$$

and so everything continues to go through as before on appropriately replacing covariates by instruments. Similarly, flexible methods to correct for endogenous sample selection similar to the two-step approach of Heckman (1979) are equally available for multiplicative-error models estimated from panel data (Jochmans 2015).

Second, valid moment conditions can equally be constructed when the covariates are not strictly exogenous. For example, when there is feedback from y_{ij} to x_{ij+1}, like in a time-series setting, (1.4) no longer holds and pseudo maximum-likelihood estimation is inconsistent. However, we still have

$$E\left[u_{ij} - u_{ij'} \,\middle|\, x_{11}, \ldots, x_{1j'}, \ldots, x_{n1} \ldots, x_{nj'}\right] = 0,$$

for all $j' < j$ (Chamberlain 1992). These sequential moment conditions can again form the basis for GMM estimation in the usual manner. In contrast, pseudo maximum-likelihood estimation fails in the presence of feedback because the profile-score equation in (1.5) is now biased at ψ_0.

C. Two-way models

A two-way model for panel data allows for fixed effects in both dimensions of the panel. In the context of (1.1), a two-way model is

$$y_{ij} = \varphi(x_{ij}; \psi_0)\, \alpha_i\, \gamma_j\, \varepsilon_{ij}, \tag{1.6}$$

where, now, α_i and γ_j represent permanent unobserved effects. The two-way model is attractive because it gives a parsimonious way to deal with aggregate shocks in both directions of the panel and to allow the unobservables $u_{ij} = \alpha_i \gamma_j \varepsilon_{ij}$ to be both heteroskedastic and correlated. Further, it gives a way to control for unobserved heterogeneity in both directions of the data. This latter observation has been an important motivation for their use with matched employer-employee data (Abowd, Kramarz and Margolis 1999), where α_i and γ_j are measures of quality, and with trade data (Harrigan 1996; Anderson and van Wincoop 2003), where they capture so-called multilateral resistance terms. Indeed, the inclusion of fixed effects in both directions has become standard practice in this literature (Head and Mayer 2014).

Despite their popularity in applied work, estimators of two-way models and their asymptotic properties are scarce, especially beyond the linear-regression context. Of course, under asymptotics where either n or m is treated as fixed it is straightforward to set up GMM estimators based on the moment conditions from the previous subsection. However, the typical sampling framework where a two-way model is called for features data where n and m are of the same order of magnitude, calling for an asymptotic scheme where both n and m grow large jointly.

Here we will derive moment conditions for ψ_0 under the maintained assumption of strict exogeneity of the covariates, that is, throughout the remainder we assume

$$E[\varepsilon_{ij} | \alpha_1, \ldots, \alpha_n; \gamma_1, \ldots, \gamma_m; x_{11}, \ldots, x_{nm}] = 1. \tag{1.7}$$

This condition can be relaxed in essentially the same ways as were discussed above. We return to this in the Monte Carlo illustrations below. Now, (1.7) implies that

$$E\left[u_{ij}\right|\alpha_1,\ldots,\alpha_n;\gamma_1,\ldots,\gamma_m;x_{11},\ldots,x_{nm}] = \alpha_i\,\gamma_j$$

for any i,j. Hence, as long as $E[\varepsilon_{ij}\varepsilon_{i'j'}|\alpha_1,\ldots,\alpha_n;\gamma_1,\ldots,\gamma_m;x_{11},\ldots,x_{nm}] = 1$ for different pairs of indices i,j and i',j',

$$E\left[u_{ij}\,u_{i'j'}\middle|\,\alpha_1,\ldots,\alpha_n;\gamma_1,\ldots,\gamma_m;x_{11},\ldots,x_{nm}\right] = (\alpha_i\,\gamma_j)\,(\alpha_{i'}\,\gamma_{j'}) = \alpha_i\alpha_{i'}\,\gamma_j\gamma_{j'},$$
$$E\left[u_{ij'}\,u_{i'j}\middle|\,\alpha_1,\ldots,\alpha_n;\gamma_1,\ldots,\gamma_m;x_{11},\ldots,x_{nm}\right] = (\alpha_i\,\gamma_{j'})\,(\alpha_{i'}\,\gamma_j) = \alpha_i\alpha_{i'}\,\gamma_j\gamma_{j'},$$

and so the conditional moment condition

$$E[u_{ij}\,u_{i'j'} - u_{ij'}\,u_{i'j}|x_{11},\ldots,x_{nm}] = 0 \tag{1.8}$$

holds. This condition is the two-way counterpart to (1.4). The need to involve pairs of variables in both dimensions, that is, pairs (i,i') and pairs (j,j') arises from the presence of nuisance parameters in both dimensions. More precisely, an appropriate comparison of data pairs effectively differences-out the nuisance parameters in each dimension, and thus paves the way for the construction of GMM estimators.

To give an example, the two-way version of (1.2) would be

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{i<i'}\sum_{j<j'}\big\{(x_{ij}-x_{ij'})-(x_{i'j}-x_{i'j'})\big\}\big\{u_{ij}(\psi)\,u_{i'j'}(\psi)-u_{ij'}(\psi)\,u_{i'j}(\psi)\big\} = 0.$$

We will turn to the construction of more general GMM estimators and their statistical properties in the next section.

In independent work, Charbonneau (2013) obtained a particular unconditional version of the moment in (1.8) for the static exponential-regression model using a somewhat different argument, but did not pursue the idea further. Rather, she focused on constructing conditional maximum-likelihood estimators (Andersen 1970; Chamberlain 1980; Hausman, Hall and Griliches 1984) for Poisson, negbin1, and Gamma models. Presumably, the statistical properties of these estimators can be derived using similar methods as the ones developed in the next section. Of course, for each of these models, the outcome variable can be written as in (1.6), and so the conditional-mean parameters can be estimated without the need for full distributional assumptions by GMM estimators constructed from the moment condition in (1.8) above.

While in the one-way case it is known that the conditional maximum-likelihood estimator of ψ_0 in the Poisson model co-incides with the fixed-effect maximum likelihood estimator (see, e.g., Lancaster 2002), this is no longer the case in the two-way model.[3] Thus, these conditional maximum-likelihood estimators will generally not have a pseudo maximum-likelihood interpretation, and so they need not be consistent under misspecification. Indeed, Charbonneau (2013) shows that the score equation of the Poisson maximum-likelihood estimator in the two-way model is biased, in general.

This implies that, contrary to in the one-way case, the estimating equations of the maximum-likelihood estimator are not nested in the GMM framework. Fernández-Val and Weidner (2014) showed that, under sampling and regularity conditions, in a correctly-specified static Poisson model, the bias in the score is asymptotically-negligible under asymptotics where $n, m \to \infty$ at the same rate. However, this is not the case for other non-negative limited dependent-variable models, and need not apply to the corresponding pseudo maximum-likelihood estimator.

A related difficulty with the fixed-effect estimator in the two-way model is that its score equation is no longer available in closed form. This makes computating

[3] The same does not hold for negative-binomial models and other models, where maximum likelihood is subject to the incidental-parameter problem; see Dhaene and Jochmans (2011) for bias calculations and corrections for one-way models.

the point estimates complicated, as all of $\{\alpha_i\}$, $\{\gamma_j\}$, and ψ_0 need to be estimated jointly. First, a normalization on the effects is needed. Indeed, if $u_{ij} = \alpha_i \gamma_j \varepsilon_{ij}$, then $u_{ij} = \widetilde{\alpha}_i \widetilde{\gamma}_j \varepsilon_{ij}$ with $\widetilde{\alpha}_i = \alpha_i \, c$ and $\widetilde{\gamma}_j = \gamma_j \, c^{-1}$ is an equivalent representation. This ambiguity can be fixed by, for example, forcing $\prod_i \alpha_i = \prod_j \gamma_j$ or setting $\alpha_1 = 1$. Guimarães and Portugal (2009) propose a zig-zag estimation routine to update parameter estimates sequentially (a similar algorithm featured earlier in Heckman and MaCurdy 1980). Apart from the obvious computational burden, such an approach has the severe disadvantage that it is not guaranteed to converge to the maximum-likelihood estimator. Furthermore, it does not yield correct standard errors, as the Hessian matrix is not block diagonal. See Greene (2004) for a discussion on these issues.

D. Estimation of fixed effects

In some cases, one may wish to estimate the parameters $\{\alpha_i\}$ and $\{\gamma_j\}$. Although the GMM approach yields moment conditions for ψ_0 that do not involve $\{\alpha_i\}$ and $\{\gamma_j\}$, these effects can still be estimated. However, it should be recalled that they may not represent good estimates of the true effects, as they will be based on few observations.

The fact that the α_i and γ_j cannot be estimated precisely has implications for the estimation of functionals of the distribution of $\{\alpha_i, \gamma_j\}$, such as their means and correlation (Andrews et al. 2008; Fernández-Val and Lee 2013), and also for the performance of conditional-variance specification tests along the lines suggested by Manning and Mullahy (2001) and Santos Silva and Tenreyro (2006). Of course, given that our model places restrictions on the conditional mean of the outcome only, and that identification is not driven by parametric assumptions on the conditional variance or on the relation between the covariates and the fixed effects, one could argue that such tests serve little purpose to begin with in the current setting.

In nonlinear panel models the presence of fixed effects generally complicates the estimation of average marginal effects rather substantially for the same reason as just described above (Dhaene and Jochmans 2014; Fernández-Val and Weidner 2014). Our multiplicative model presents an interesting exception. The average marginal effect for continuous x_{ij}, for example, equals

$$\lim_{n,m\to\infty} \frac{1}{nm} \sum_i \sum_j \frac{\partial E[y_{ij}|\alpha_i, \gamma_j, x_{ij}]}{\partial x_{ij}} = \lim_{n,m\to\infty} \frac{1}{nm} \sum_i \sum_j x_{ij} \, \frac{\partial \varphi(x_{ij}; \psi_0)}{\partial x_{ij}} \, \alpha_i \, \gamma_j,$$

which clearly involves all fixed effects. Hence, a naive plug-in estimator of this quantity will perform poorly, even if ψ_0 would be known. However, given ψ_0, we observe the errors

$$u_{ij} = \frac{y_{ij}}{\varphi(x_{ij}; \psi_0)} = \alpha_i \gamma_j \, \varepsilon_{ij}.$$

This suggests estimating the average marginal effect by

$$\frac{1}{nm} \sum_i \sum_j x_{ij} \, \frac{\partial \varphi(x_{ij}; \psi_{nm})}{\partial x_{ij}} \, u_{ij}(\psi_{nm}).$$

Under conventional smoothness and dominance conditions, this estimator has standard asymptotic properties.

II. GMM estimation of the two-way model

A. Choice of moments and implementation

Let $\rho = n(n-1)m(m-1)/4$ denote the number of distinct quads in the data and write $w_{ij} = (y_{ij}, x_{ij})$. For a function $p(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)$ that is antisymmetric in both (i, i') and (j, j'), let

$$h(w_{ij}, w_{ij'}, w_{i'j}, w_{i'j'}; \psi) = p(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi) \big\{ u_{ij}(\psi) u_{i'j'}(\psi) - u_{i'j}(\psi) u_{ij'}(\psi) \big\}.$$

Then

$$s(\psi) = \rho^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{i<i'} \sum_{j<j'} h(w_{ij}, w_{ij'}, w_{i'j}, w_{i'j'}; \psi) \tag{2.1}$$

can serve as the basis for a GMM estimator of ψ_0 that minimizes the quadratic form

$$s(\psi)' W s(\psi),$$

for some chosen positive-definite weight matrix W.

Antisymmetry of p is without loss of generality. It is useful as it makes h, the kernel of $s(\psi)$, permutation invariant in both dimensions of the panel. Our simple example from the previous section, for example, had

$$p(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi) = (x_{ij} - x_{ij'}) - (x_{i'j} - x_{i'j'}). \tag{2.2}$$

This is just a symmetrized version of the unconditional moment condition using x_{ij} as an instrument, i.e., of $E[x_{ij}\{u_{ij}(\psi)u_{i'j'}(\psi) - u_{i'j}(\psi)u_{ij'}(\psi)\}] = 0$. More general choices for p can be useful for efficiency considerations, however, when turning conditional moments into unconditional ones (Chamberlain 1987). One practical and simple choice of unconditional moment conditions that generalizes (2.2) follows from arguments of Newey (1990) and Donald, Imbens and Newey (2003) and has

$$p(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi) = (r(x_{ij}) + r(x_{i'j'})) - (r(x_{i'j}) + r(x_{ij'})) \tag{2.3}$$

for a vector of approximating functions $x \mapsto r(x)$. In the simplest case, $r(x)$ would just be a power series.[4] Often, a relatively small set of moment conditions is used and, indeed, much information can already be contained in these. With power

[4] Under appropriate conditions and i.i.d. sampling, letting the dimension of $r(x)$ increase with the sample size implies that the optimally-weighted GMM estimator attains the semiparametric efficiency bound of Chamberlain (1987); see Donald, Imbens and Newey (2003).

series, this means that $r(x)$ will consist of low-order polynomials and interaction terms. The resulting estimator is a nonlinear version of a two-stage least-squares estimator where the first stage consists of a regression of the endogenous variables on low-order polynomials in the instruments (see Kelejian 1971), which is very popular in applied work.

In any case, the GMM estimator we are interested in here has the generic form

$$\psi_{nm} = \arg\min_{\psi} s(\psi)' W s(\psi), \tag{2.4}$$

where $s(\psi)$ is constructed as in (2.1). W may depend on the data. In particular, the optimally-weighted estimator is the two-step estimator that sets $W = V^{-1}$ where V is a full-rank matrix such that

$$V^{-1/2} \sqrt{nm}\, s(\psi_0) \xrightarrow{d} N(0, I)$$

as $n, m \to \infty$, where I denotes the identity matrix of suitable dimension. Rather than the two-step estimator we could also work with the continuously-updating GMM estimator of Hansen, Heaton and Yaron (1996), defined as a minimizer of

$$s(\psi)' V(\psi)^{-1} s(\psi),$$

in obvious notation.

B. Estimation and inference

Under a set of regularity conditions, ψ_{nm} is consistent and asymptotically normal. The asymptotic distribution depends on whether both $n, m \to \infty$ or only one of the indices diverges. Because the most relevant asymptotic scheme for our purposes is the joint divergence of n and m, that is, double asymptotics, we consider only that case here.

To perform statistical inference, let

$$\upsilon_{ij} = \frac{4}{(n-1)(m-1)} \sum_{i' \neq i} \sum_{j' \neq j} h(w_{ij}, w_{ij'}, w_{i'j}, w_{i'j'}; \psi_{nm}),$$

and construct

$$V = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \upsilon_{ij}\, \upsilon_{ij}', \quad \Upsilon = \rho^{-1} \sum_{i=1}^{n} \sum_{i<i'} \sum_{j=1}^{m} \sum_{j<j'} \frac{\partial h(w_{ij}, w_{ij'}, w_{i'j}, w_{i'j'}; \psi_{nm})}{\partial \psi}.$$

Then, under a set of standard regularity conditions we have that, as $n, m \to \infty$ so that $n/m \to c$ for some constant c,

$$\Omega^{-1/2} \sqrt{nm}\, (\psi_{nm} - \psi_0) \xrightarrow{d} N(0, I) \tag{2.5}$$

for $\Omega = (\Upsilon' W \Upsilon)^{-1} (\Upsilon' W V W \Upsilon)(\Upsilon' W \Upsilon)^{-1}$.

Furthermore, for the optimally-weighted criterion we have

$$nm\, s(\psi_{nm})\, V^{-1} s(\psi_{nm}) \xrightarrow{p} \chi^2_o \tag{2.6}$$

where o is the number of overidentifying restrictions and χ^2_o is the chi-squared distribution with o degrees of freedom. The result in (2.6) can be used to test overidentifying restrictions.

The assumptions used to validate (2.5) include the existence of moments of the data of suitable order and rule out dependence in the outcome variables conditional on the fixed effects and the covariates but do not require them to be identically distributed. The errors can be heteroskedastic but are taken to be serially uncorrelated (conditional on the fixed effects). This is a reasonable assumption in many instances. The convergence result in (2.5) can be extended to deal with correlated errors but the form of the asymptotic-variance matrix, Ω, will change as a result of this correlation. However note that, while we require the shocks ε_{ij} to be uncorrelated, the disturbances $u_{ij} = \alpha_i \gamma_j \varepsilon_{ij}$ are correlated at the (i, j) level due to the presence of the fixed effects, and that no clustering is necessary to obtain valid standard errors for ψ_{nm}.

The distribution theory readily extends to the case where x_{ij} is endogenous but one has instruments z_{ij}, requiring only to replace $p(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)$ by $p(z_{ij}, z_{ij'}, z_{i'j}, z_{i'j'}; \psi)$.

C. *Numerical illustrations*

A prime example of the models covered by the approach in this paper is the exponential-regression model with fixed effects. Hence, it is useful to consider the performance of the GMM estimator by means of a Monte Carlo experiment set up around the specification

$$y_{ij} = \exp(x_{ij}\psi_0)\, \alpha_i\, \gamma_j\, \varepsilon_{ij}. \tag{2.7}$$

Throughout we fix $\psi_0 = 1$. We report results for panels of dimension $n = m = 50$. Different sample sizes gave similar results.

In each Monte Carlo replication we generated

$$x_{ij} \sim N(0,1), \qquad \alpha_i \sim \log N(0,1), \qquad \gamma_j \sim \log N(0,1).$$

We then drew

$$\varepsilon_{ij} \sim \log N\left(-\log\sqrt{1+\sigma_{ij}^2},\, \log\left(1+\sigma_{ij}^2\right)\right)$$

for five different choices of σ_{ij}. Similar to Santos Silva and Tenreyro (2006), the designs considered are

$$\begin{aligned}
&\text{Design 1:} \qquad && \sigma_{ij}^2 = 1,\\
&\text{Design 2:} \qquad && \sigma_{ij}^2 = \left(\exp(x_{ij}\psi_0)\, \alpha_i\, \gamma_j\right)^{-1},\\
&\text{Design 3:} \qquad && \sigma_{ij}^2 = \left(\exp(x_{ij}\psi_0)\, \alpha_i\, \gamma_j\right),\\
&\text{Design 4:} \qquad && \sigma_{ij}^2 = \left(\exp(x_{ij}\psi_0)\, \alpha_i\, \gamma_j\right)^{-2},\\
&\text{Design 5:} \qquad && \sigma_{ij}^2 = \left(\exp(x_{ij}\psi_0)\, \alpha_i\, \gamma_j\right)^{2}.
\end{aligned}$$

Then $\mathrm{var}(\varepsilon_{ij}) = \sigma_{ij}^2$ and $E[\varepsilon_{ij}] = 1$. Moreover, ε_{ij} is mean-independent of x_{ij}, but conditionally heteroskedastic in all but the first design.

We estimated ψ_0 using the GMM estimator developed here with $r(x) = x$ that is, with (2.1) combined with (2.2). This is the simplest possible choice. We also report results for the fixed-effect Poisson pseudo maximum-likelihood estimator (PMLE) and the fixed-effect least-squares estimator (OLS) of the log-linearized equation

$$\log y_{ij} = x_{ij}\psi_0 + \log \alpha_i + \log \gamma_j + \log \varepsilon_{ij}.$$

The latter two are included because they are popular in the trade literature. Pseudo maximum-likelihood estimation based on negative-binomial and other models has also been discussed in this literature (see, e.g., Head and Mayer 2014). We do not consider such estimators here because, in models with fixed effects, most of these estimators are subject to incidental-parameter bias (Dhaene and Jochmans 2011). Hence, they are not appropriate to estimate gravity equations from panel data.

Table 1 gives the mean and standard deviation for these three estimators, as obtained over $1,000$ Monte Carlo replications. For GMM, we also provide the empirical coverage rate of 95% confidence intervals (ci).

The table replicates the usual finding that log-linearization induces endogeneity bias when errors are not independent of covariates. Indeed, except for Design 1, OLS is inconsistent and heavily biased. In contrast for all designs, GMM has small bias. Furthermore, its confidence intervals display good coverage. Interestingly, joint maximization of the Poisson pseudo log-likelihood over ψ and all fixed effects yields estimates with finite-sample behavior close to that of GMM. This suggests that the favorable robustness properties of this estimator in the one-way model (Wooldridge 1999) may carry over to the two-way case. Theoretical results on this would be a nice addition to the literature.

Table 1—Simulation results for an exponential-regression model

	mean			sd			ci
Design	OLS	PMLE	GMM	OLS	PMLE	GMM	GMM
1	1.001	1.000	1.003	.017	.061	.043	.962
2	1.249	1.000	1.001	.028	.011	.021	.951
3	.749	.948	.974	.028	.190	.136	.879
4	1.501	1.000	1.002	.053	.008	.028	.912
5	.499	.862	.903	.028	.167	.094	.832

We also experimented with an instrumental-variable specification of (2.7) where

$$\binom{\log \varepsilon_{ij}}{\eta_{ij}} \sim N\left(\begin{pmatrix} -\log\sqrt{2} \\ 0 \end{pmatrix} \begin{pmatrix} \log 2 & \varrho\sqrt{\log 2} \\ \varrho\sqrt{\log 2} & 1 \end{pmatrix} \right),$$

for $|\varrho| < 1$, and the covariate is generated as

$$x_{ij} = z_{ij} + \eta_{ij}, \qquad z_{ij} \sim N\big(0, (1-\varrho^2)\,\log 2\big).$$

Here, the errors satisfy $E[\varepsilon_{ij}] = 1$ and $\mathrm{var}(\varepsilon_{ij}) = 1$, and $\eta_{ij} \sim N(0,1)$, but x_{ij} is endogenous unless $\varrho = 0$. Indeed,

$$x_{ij} = \sqrt{(1-\varrho^2)\,\log 2}\,\epsilon^1_{ij} + \sqrt{(1-\varrho^2)\,\log 2}\,\epsilon^2_{ij} + \sqrt{\varrho^2 \log 2}\,(\log \varepsilon_{ij} + \log\sqrt{2})$$

for independent standard-normal variates ϵ^1_{ij} and ϵ^2_{ij}. The first right-hand side term is the impact of the instrument z_{ij} on x_{ij}. The second right-hand side term is a random shock, which contributed to x_{ij} in the same way as does z_{ij}. The third right-hand side term, finally, is the impact of ε_{ij} on x_{ij}. The effect of this term is linear in ϱ. We consider five designs that differ only in the value that ϱ takes. The remainder of the data generating process is as before, and so is the sample size.

Table 2, which has the same layout as Table 1, contains the results. Besides results for the GMM estimator, results for least squares and pseudo maximum

TABLE 2—SIMULATION RESULTS FOR AN EXPONENTIAL INSTRUMENTAL-VARIABLE MODEL

	mean			sd			ci
ϱ	OLS	PMLE	GMM	OLS	PMLE	GMM	GMM
0	1.000	.997	1.001	.015	.058	.053	.957
−.50	.714	.712	1.001	.015	.047	.063	.950
−.25	.871	.869	1.002	.014	.052	.055	.959
.25	1.129	1.125	.997	.014	.056	.054	.964
.50	1.285	1.282	.999	.014	.062	.061	.967

likelihood are also given. The latter two are, however, not designed to handle endogeneity. They are reported as a means to appreciate the degreee of bias endogeneity induces in our data generating process. The table shows this bias is substantial, ranging up to nearly 30%. GMM has negligible bias (relative to its standard deviation) for all designs considered. Furthermore, the associated confidence intervals enjoy excellent coverage.

III. Gravity estimates

As an empirical application we estimated the Anderson and van Wincoop (2003) gravity model. This model is essentially an exponential-regression model as in (2.7) that links trade flows from country i to country j to a linear index of distance measures, $x'_{ij}\psi$, and multilateral resistance terms or exporter and importer effects, α_i and γ_j, respectively.

TABLE 3—TRADE DATA: DESCRIPTIVE STATISTICS

	full sample		positive-trade sample	
	mean	std	mean	std
trade decision	0.5236	0.4995	—	—
trade volume	172130	1829058	328752	2517607
log distance	8.7855	0.7418	8.6950	0.7728
common border	0.0196	0.1387	0.0236	0.1519
common language	0.2097	0.4071	0.2128	0.4093
colonial past	0.1705	0.3761	0.1689	0.3747
free trade agreement	0.0155	0.1234	0.0262	0.1598

We use the data of Santos Silva and Tenreyro (2006). These data contain information on 136 countries, giving a total of $136 \times 135 = 18,360$ directed trade flows. About 52% of these bilateral flows are positive. As outcome variable we use bilateral trade, measured in $1,000$ dollars. As distance measures we use an actual geographical distance together with a set of dummies that aim to capture other factors of relatedness. Moreover, we include dummies that indicate whether or not countries i and j share a border, speak the same language, have a colonial past, and fall under a joint free-trade agreement. Table 3 contains summary statistics for all variables. A more detailed presentation of the data is given in Santos Silva and Tenreyro (2006).

A peculiar feature of trade data is that all regressors take on only non-negative values. With $\varphi(x;\psi) = \exp(x'\psi)$, this means that

$$\|s(\psi)\| \to 0 \quad \text{as} \quad \|\psi\| \to +\infty$$

if $s(\psi)$ is set as in (2.1). At the same time, the Jacobian matrix will converge to zero. This makes (2.1) unattractive to work with for this type of data. To circumvent this issue we work with

$$\rho^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{\substack{i<i' \\ i' \neq j}} \sum_{\substack{j<j' \\ j' \neq i,i'}} h(w_{ij}, w_{ij'}, w_{i'j}, w_{i'j'};\psi), \tag{3.1}$$

where we set $h(w_{ij}, w_{ij'}, w_{i'j}, w_{i'j'};\psi)$ equal to

$$\{(x_{ij}+x_{i'j'})-(x_{i'j}+x_{ij'})\}\{y_{ij}y_{i'j'}\,\varphi(x_{i'j};\psi)\varphi(x_{ij'};\psi)-y_{i'j}y_{ij'}\,\varphi(x_{ij};\psi)\varphi(x_{i'j'};\psi)\}.$$

This amounts to multiplying (2.1) with $\varphi(x_{ij};\psi)\varphi(x_{ij'};\psi)\varphi(x_{i'j};\psi)\varphi(x_{i'j'};\psi)$.[5] We also adjust the summations in (3.1) relative to (2.1) to take into account that

[5]One may note that $s(\psi)$ in (3.1) converges to zero as $\|\psi\| \to -\infty$. However, the Jacobian matrix, $H(\psi) = \partial s(\psi)/\partial\psi'$, diverges to $+\infty$ at the same rate in this case. Hence, the first-order condition of the GMM minimization problem, which is $H(\psi)'Ws(\psi) = 0$, remains well behaved.

each country shows up in the data as both exporter and importer, and because countries do not trade with themselves.

We report estimates based on GMM and on pseudo maximum likelihood. The latter were taken from Santos Silva and Tenreyro (2006). Within-group estimates of the log-linearized model or (fixed-effect) nonlinear least-squares estimates of the model in levels are not reported here (see Santos Silva and Tenreyro 2006, Table 5). Within-groups requires an assumption of full independence on the random disturbances while its nonlinear version is asymptotically biased in a panel setting with importer/exporter effects.

Table 4—Trade flows: gravity estimates

outcome variable: trade flows (in $1,000$ US dollars)				
	all flows		positive flows	
	GMM	PMLE	GMM	PMLE
log distance	$-.781$	$-.750$	$-.799$	$-.770$
	(.096)	(.041)	(.103)	(.042)
common border	.125	.370	.109	.352
	(.095)	(.091)	(.099)	(.090)
common language	.521	.383	.534	.418
	(.089)	(.093)	(.087)	(.094)
colonial past	.206	.079	.190	.038
	(.168)	(.134)	(.176)	(.134)
free trade agreement	.253	.376	.256	.374
	(.153)	(.077)	(.155)	(.076)

The first columns (“all flows”) in Table 4 provides the estimation results based on the full sample (including zero trade flows), together with standard errors in brackets. The sign of all estimated coefficients is in accordance with expectations. An increase in geographical distance has a negative ceteris paribus impact on the size of expected trade flows. The magnitude of this effect is similar to that found by Santos Silva and Tenreyro (2006), and is strongly statistically significant. Sharing a border or spoken language, having a colonial past, or being in a free trade agreement all tend to increase the magnitude of trade flows. As compared to Santos Silva and Tenreyro (2006), our estimates give less economic importance to the impact on trade of having a border in common (conditional on the other regressors, and geographical distance in particular). With a p-value of .1895, the common-border effect is further also statically not distinguishable from zero at all conventional significance levels. In contrast, speaking the same language has a large and strongly significant impact on trade. Colonial heritage seems to contribute very little in determining current trade flows, and we find no statistical evidence for it in the data used here. Finally, we find a positive impact of free trade agreements on trade intensity. This effect is statically significant only at the 10% level.

A considerable part of country pairs in the data do not trade. Zero trade flows are difficult to explain within the traditional gravity framework. While exponential-regression models have no technical difficulty in dealing with zeros, the discrepancy between stylized facts and theory suggests the gravity model to be incorrectly specified. An alternative and very simple explanation for zero trade would be a measurement issue. That is, small but non-zero trade flows show up as exact zeros in the data. However, even if so, the assumption that such trade flows are missing at random is very difficult to support. For completeness, the second set of columns ("positive flows") in Table 4 contains the estimation results based on the subsample of positive trade flows only, thus excluding all zero trade flows. The results change very little.

Work by Chaney (2008) and Helpman, Melitz and Rubinstein (2008) models trading as a two-stage process, where the first stage is a decision of whether or not to initiate trade and the second-stage is a choice of trade intensity. With importer and exporter fixed effects, such a system of equations is difficult to estimate. Indeed, even in a fully parametric setting (as in Helpman, Melitz and Rubinstein 2008), maximum likelihood suffers from large bias. An interesting agenda for future research would be to derive estimators with good theoretical properties for such extended gravity models, building on the theoretical results presented here.

Concluding remarks

We have introduced a differencing strategy to construct GMM estimators for a class of nonlinear panel models with two-way fixed effects. The approach can be seen as an intuitive extension to the well-known first-differencing strategy that is conventional in standard (one-way) fixed-effect models. More generally, it can be extended to multiway models. For example, if agents are observed over multiple time periods or in different markets $k = 1, \ldots, l$ we have three-dimensional data. The model

$$y_{ijk} = \varphi(x_{ijk}; \psi_0)\, u_{ijk}, \qquad u_{ijk} = \alpha_i \gamma_j \delta_k\, \varepsilon_{ijk},$$

in obvious notation, allows for heterogeneity in each of the three dimensions. A mean-independence assumption implies that

$$E[(u_{ijk}\, u_{i'j'k'}\, u_{i''j''k''}) - (u_{ij'k''}\, u_{i'j''k}\, u_{i''jk'})|x_{111}, \ldots, x_{nml}] = 0.$$

GMM estimators follow in the same way as before. Their asymptotic behavior can be deduced from (2.5).

References

Aaronson, D. L., L. Barrow, and W. Sander. 2007. “Teacher and student achievement in the Chicago public high schools.” *Journal of Labor Economics*, 25: 95–135.

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, Series B*, 32: 283–301.

Anderson, J. E. 2011. "The gravity model." *Annual Review of Economics*, 3: 133–160.

Anderson, J. E., and E. van Wincoop. 2003. "Gravity with gravitas: A solution to the border puzzle." *American Economic Review*, 93: 170–192.

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 2008. "High wage workers and low wage firms: Negative assortative matching or limited mobility bias." *Journal of the Royal Statistical Society, Series A*, 171: 673–697.

Chamberlain, G. 1980. "Analysis of covariance with qualitative data." *Review of Economic Studies*, 47: 225–238.

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

Chamberlain, G. 1992. "Comment: Sequential moment restrictions in panel data." *Journal of Business & Economic Statistics*, 10: 20–26.

Chaney, T. 2008. "Distorted gravity: The intensive and extensive margins of international trade." *American Economic Review*, 98: 1707–1721.

Charbonneau, K. B. 2013. "Multiple fixed effects in theoretical and applied econometrics." PhD diss. Princeton University.

Dhaene, G., and K. Jochmans. 2011. "Profile-score adjustments for nonlinear fixed-effect models." Mimeo.

Dhaene, G., and K. Jochmans. 2014. "Split-panel jackknife estimation of fixed-effect models." *Review of Economic Studies*, forthcoming.

Donald, S. G., G. W. Imbens, and W. K. Newey. 2003. "Empirical likelihood estimation and consistent tests with conditional moment restrictions." *Journal of Econometrics*, 117: 55–93.

Fernández-Val, I., and J. Lee. 2013. "Panel data models with nonadditive unobserved heterogeneity: Estimation and inference." *Quantitative Economics*, 4: 453–481.

Fernández-Val, I., and M. Weidner. 2014. "Individual and time effects in nonlinear panel data models with large N, T." *CeMMAP Working Paper CWP33/14*.

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

Graham, B. G. 2013. "Comment." *Journal of Business & Economic Statistics*, 31: 266–270.

Graham, B. S. 2014. "An empirical model of network formation: Detecting homophily when agents are heterogeneous." Mimeo.

Greene, W. H. 2004. "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects." *Econometrics Journal*, 7: 98–119.

Guimarães, P., and P. Portugal. 2009. "A simple feasible alternative procedure to estimate models with high-dimensional fixed effects." *IZA Discussion Paper Series No. 3935*.

Hahn, J., and W. K. Newey. 2004. "Jackknife and analytical bias reduction for nonlinear panel models." *Econometrica*, 72: 1295–1319.

Hansen, L. P., J. Heaton, and A. Yaron. 1996. "Finite-sample properties of some alternative GMM estimators." *Journal of Business and Economic Statistics*, 14: 262–280.

Harrigan, J. 1996. "Openness to trade in manufactures in the OECD." *Journal of International Economics*, 40: 23–39.

Hausman, J. A., B. H. Hall, and Z. Griliches. 1984. "Econometric models for count data with an application to the patents-R&D relationship." *Econometrica*, 52: 909–938.

Head, K., and T. Mayer. 2014. "Gravity equations: Workhorse, toolkit, and cookbook." In *Handbook of International Economics*. Vol. 4, , ed. G. Gopinath, E. Helpman and K. Rogoff, Chapter 3, 131–195. Elsevier.

Heckman, J. J. 1979. "Sample selection bias as a specification error." *Econometrica*, 47: 153–161.

Heckman, J. J., and T. E. MaCurdy. 1980. "A life cycle model of female labor supply." *Review of Economic Studies*, 47: 47–74.

Helpman, E., M. Melitz, and Y. Rubinstein. 2008. "Estimating trade flows: Trading partners and trading volumes." *Quarterly Journal of Economics*, 123: 441–487.

Jochmans, K. 2015. "Multiplicative-error models with sample selection." *Journal of Econometrics*, 184: 315–327.

Kelejian, H. H. 1971. "Two-stage least squares and econometric systems linear in parameters but nonlinear in endogenous variables." *Journal of the American Statistical Association*, 66: 373–374.

Lancaster, T. 2002. "Orthogonal parameters and panel data." *Review of Economic Studies*, 69: 647–666.

Manning, W. G., and J. Mullahy. 2001. "Estimating log models: To transform or not to transform?" *Journal of Health Economics*, 20: 461–494.

Newey, W. K. 1990. "Efficient instrumental variables estimation of nonlinear models." *Econometrica*, 58: 809–837.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica*, 73: 417–458.

Santos Silva, J. M. C., and S. Tenreyro. 2006. "The log of gravity." *Review of Economics and Statistics*, 88: 641–658.

Tinbergen, J. 1962. *The World Economy: Suggestions for an International Economic Policy.* Twentieth Century Fund.

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

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