TWO-WAY MODELS FOR GRAVITY

Koen Jochmans*

Draft: February 2, 2015 Revised: November 18, 2015
This version: April 19, 2016

Empirical models for dyadic interactions between n agents often feature agent-specific parameters. Fixed-effect estimators of such models generally have bias of order n^{-1}, which is non-negligible relative to their standard error. Therefore, confidence sets based on the asymptotic distribution have incorrect coverage. This paper looks at models with multiplicative unobservables and fixed effects. We derive moment conditions that are free of fixed effects and use them to set up estimators that are n-consistent, asymptotically normally-distributed, and asymptotically unbiased. We provide Monte Carlo evidence for a range of models. We estimate a gravity equation as an empirical illustration.
JEL Classification: C14, C23, F14

Empirical models for dyadic interactions between n agents frequently contain agent-specific fixed effects. The inclusion of such effects captures unobserved characteristics that are heterogeneous across agents. One leading example is a gravity equation for bilateral trade flows between countries; they feature both importer and exporter fixed effects at least since the work of Harrigan (1996) and Anderson and van Wincoop (2003). While such two-way models are intuitively

* Address: Sciences Po, Département d'économie, 28 rue des Saints Pères, 75007 Paris, France; koen.jochmans@sciencespo.fr.
Acknowledgements: I am grateful to the editor and to three referees for comments and suggestions, and to Maurice Bun, Thierry Mayer, Martin Weidner, and Frank Windmeijer for fruitful discussion. This work was partially supported by the SAB grant 'Econometric analysis of linked data'. Part of the work on this paper was done while I was visiting the London School of Economics, whose hospitality is gratefully acknowledged.

attractive and their use is widespread, there is little to no theoretical work on the statistical properties of the corresponding estimators.

This paper considers estimation and inference for nonlinear two-way models with multiplicative unobservables and fixed effects. Such models are well suited for studying non-negative outcomes in a variety of contexts. Count data and duration data are two obvious and important examples. Other examples are constant-elasticity models, life-cycle models for consumption, and binary-choice models with multiplicative effects. Our approach is semiparametric in that it requires a conditional moment restriction only and is sufficiently general to cover instrumental-variable models although, for conciseness, we do not cover the latter in detail here. Building on an insight of Charbonneau (2013), we derive moment conditions that difference-out the fixed effects. Under regularity conditions the associated generalized method-of-moment (GMM) estimators are consistent and converge at the rate n^{-1} to a normal random variable whose variance can be estimated. Extensive numerical experiments show that our asymptotic theory provides a good approximation to the small-sample behavior of the estimators. Furthermore, in experiments with exponential-regression models, they are found to provide more reliable inference than the Poisson pseudo maximum likelihood (Gouriéroux, Monfort and Trognon, 1984*a*). As an empirical application we estimate a gravity equation in levels (as advocated by Santos Silva and Tenreyro 2006), controlling for multilateral resistance terms.

There is related work by Fernández-Val and Weidner (2016) on likelihood-based estimation of two-way models. They show that (under regularity conditions) the bias of the fixed-effect estimator of two-way models, in general, is $O(n^{-1})$ and needs to be corrected for in order to perform asymptotically-valid inference. Our approach is different as we work with moment conditions that are free of fixed effects, implying the associated estimators to be asymptotically unbiased. Also, the class of models considered by Fernández-Val and Weidner (2016) and the one under study here are different, and they are not nested.[1] In the likelihood setting, a possible alternative may be to work with a conditional likelihood. Charbonneau (2013) investigates this possibility for several models for count data.

I. Multiplicative models for dyadic data

We have data on dyadic interactions between n agents. Let (y_{ij}, x_{ij}) denote the observation on dyad (i, j). We allow for directed interactions, so that (y_{ij}, x_{ij}) need not be equal to (y_{ji}, x_{ji}), and include self links, that is, (y_{ii}, x_{ii}).[2] Suppose that

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

where φ is a function known up to the parameter vector ψ_0, and u_{ij} is a latent disturbance. We will assume that

$$u_{ij} = \alpha_i\, \gamma_j\, \varepsilon_{ij}, \tag{1.2}$$

where α_i and γ_j represent permanent unobserved effects and ε_{ij} is an idiosyncratic disturbance that is independent across both i and j. Independence will only be used to establish asymptotic normality and can be relaxed, as discussed in more detail below. Note that, besides controlling for unobserved heterogeneity, this two-way model gives a simple framework to deal with aggregate shocks. Moreover, the presence of α_i and γ_j implies that u_{ij} is heteroskedastic and correlated across both i and j. We will treat α_i and γ_j as fixed, that is, throughout, we condition on them.[3]

[1] Our results are applicable to $n \times m$ panel data under asymptotics where $n, m \to \infty$ jointly; see a previous version of this paper. This can be useful for modelling linked data between two different types of agents, such as firms and workers or teachers and students. The formulae to follow require only minor and obvious modification, and the sampling scheme in Assumption 3 needs to be redefined appropriately.

[2] In the absence of self links it suffices to alter all expressions below by adjusting the range of the sums and by rescaling appropriately to obtain a degrees-of-freedom correction.

[3] We omit the qualifier 'almost surely' from all probabilistic statements.

Our aim is to estimate the parameter ψ_0 under the conditional-mean restriction

$$E[\varepsilon_{ij}|x_{11},\ldots,x_{nn}] = 1. \tag{1.3}$$

Everything that follows extends to the setting where $E[\varepsilon_{ij}|z_{11},\ldots,z_{nn}] = 1$ for instrumental variables $z_{11},\ldots,z_{nn}$, with obvious modification to the formulae and subject to suitably adjusted regularity conditions. For conciseness, we maintain (1.3) here.[4]

To construct an estimator of ψ_0 that will have good statistical properties as $n \to \infty$ we construct moment conditions that are free of fixed effects. This can be done by extending a recent finding due to Charbonneau (2013) for the exponential-regression model to the more general framework entertained here. We do so by following the intuition underlying the work of Chamberlain (1992) and Wooldridge (1997) for one-way models. First observe that (1.3) implies that

$$E\left[u_{ij}\right| x_{11},\ldots,x_{nn}] = \alpha_i\,\gamma_j$$

for any i,j. Furthermore, as $E[\varepsilon_{ij}\varepsilon_{i'j'}|x_{11},\ldots,x_{nn}] = 1$ for different pairs of indices i,j and i',j',

$$\begin{aligned} E\left[u_{ij}\,u_{i'j'}\right| x_{11},\ldots,x_{nn}\big] &= (\alpha_i\,\gamma_j)\,(\alpha_{i'}\,\gamma_{j'}) = \alpha_i\alpha_{i'}\,\gamma_j\gamma_{j'}, \\ E\left[u_{ij'}\,u_{i'j}\right| x_{11},\ldots,x_{nn}\big] &= (\alpha_i\,\gamma_{j'})\,(\alpha_{i'}\,\gamma_j) = \alpha_i\alpha_{i'}\,\gamma_j\gamma_{j'}. \end{aligned}$$

By differencing these equations we then obtain the conditional moment condition

$$E[u_{ij}\,u_{i'j'} - u_{ij'}\,u_{i'j}|x_{11},\ldots,x_{nn}] = 0, \tag{1.4}$$

[4]A previous version of this paper contains simulation results for an instrumental-variable model.

which does not involve any of the nuisance parameters, and holds for all

$$\varrho = \binom{n}{2}\binom{n}{2} = \left(\frac{n!}{2!\,(n-2)!}\right)^2 = \frac{n^2(n-1)^2}{4}$$

unique choices for (i,i') and (j,j'). Equation (1.4) is the two-way counterpart to Chamberlain (1992) and Wooldridge (1997). It effectively differences-out each of the fixed effects. As such, the conditional moment condition in (1.4) paves the way for the construction of GMM estimators of ψ_0 set up from unconditional moments conditions implied by it. Such estimators are the topic of the next section.

An issue that we do not address here is semiparametrically-efficient estimation. The classic results of Chamberlain (1987) do not apply to the current framework. Furthermore, calculations of the moment conditions implied by the formulae in Chamberlain (1987) for some parametric specifications of (1.1)–(1.3) for 2×2 data, such as the Poisson model and negative-binomial model, reveal that these moments depend on the fixed effects. See the Supplementary Material for detailed calculations.

II. Estimation

Equation (1.4) implies that the unconditional moment condition

$$E[\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)(u_{ij}\, u_{i'j'} - u_{ij'}\, u_{i'j})] = 0, \tag{2.1}$$

where ϕ is a chosen (vector) function, holds for all ϱ choices of i, i', j, j'. An intuitive way of obtaining an estimating equation for ψ_0 then is to work with the empirical counterpart of the average of (2.1) over all ϱ choices. By letting $u_{ij}(\psi) = y_{ij}/\varphi(x_{ij};\psi)$, this empirical moment at a given value ψ is the U-statistic

$$s(\psi) = \varrho^{-1} \sum_{i=1}^{n} \sum_{i<i'} \sum_{j=1}^{n} \sum_{j<j'} \phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)(u_{ij}(\psi)\, u_{i'j'}(\psi) - u_{ij'}(\psi)\, u_{i'j}(\psi))$$

where, without loss of generality, we have assumed that the kernel function, $\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)(u_{ij}(\psi)\, u_{i'j'}(\psi) - u_{ij'}(\psi)\, u_{i'j}(\psi))$, is permutation invariant in both (i, i') and (j, j'). A GMM estimator of ψ_0 is

$$\psi_n = \arg\min_{\psi \in \mathcal{S}} s(\psi)'\, \Omega_n\, s(\psi),$$

where $\mathcal{S}$ is the parameter space searched over and Ω_n is a chosen positive-definite weight matrix. As usual for GMM estimators, Ω_n defines a distance metric for the moment conditions in case of overidentification, that is, when the dimension of ϕ exceeds the dimension of ψ.

We now provide distribution theory for this estimator. All proofs are collected in the Supplementary Material to this paper.

We start by imposing standard regularity conditions.

Assumption 1. *The set $\mathcal{S}$ is compact and ψ_0 is interior to it. The functions φ and ϕ are continuously-differentiable in ψ with derivatives φ' and ϕ'. There exists a positive definite matrix Ω such that $\Omega_n \xrightarrow{p} \Omega$ as $n \to \infty$.*

The next assumption relates to identification of ψ_0. We introduce the matrix

$$\Sigma = -\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[w_{ij}\, \tau_{ij}(x_{ij}; \psi_0)'],$$

where we define the random variable w_{ij} as

$$w_{ij} = \frac{4}{(n-1)^2} \sum_{i' \neq i} \sum_{j' \neq j} \phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)\, \alpha_i \alpha_{i'}\, \gamma_j \gamma_{j'}$$

and let $\tau(x_{ij}; \psi) = \varphi'(x_{ij}; \psi)/\varphi(x_{ij}; \psi)$.

Assumption 2. *With $\overline{s}(\psi) = \lim_{n\to\infty} s(\psi)$, $\|\overline{s}(\psi_k)\| \to 0$ implies $\|\psi_k - \psi_0\| \to 0$ for any sequence of vectors $\{\psi_k\}$ from $\mathcal{S}$. The matrix Σ has maximal column rank.*

Sampling is governed by the next assumption.

Assumption 3. *The n observations are sampled independently.*

Assumption 3 allows for dependence between dyads that have observations in common, which is important in applications.

The next assumption collects moment conditions that allow the application of a law of large numbers. We let $\sigma_{ij}^2 = \mathrm{var}(\varepsilon_{ij}|x_{11}, \ldots, x_{nn})$.

Assumption 4. *There exist finite constants C_u and C_ϕ, independent of ψ, such that $E[\|u_{ij}(\psi)\|^8] < C_u$ and $E[\|\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)\|^8] < C_\phi$ for all ψ in $\mathcal{S}$, and the constants α_i, γ_i are finite for all i. There exists a finite constant C_σ such that $E[\varepsilon_{ij}^4|x_{11}, \ldots, x_{nn}] < C_\sigma$, and the conditional variance σ_{ij}^2 is positive and has finite fourth-order moment.*

Assumptions 1–4 allow us to derive a consistency result for ψ_n.

Theorem 1 (Consistency)**.** *If Assumptions 1–4 hold, $\psi_n \xrightarrow{p} \psi_0$ as $n \to \infty$.*

To see why the dependence between dyads that have observations in common is not a hinder for consistency, note that $\mathrm{var}(s(\psi))$ is an average over $O(n^8)$ combinations of observations. Of these, $O(n^7)$ have at least one observation in common. Therefore, $\mathrm{var}(s(\psi)) = O(n^{-1})$, from which the convergence result follows. We note that Theorem 1 continues to go through when the disturbances ε_{ij} are dependent across i or j (or both).

Moving on to deriving the convergence rate and asymptotic distribution requires establishing the large-sample behavior of the empirical moment conditions. This is not immediate because the data are not identically distributed and can be strongly correlated across both i and j. We exploit the U-statistic structure of $s(\psi)$ to show that

$$n\, s(\psi_0) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\, (\varepsilon_{ij} - 1) + o_p(1). \tag{2.2}$$

The dominant right-hand side term is a Hájek projection (van der Vaart 2000, Section 11.3). The summands in (2.2) are all zero-mean random variables that are independent conditional on $x_{11}, \dots, x_{nn}$. Equation (2.2) states that $s(\psi_0)$ is asymptotically equivalent to its Hájek projection. Thus, $\operatorname{var}(s(\psi_0)) = O(n^{-2})$, and so we get $n \|\psi_n - \psi_0\| = O_p(1)$. Moreover, a suitable central limit theorem allows to establish that

$$n\, s(\psi_0) \xrightarrow{d} N(0, V), \qquad V = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[w_{ij} w_{ij}' \, \sigma_{ij}^2],$$

as $n \to \infty$.

The last ingredient needed for asymptotic normality is a convergence result for $S(\psi) = \partial s(\psi)/\partial \psi'$, the Jacobian of the empirical moment conditions. The next assumption collects sufficient additional conditions to ensure that $S(\psi_n) \xrightarrow{p} \Sigma$ as $n \to \infty$.

Assumption 5. *There exist finite constants C_u and C_ϕ, independent of ψ, such that $E[\|\tau(x_{ij};\psi)\|^8] < C_\tau$ and $E[\|\phi'(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'};\psi)\|^8] < C_{\phi'}$ for all ψ in $\mathcal{S}$.*

An expansion of the first-order conditions of the GMM estimation problem around ψ_0 then yields the following result.

Theorem 2 (Asymptotic normality)**.** *If Assumptions 1–5 hold and V is positive definite, then*

$$n\,(\psi_n - \psi_0) \xrightarrow{d} N(0, \Upsilon)$$

as $n \to \infty$, where the covariance matrix is $\Upsilon = (\Sigma'\Omega\Sigma)^{-1}(\Sigma'\Omega V \Omega\Sigma)(\Sigma'\Omega\Sigma)^{-1}$.

As usual, the asymptotic variance is minimized by setting $\Omega_n = V_n^{-1}$ where V_n is a consistent estimator of V.

The asymptotic variance Υ can be estimated by

$$\Upsilon_n = (S_n'\Omega_n S_n)^{-1}(S_n'\Omega_n V_n \Omega_n S_n)(S_n'\Omega_n S_n)^{-1},$$

where $S_n = S(\psi_n)$ is the Jacobian of the empirical moment conditions evaluated at the point estimator and

$$V_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{\upsilon}_{ij} \hat{\upsilon}_{ij}'$$

for

$$\hat{\upsilon}_{ij} = \frac{4}{(n-1)^2} \sum_{i' \neq i} \sum_{j' \neq j} \phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_n)(\hat{u}_{ij}\, \hat{u}_{i'j'} - \hat{u}_{ij'}\, \hat{u}_{i'j})$$

with $\hat{u}_{ij} = u_{ij}(\psi_n)$. The moment conditions in Assumptions 4–5 imply that $\|\Upsilon_n - \Upsilon\| = o_p(1)$ as $n \to \infty$, operationalizing our estimator as a tool for statistical inference.

An interesting extension of Theorem 2 would be to allow for the errors ε_{ij} to be dependent at the (i, j) level. If left unrestricted, this additional dependence would slow down the convergence rate of ψ_n from n^{-1} to $n^{-1/2}$ (see Hansen 2007 for a discussion on this in the linear model) and would lead to a more complicated expression for the variance of the moment conditions V (see Cameron, Gelbach and Miller 2011). We leave a detailed analysis of two-way clustering in the current context for future research.

III. Numerical experiments

We consider the performance of our estimator in a series of simulation experiments centered around exponential-regression models. For such models, the Poisson pseudo maximum-likelihood estimator can serve as a useful benchmark. We write

$$\mu_{ij} = e^{x_{ij}' \psi_0}\, \alpha_i\, \gamma_j.$$

We consider data generating processes for count data, continuous outcomes, and mixed continuous/discrete outcomes.

To simulate count data we use the Poisson model and the negative-binomial (negbin) model. In the former model, the conditional mean and variance both equal the arrival rate, μ_{ij}. The negative-binomial model is a mixture model over

Poisson models, where the arrival rate has a Gamma distribution with positive shape and scale parameters θ and $p_{ij} = (1 + \mu_{ij}/\theta)^{-1}$, respectively. In this case $\text{var}(y_{ij}|x_{ij}) = \mu_{ij} + \theta\mu_{ij}^2$, and the variance exceeds the mean. By setting $\theta \in \{1, 5, 10\}$ we will look at data generating processes with varying degree of overdispersion.

To generate non-negative continuous outcomes we use an exponential-regression model with log-normal disturbances. More precisely, we draw $y_{ij} = \mu_{ij}\,\varepsilon_{ij}$, where

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

for $\sigma_{ij}^2 > 0$. This implies that $E[\varepsilon_{ij}|x_{ij}] = 1$ and $\text{var}(\varepsilon_{ij}|x_{ij}) = \sigma_{ij}^2$. We will take $\sigma_{ij}^2 \in \{1, \mu_{ij}^{-1}, 1 + \mu_{ij}^{-1}, \mu_{ij}^{-2}\}$. These cases correspond to $\text{var}(y_{ij}|x_{ij})$ being in $\{\mu_{ij}^2, \mu_{ij}, \mu_{ij}(1+\mu_{ij}), 1\}$. The first specification has homoskedastic errors. The second specification has Poisson-type errors, with the conditional mean equaling the conditional variance, and the third specification gives an overinflated variance as in a negative-binomial model with $\theta = 1$. The fourth specification, finally, gives homoskedastic outcomes. In this model, $\Pr(y_{ij} = 0|\, x_{ij}) = 0$.

The next model has a mixed discrete/continuous outcome distribution with a mass point at zero. We follow Santos Silva and Tenreyro (2011) and generate the outcome y_{ij} from a χ^2 distribution with d_{ij} degrees of freedom, where d_{ij} is drawn from a negative-binomial distribution with shape parameter θ and scale parameter $p_{ij} = (1 + \mu_{ij}/\theta)^{-1}$. This implies that $\Pr(y_{ij} = 0|\, x_{ij}) = (1 - p_{ij})^{\theta}$ is non-zero. We will refer to this model as the inflated model and will generate data with $\theta \in \{5, 15\}$.

Taken together, this yields ten different data generating processes that represent well the various situations where exponential-regression models have been used in empirical work.

The conditional mean is set as follows. We first draw $(\log\alpha_i, \log\gamma_i)$ from a bivariate normal distribution with zero mean and unit variances and correlation ρ.

We then generate a bivariate regressor $x_{ij} = (x_{ij1}, x_{ij2})'$ from a distribution that depends on the fixed effects. To do this we proceed sequentially. We first draw the binary variable $x_{ij2} = v_i v_j = x_{ji2}$, where $v_i = 1\{\log \alpha_i - \log \gamma_i \geq t_\rho\}$ and the threshold t_ρ is set such that $\Pr(v_i = 1) = \sqrt{1/2}$, so $\Pr(x_{ij2} = 1) = 1/2$. We then draw the second regressor, x_{ij1}, from a mixture of two skew-normal distributions (Azzalini 1985). Moreover, we draw x_{ij1} from a normal distribution with mean 1 and variance 1 when $x_{ij2} = 0$ and from a right-skewed normal distribution (with noncentrality parameter set to 3) with mean -1 and variance 1 when $x_{ij2} = 1$. In this way we introduce dependence between both regressors and between the regressors and the fixed effects. Furthermore, x_{ij} and $x_{i'j'}$ are dependent unless $\{i, j\}$ and $\{i', j'\}$ are disjoint. Below we report simulation results for $\rho = -1/4$. Throughout we fix $\psi_0 = (\psi_1, \psi_2)' = (-1, 1)'$.

We present results for two just-identified GMM estimators. The first estimator (GMM1) has $\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)$ set equal to

$$(x_{ij} - x_{ij'}) - (x_{i'j} - x_{i'j'}),$$

while the second estimator (GMM2) uses

$$\big\{(x_{ij} - x_{ij'}) - (x_{i'j} - x_{i'j'})\big\} \times \varphi(x_{ij}, \psi)\, \varphi(x_{i'j'}, \psi)\, \varphi(x_{i'j}, \psi)\, \varphi(x_{ij'}, \psi).$$

Apart from being intuitive and obvious choices, they can be motivated through moment calculations using the formulae in Chamberlain (1987) for one-quad data. In our context, these moment conditions depend on the fixed effects, in general. GMM1 uses Chamberlain's moments obtained under the assumption that errors are homoskedastic and no fixed effects are present. Similarly, GMM2 uses an approximation to his moments under the assumption that the data are Poisson distributed and no fixed effects are present. Detailed calculations are collected in the Supplementary Material.

We also report results for the Poisson pseudo maximum-likelihood estimator

(PMLE), which is widely used in applied work but whose sampling properties in two-way models have not been well studied (see Gouriéroux, Monfort and Trognon 1984*a*,*b* and Santos Silva and Tenreyro 2006).[5] The PMLE estimator can be slow to compute in large samples as the number of parameters to estimate grows with n and the $(2n + \dim \psi) \times (2n + \dim \psi)$ Hessian matrix is not block diagonal; see Guimarães and Portugal (2010), for example. Although the number of moment conditions for GMM does not depend on n, brute-force evaluation of $s(\psi)$ requires $O(n^4)$ operations. In large samples, such an approach may be infeasible. Fortunately, brute-force evaluation can be avoided, and the GMM estimators can be computed quite rapidly. A more detailed discussion on this is provided in the Supplementary Material. Here we just note that the average time required to compute the point estimate and the standard error in our designs with $n = 25$ was broadly 1.00 seconds for PMLE, 0.05 seconds for GMM1, and 0.25 seconds for GMM2. For $n = 100$ the average computational time was 110 seconds for PMLE, 4.25 seconds for GMM1, and 21 seconds for GMM2. So, GMM1 and GMM2 are roughly 20 times and 5 times faster to compute than PMLE, respectively. MATLAB code for point estimation and inference based on GMM1 and GMM2 is available as supplementary material to this paper.

In Table 1 we present results for $n = 25$ while in Table 2 we provide results for $n = 100$. Each table contains the median bias, the interquartile range, and the actual coverage rate of 95% confidence intervals for the all three estimators and for all ten designs considered. Also reported are L-estimates (Hosking 1990) of the standard deviation of each estimator based on the interdecile range and the presumption of normality.[6] These are robust estimates with a high breakdown

[5] Theoretical results for the Poisson maximum-likelihood estimator in $n \times m$ panel models under asymptotics where n and m grow at the same rate follow from Fernández-Val and Weidner (2016). The behavior of the estimator under more general asymptotics is currently unknown. The PMLE estimator has received a substantial amount of attention in the trade literature. However, to the best of my knowledge, the numerical evaluations in that literature do not look at dyadic data and do not consider data generating processes that include fixed effects.

[6] Denote the interdecile range across the Monte Carlo replications by IDR and let $\operatorname{erf}(a)$ be the error function at a. Then the L-estimator of the standard error of a normal random variable equals the ratio $\mathrm{IDR}/(2\sqrt{2}\operatorname{erf}^{-1}(.80))$.

point. All simulation results were obtained over $10,000$ Monte Carlo replications. All the regressors, disturbances, and fixed effects are redrawn in each Monte Carlo replication.

All estimators perform well in terms of bias and interquartile range. Across all models and designs, none uniformly dominates. This is not surprising given the large differences between the various designs. Turning to inference we see that our asymptotics provide a rather good approximation to the small-sample behavior of both GMM estimators for both samples sizes considered. Moreover, the observed coverage rates are close to their theoretical level of .95. The coverage rates of PMLE are more volatile. Inn several of the designs, and especially for $n = 25$, they are quite a bit smaller than their theoretical values. Analogously, the t-test based on PMLE heavily overrejects under the null. Consequently, inference based on this estimator is less reliable.[7]

Finally, to assess the sensitivity of the estimators to measurement error in the outcome variable, we also investigate their performance in the log-normal model from above when we only observe y_{ij} rounded to the nearest integer value, as in Santos Silva and Tenreyro (2006). The results, for $n = 25$, are in Table 3. Note that all estimators lose their theoretical validity and so none of them is guaranteed to be consistent in this case. All estimators are now more biased, notably for ψ_2. Regarding ψ_1, PMLE and GMM2 continue to perform well and behave very similarly. The GMM1 estimator of ψ_1 suffers from larger bias. With regard to inference, the coverage rates for PMLE are broadly unaffected by the rounding errors and continue to be too low. Those of GMM1 worsen somewhat due to the presence of bias, while those of GMM2 continue to provide very reliable inference throughout.

Our simulation study suggests that our GMM estimators present a viable option for inference in exponential-regression models with two-way fixed effects. Like

[7]Due to the estimation of the fixed effects, the score contributions of PMLE are strongly correlated across observations. The variance estimator fails to capture this and so delivers standard errors that tend to be too small. This implies that confidence bounds are too narrow.

Table 1—Simulation results for $n = 25$

Model	PMLE		GMM1		GMM2		PMLE		GMM1		GMM2	
	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2
	median bias						interquartile range					
Poisson	-.0004	-.0003	-.0095	-.0106	-.0006	.0010	.0329	.3539	.0974	.5373	.0395	.3841
Negbin												
1	.0033	.0057	-.0056	.0138	.0117	.0136	.1877	.7100	.1710	.7650	.2396	.8067
5	.0020	.0050	-.0074	.0026	.0035	.0089	.0911	.4664	.1206	.6031	.1173	.5134
10	.0001	-.0012	-.0076	-.0097	.0004	.0009	.0689	.4105	.1088	.5747	.0891	.4480
Normal												
1	.0060	-.0030	.0024	-.0054	.0130	-.0010	.1748	.5545	.1229	.4811	.2200	.6184
μ^{-1}	-.0005	-.0003	-.0115	-.0158	-.0002	.0017	.0326	.3413	.0782	.4345	.0379	.3631
$1+\mu^{-1}$	.0056	.0019	-.0071	-.0148	.0128	.0131	.1792	.6531	.1468	.6333	.2285	.7260
μ^{-2}	-.0011	.0238	-.0238	.0130	-.0003	.0301	.0152	.3478	.0851	.5611	.0132	.3598
Inflated												
5	.0006	-.0027	-.0196	-.0218	.0027	.0042	.1031	.6847	.1721	.9213	.1271	.7526
15	-.0014	.0118	-.0189	-.0095	-.0001	.0142	.0756	.6366	.1627	.8896	.0950	.7015
	standard deviation (L-estimates)						coverage rate (95%)					
Poisson	.0248	.2746	.0757	.4244	.0294	.2971	.9213	.9156	.9480	.9511	.9544	.9394
Negbin												
1	.1399	.5455	.1315	.5915	.1789	.6183	.8518	.8761	.9476	.9361	.9380	.9398
5	.0683	.3535	.0912	.4720	.0879	.3968	.8691	.9016	.9504	.9466	.9450	.9009
10	.0516	.3184	.0829	.4453	.0665	.3523	.8728	.9039	.9510	.9487	.9406	.8956
Normal												
1	.1310	.4205	.0949	.3577	.1673	.4771	.8360	.8571	.9470	.9297	.9313	.9518
μ^{-1}	.0241	.2680	.0605	.3432	.0287	.2855	.9263	.8859	.9305	.9334	.9549	.9080
$1+\mu^{-1}$	.1340	.4973	.1106	.4867	.1715	.5611	.8376	.8571	.9373	.9197	.9307	.9418
μ^{-2}	.0125	.2976	.0688	.4569	.0107	.3034	.9396	.8963	.8820	.9224	.9636	.9221
Inflated												
5	.0771	.5433	.1368	.7143	.0985	.5945	.8772	.8772	.9244	.9179	.9440	.8857
15	.0572	.5080	.1274	.7029	.0713	.5512	.8947	.8842	.9294	.9175	.9460	.8909

Table 2—Simulation results for $n = 100$

Model	PMLE		GMM1		GMM2		PMLE		GMM1		GMM2	
	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2
	median bias						interquartile range					
Poisson	-.0002	-.0028	-.0017	-.0065	-.0001	-.0028	.0073	.0787	.0246	.1326	.0082	.0806
Negbin												
1	.0023	.0141	-.0006	.0048	.0023	.0173	.0528	.1951	.0465	.1965	.0680	.2180
5	.0000	.0005	-.0006	-.0019	.0005	-.0004	.0259	.1153	.0314	.1448	.0322	.1241
10	.0001	.0013	-.0004	-.0001	.0006	.0023	.0191	.0991	.0285	.1378	.0239	.1057
Normal												
1	.0000	.0005	-.0006	-.0019	.0005	-.0004	.0259	.1153	.0314	.1448	.0322	.1241
μ^{-1}	-.0001	-.0009	-.0017	-.0060	.0001	-.0008	.0070	.0797	.0225	.1206	.0087	.0810
$1+\mu^{-1}$	.0021	-.0040	-.0021	-.0050	.0034	-.0030	.0533	.1895	.0430	.1869	.0665	.2086
μ^{-2}	-.0001	.0042	-.0057	.0049	.0000	.0040	.0035	.0938	.0291	.1918	.0028	.0916
Inflated												
5	.0007	-.0009	-.0019	-.0048	.0008	-.0015	.0286	.1617	.0467	.2342	.0352	.1717
15	.0001	.0039	-.0019	-.0015	.0006	.0029	.0194	.1437	.0424	.2274	.0235	.1550
	standard deviation (L-estimates)						coverage rate (95%)					
Poisson	.0054	.0605	.0183	.1004	.0060	.0616	.9428	.9408	.9500	.9472	.9608	.9584
Negbin												
1	.0409	.1468	.0340	.1466	.0518	.1615	.9068	.9248	.9604	.9532	.9420	.9040
5	.0195	.0859	.0229	.1086	.0239	.0932	.9152	.9332	.9588	.9568	.9568	.9128
10	.0145	.0754	.0205	.1048	.0175	.0792	.9172	.9308	.9496	.9480	.9588	.9376
Normal												
1	.0195	.0859	.0229	.1086	.0239	.0932	.9152	.9332	.9588	.9568	.9568	.9128
μ^{-1}	.0053	.0596	.0173	.0930	.0062	.0621	.9452	.9428	.9380	.9572	.9652	.9560
$1+\mu^{-1}$	.0402	.1415	.0332	.1401	.0500	.1550	.9212	.9224	.9500	.9448	.9524	.9136
μ^{-2}	.0027	.0698	.0231	.1509	.0021	.0715	.9500	.9472	.8948	.9560	.9600	.9536
Inflated												
5	.0210	.1229	.0356	.1785	.0257	.1298	.9148	.9336	.9392	.9456	.9520	.9284
15	.0146	.1106	.0320	.1730	.0174	.1146	.9248	.9420	.9408	.9484	.9564	.9560

TABLE 3—SIMULATION RESULTS WITH ROUNDING ERROR FOR $n = 25$

	PMLE		GMM1		GMM2		PMLE		GMM1		GMM2	
	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2	ψ_1	ψ_2
	median bias						interquartile range					
1	-.0060	.2259	-.0963	.3296	.0122	.2415	.1765	.6632	.1775	.7596	.2174	.7471
μ^{-1}	-.0083	.1375	-.0648	.1592	-.0019	.1524	.0336	.4095	.1177	.6866	.0382	.4561
$1+\mu^{-1}$	-.0067	.1421	-.0685	.1597	.0075	.1632	.1789	.7278	.1899	.8634	.2269	.8125
μ^{-2}	-.0069	.1095	-.0619	.1001	-.0019	.1209	.0169	.4187	.1186	.7777	.0137	.4397
	standard deviation (L-estimates)						coverage rate (95%)					
1	.1299	.5111	.1347	.5808	.1672	.5843	.8436	.8171	.8719	.8657	.9359	.9414
μ^{-1}	.0255	.3297	.0941	.5345	.0291	.3583	.9012	.8604	.8576	.9114	.9545	.9131
$1+\mu^{-1}$	.1352	.5594	.1463	.6608	.1730	.6260	.8378	.8460	.9045	.9138	.9336	.9382
μ^{-2}	.0136	.3449	.0961	.6192	.0112	.3674	.8691	.8745	.8285	.9037	.9606	.9183

PMLE they have small bias in a broad range of data generating processes. At the same time, where the quality of inference based on PMLE tends to vary with the particular data generating process at hand, the observed coverage rates induced by GMM are consistently close to their theoretical rates across all designs considered, and this so even in relatively small samples. Furthermore, the GMM estimators, and GMM2 in particular, appear to be fairly robust to rounding errors in the outcome variable, much as PMLE (Santos Silva and Tenreyro 2006), and continue to provide excellent inference. This may be an issue in empirical applications. Of course, the optimal choice of GMM estimator depends on the application at hand. Moment calculations as those described above can be of use here.

IV. Empirical application

We use data of Santos Silva and Tenreyro (2006) to estimate a gravity equation with multilateral resistance terms (Anderson and van Wincoop, 2003) in levels. These data contain information on 136 countries, giving $136 \times 135 = 18,360$ directed trade flows. About 52% of these flows are positive. As outcome variable we use bilateral trade, measured in $1,000$ U.S. dollars. As distance measures we use (the logarithm of) 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 common border, speak the same language, have a colonial history, and take part in a common free-trade agreement. Table 4 provides summary statistics for all variables in the full sample and in the subsample of positive trade flows.

Table 5 provides point estimates and standard errors (in parentheses) for GMM (GMM2 from the simulations[8,9]) and PMLE, both when using the full sample

[8] GMM1 as defined above is not well suited for these data. As all regressors are non-negative we have that $\|s(\psi)\| \to 0$ and $\|S(\psi)\| \to 0$ as (one or more of) the elements of ψ grow large. A similar issue arises in the one-way model and is discussed in Wooldridge (1997, Endnote 2). One possible adjustment to the moment condition is to transform x_{ij} into $x_{ij} - \overline{x}$, where $\overline{x}$ is the overall mean of the regressors, and premultiply $(x_{ij} - x_{i'j'}) - (x_{i'j} - x_{ij'})$ by $\varphi(\overline{x}, \psi)$.

[9] Another reason to prefer GMM2 in the context of the empirical application is its relatively good

TABLE 4—SUMMARY 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	0.0251	0.1563	0.0445	0.2063

(trade ≥ 0) and when using the subsample of positive trade flows (trade > 0). We also provide results for the fixed-effect ordinary least-squares (OLS) estimator of the log-linearized gravity equation.

Overall, GMM and PMLE provide similar point estimates, taking into account standard errors. This is the case both for the full sample and for the subsample of positive trade flows. Both estimators find that geographical distance tends to decrease trade while sharing a common language tends to increase trade. The estimated elasticities range between $-.75$ and $-.77$; and between .38 and .50, respectively. PMLE additionally finds sharing a common border to be a statistically-significant driver behind the magnitude of trade flows. The GMM estimate of the common-border effect is smaller and the associated standard error does not allow to distinguish it from zero at conventional significance levels. The difference between the two estimates is not unreasonably large when taking into account estimation noise. These findings are in line with the simulation results reported on in the previous section. The OLS point estimates differ most greatly on geographical distance and the importance of colonial ties, with both point estimates being larger in magnitude. The remaining point estimates are similar, again taking into account standard errors. For discussion on the appropriateness of working with a log-linearized estimating equation in the gravity context, see Santos Silva and Tenreyro (2006).

performance in the simulations with measurement error in the outcome variable.

Table 5—Gravity estimates

outcome variable: trade volume (in 1,000 U. S. dollars)					
	GMM		PMLE		OLS
	trade ≥ 0	trade > 0	trade ≥ 0	trade > 0	trade > 0
log distance	-.751	-.767	-.750	-.770	-1.347
	(.057)	(.059)	(.041)	(.042)	(.031)
common border	.149	.135	.370	.352	.174
	(.077)	(.078)	(.091)	(.090)	(.130)
common language	.491	.500	.383	.418	.406
	(.093)	(.092)	(.093)	(.094)	(.068)
colonial past	.213	.198	.079	.038	.666
	(.121)	(.121)	(.134)	(.134)	(.070)
free trade agreement	.330	.335	.376	.374	.310
	(.125)	(.125)	(.077)	(.076)	(.098)

References

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

Azzalini, Adelchi. 1985. "A class of distributions which includes the normal ones." *Scandinavian Journal of Statistics*, 12: 171–178.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust inference with multiway clustering." *Journal of Business & Economic Statistics*, 29: 238–249.

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

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

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

Fernández-Val, Ivan, and Martin Weidner. 2016. "Individual and time effects in nonlinear panel data models with large N, T." *Journal of Econometrics*, 192: 291–312.

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

Gouriéroux, Christian, Alain Monfort, and Alain Trognon. 1984*b*. "Pseudo maximum likelihood methods: Theory." *Econometrica*, 52: 681–700.

Guimarães, Paulo, and Pedro Portugal. 2010. "A simple feasible procedure to fit models with high-dimensional fixed effects." *Stata Journal*, 10: 628–649.

Hansen, Christian. 2007. "Asymptotic properties of a robust variance matrix estimator for panel data when T is large." *Journal of Econometrics*, 141: 597–620.

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

Hosking, Jonathan R. M. 1990. "L-moments: Analysis and estimation of distributions using linear combinations of order statistics." *Journal of the Royal Statistical Society, Series B*, 52: 105–124.

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

Santos Silva, João M. C., and Silvana Tenreyro. 2011. "Further simulation evidence on the performance of the Poisson-PML estimator." *Economics Letters*, 112: 220–222.

van der Vaart, Aad W. 2000. *Asymptotic Statistics.* Cambridge University Press.

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