SciencesPo
DEPARTMENT OF ECONOMICS

Discussion paper 2015-09

TWO-WAY MODELS FOR GRAVITY

Koen Jochmans

Sciences Po Economics Discussion Papers

TWO-WAY MODELS FOR GRAVITY

Koen Jochmans*

Draft: February 2, 2015 Revised: November 18, 2015

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 first derive moment conditions that are free of fixed effects. We next use these moment conditions to set up estimators that are n-consistent, asymptotically normally-distributed, and asymptotically unbiased. We provide Monte Carlo evidence for a range of models and we estimate a gravity equation with multilateral resistance terms as an empirical illustration.

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. A leading example is a gravity equation for bilateral trade flows between countries. Such models feature both importer and exporter effects at least since the work of Anderson and van Wincoop (2003). While such two-way models are intuitively 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

* 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. Discussions with Maurice Bun, Thierry Mayer, and Martin Weidner were much appreciated. This work was partially supported by the SAB grant 'Econometric analysis of linked data'. Part of the work on this paper was done while the author was visiting the London School of Economics, whose hospitality is gratefully acknowledged.

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 et al., 1984). 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 (2014) 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. Further, the class of models considered by Fernández-Val and Weidner (2014) and the one under study here are different, and they are not nested. Also, they work under sampling conditions that would be difficult to maintain in the current context.[1]

In the likelihood setting, an alternative for some models can be to work with a conditional likelihood (Andersen 1970). Charbonneau (2013) investigates this possibility for several models for count data, but she does not provide distribution theory.

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. 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]

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

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

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',

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

By differencing these equations we then obtain the conditional moment condition

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

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.[5]

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. We now provide distribution theory for this estimator. All proofs are collected in the Appendix.

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 $\overline{s}(\psi) = \lim_{n\to\infty} s(\psi_k)$, $\|\overline{s}(\psi)\| \to 0$ 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.

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

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 $n \to \infty$.*

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 summands in (2.2) are all zero-mean random variables that are independent conditionally on $x_{11}, \ldots, x_{nn}$. A suitable central limit theorem then 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 $S(\psi_n) \xrightarrow{p} \varSigma$ 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*

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

as $n \to \infty$, where the covariance matrix is $\varUpsilon = (\varSigma' \varOmega \varSigma)^{-1} (\varSigma' \varOmega V \varOmega \varSigma)(\varSigma' \varOmega \varSigma)^{-1}$.

As usual, the asymptotic variance is minimized by setting $\varOmega_n = V_n^{-1}$ 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

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

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.

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

$$\mu_{ij} = \exp(x_{ij}' \psi_0)\, \alpha_i\, \gamma_j.$$

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

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 $\mathrm{var}(\varepsilon_{ij}|x_{ij}) = \sigma_{ij}^2$. We will take $\sigma_{ij}^2 \in \{1, \mu_{ij}^{-1}, 1+\mu_{ij}^{-1}, \mu_{ij}^{-2}\}$. These cases correspond to $\mathrm{var}(y_{ij}|x_{ij})$ 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}. 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 $\Pr(v_i = 1) = \sqrt{1/2}$, and so $\Pr(x_{ij2} = 1) = 1/2$.

We then draw the second regressor, x_{ij1}, from a normal distribution with mean $-(2x_{ij2}-1)$ and variance one. So, the two regressors are negatively correlated and depend on 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 one-step GMM estimators based on two sets of moment conditions. The first estimator (GMM1) has $\phi(x_{ij},x_{ij'},x_{i'j},x_{i'j'};\psi)$ set equal to

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

while the second estimator (GMM2) uses

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

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.[6] In Table 1 we present the median bias, interquartile range, and coverage rates of 90% and 95% confidence intervals for these three estimators and for all designs considered. All results were obtained over $1,000$ Monte Carlo replications and for data of size $n=25$.

All estimators perform well and similar in terms of bias and interquartile range. Across all models and designs, none uniformly dominates. Turning to inference we see that our asymptotic results provide a good approximation to the small-sample behavior of both GMM estimators, even though the sample size considered is very small. Moreover, the observed coverage rates are close to their theoretical levels of 90% and 95%. This is true for all designs. The coverage rates of PMLE stay far below their theoretical values in almost all designs, and so inference based on this estimator is not 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

Table 1—Simulation results

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	-.001	.008	-.002	.005	.000	.007	.026	.208	.051	.330	.033	.248
Negbin												
1	.008	-.002	-.004	.016	.028	-.072	.148	.564	.129	.553	.179	.657
5	.004	.013	-.003	.024	.004	.019	.069	.347	.075	.412	.090	.402
10	.001	-.018	-.002	.011	.003	-.025	.053	.279	.066	.374	.065	.328
Normal												
1	.014	-.008	.002	-.013	.022	-.026	.116	.488	.103	.376	.153	.568
μ^{-1}	.000	.008	-.002	.016	.001	.008	.027	.213	.049	.287	.033	.234
$1+\mu^{-1}$	.006	-.026	-.001	-.022	.019	-.058	.133	.512	.111	.477	.167	.555
μ^{-2}	.001	.007	-.008	.037	-.002	.007	.014	.156	.043	.321	.015	.168
Inflated												
5	-.001	.012	-.012	-.009	.001	.009	.080	.487	.104	.601	.097	.526
15	-.004	-.017	-.011	-.011	-.003	-.018	.060	.402	.093	.594	.075	.472
	coverage rate (90%)						coverage rate (95%)					
Poisson	.848	.849	.897	.882	.940	.881	.908	.920	.953	.941	972	.950
Negbin												
1	.667	.790	.880	.871	.858	.868	.751	.862	.936	.940	.906	.928
5	.701	.817	.882	.862	.850	.884	.778	.890	.937	.913	.915	.928
10	.742	.822	.901	.876	.862	.888	.817	.888	.955	.930	.937	.934
Normal												
1	.697	.788	.876	.877	.866	.872	.768	.857	.938	.938	.922	.926
μ^{-1}	.826	.855	.873	.872	.917	.895	.898	.913	.934	.927	.967	.940
$1+\mu^{-1}$	.671	.803	.879	.848	.861	.876	.753	.873	.929	.917	.924	.931
μ^{-2}	.887	.867	.830	.859	.941	.921	.937	.933	.886	.901	.972	.954
Inflated												
5	.759	.823	.871	.850	.899	.883	.845	.893	.925	.908	.952	.936
15	.801	.843	.876	.863	.914	.900	.870	.910	.938	.928	.956	.942

from above when we only observe y_{ij} rounded to the nearest integer value, as in Santos Silva and Tenreyro (2006). Note that none of the estimators is guaranteed to be consistent in this case. The results are in Table 2. In terms of bias and spread, PMLE and GMM2 continue to perform well and behave very similarly. GMM1 is somewhat more biased, especially with respect to ψ_2. 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 slightly due to the presence of bias, while those of GMM2 continue to provide very reliable inference throughout.

Table 2—Simulation results with rounding error

	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	.001	.084	-.038	.218	.014	.069	.136	.494	.121	.501	.166	.562
μ^{-1}	-.009	.060	-.034	.144	-.007	.070	.030	.236	.061	.391	.034	.272
$1+\mu^{-1}$	.006	.060	-.026	.141	.021	.041	.129	.539	.131	.582	.170	.603
μ^{-2}	-.008	.059	-.026	.135	-.006	.062	.016	.192	.054	.407	.017	.209
	coverage rate (90%)						coverage rate (95%)					
1	.669	.786	.820	.802	.845	.880	.768	.860	.900	.875	.912	.927
μ^{-1}	.807	.819	.761	.818	.921	.874	.875	.887	.842	.898	.957	.926
$1+\mu^{-1}$	.677	.785	.849	.826	.851	.877	.768	.857	.924	.898	.919	.925
μ^{-2}	.794	.823	.731	.801	.889	.889	.859	.903	.816	.864	.941	.945

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 3 provides summary statistics for all variables in the full sample and in the subsample of positive trade flows.

Table 3—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

Table 4 provides point estimates and standard errors (in parentheses) for GMM (GMM2 from the simulations[8]) and PMLE, both when using the full sample (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 (along with heteroskedasticity-robust standard errors).

Table 4—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)

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 unreasonable large when taking into account estimation noise, however. The OLS point estimates differ most greatly on geographical distance and the importance of colonial ties, with both point estimates being larger in magnitude than for GMM and PMLE. The remaining point estimates are similar, again taking into account standard errors.

Notes

[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.

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

[5]An issue that we do not address is semiparametrically-efficient estimation. The classic results of Chamberlain (1987) do not readily extend 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.

[6]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 (2014). 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.

[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.

[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 $\|\psi\| \to \infty$. A similar issue arises in the one-way model and is discussed in Wooldridge (1997, Endnote 2).

Appendix

Proof of Theorem 1. We establish consistency by verifying Conditions (i)–(iv) of Theorem 2.1 in Newey and McFadden (1994). Assumptions 1 and 2 imply that Conditions (i)–(iii) hold. Condition (iv) states that $s(\psi)$ converges in probability to $\overline{s}(\psi)$ uniformly on $\mathcal{S}$ and remains to be shown. By definition we need to show that

$$\lim_{n\to\infty} \Pr\left(\sup_{\psi\in\mathcal{S}} \|s(\psi) - E[s(\psi)]\| > \epsilon\right) = 0$$

for any $\epsilon > 0$.

By symmetry,

$$s(\psi) - E[s(\psi)] = \frac{\varrho^{-1}}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i'\neq i} \sum_{j'\neq j} \overline{v}(\{i, i', j, j'\}, \psi),$$

where $\overline{v}(\{i, i', j, j'\}, \psi) = v(\{i, i', j, j'\}, \psi) - E[v(\{i, i', j, j'\}, \psi)]$ and we introduce the notational shorthand

$$v(\{i, i', j, j'\}, \psi) = \phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)(u_{ij}(\psi)\, u_{i'j'}(\psi) - u_{ij'}(\psi)\, u_{i'j}(\psi)).$$

By the Cauchy-Schwarz inequality,

$$\sqrt{E[\|\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)\, u_{ij}(\psi)\, u_{i'j'}(\psi)\|^2]}$$

is bounded by

$$E[\|\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)\|^4] \times \sqrt{E[\|u_{ij}(\psi)\|^8]}\, \sqrt{E[\|u_{i'j'}(\psi)\|^8]}.$$

By Assumption 4, these terms are uniformly bounded on $\mathcal{S}$ for any i, i', j, j'. Therefore, there exists a constant C so that

$$E[\|v(\{i, i', j, j'\}, \psi)\|^2] < C.$$

This, in turn, implies that the variance of $s(\psi)$ is uniformly bounded. Chebychev's inequality yields

$$\Pr\left(\|s(\psi) - E[s(\psi)]\| > \epsilon\right) \leq \frac{E[\|s(\psi) - E[s(\psi)]\|^2]}{\epsilon}.$$

The numerator on the right-hand side is bounded by

$$\frac{\begin{array}{l}\sum_{i_1=1}^{n}\sum_{j_1=1}^{n}\sum_{i_2\neq i_1}\sum_{j_2\neq j_1} \\ \sum_{i_3=1}^{n}\sum_{j_3=1}^{n}\sum_{i_4\neq i_3}\sum_{j_4\neq j_3}\end{array} E[\|\overline{v}(\{i_1,i_2,j_1,j_2\},\psi)\|\,\|\overline{v}(\{i_3,i_4,j_3,j_4\},\psi)\|]}{16\varrho^2}$$

The covariance between $v(\{i_1,i_2,j_1,j_2\},\psi)$ and $v(\{i_3,i_4,j_3,j_4\},\psi)$ depends on how many of the indices are common across the quadruples (i_1,i_2,j_1,j_2) and (i_3,i_4,j_3,j_4). The correlation is non-zero as soon as these sets overlap. When the sets are disjoint, the terms $v(\{i_1,i_2,j_1,j_2\},\psi)$ and $v(\{i_3,i_4,j_3,j_4\},\psi)$ are independent by virtue of Assumption 3. Of the $O(n^8)$ possible combinations of observations, $O(n^7)$ combinations have dyads that overlap. Hence, uniformly on $\mathcal{S}$,

$$\Pr\left(\|s(\psi) - E[s(\psi)]\| > \epsilon\right) = O(n^{-1}),$$

which converges to zero as $n \to \infty$ for any $\epsilon > 0$. Therefore, uniform convergence of the empirical moment condition $s(\psi)$ to $\overline{s}(\psi)$ has been established. With all conditions of Theorem 2.1 in Newey and McFadden (1994) fullfilled we have established that

$$\psi_n - \psi_0 \xrightarrow{p} 0$$

as $n \to \infty$. The proof is complete. □

Proof of Theorem 2. The proof proceeds in three steps. We first show (2.2). We next establish the uniform convergence of the Jacobian matrix of the moment conditions. We then combine these results with a Taylor expansion to establish that

$$\text{(a)} \qquad n\,(\psi_n - \psi_0) = -(\Sigma'\Omega\Sigma)^{-1}\,\Sigma'\Omega\,\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(\varepsilon_{ij}-1) + o_p(1),$$

and apply a suitable central limit theorem to the right-hand side of this equation to validate Theorem 2.

(i) Asymptotic approximation of the moment conditions. At ψ_0 the empirical moment conditions are

$$\text{(b)} \qquad s(\psi_0) = \frac{\varrho^{-1}}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{i'\neq i}\sum_{j'\neq j}\phi(x_{ij},x_{ij'},x_{i'j},x_{i'j'};\psi_0)(u_{ij}\,u_{i'j'} - u_{ij'}\,u_{i'j}),$$

where we have exploited the symmetry of $s(\psi_0)$ in (i,i') and (j,j'). A small calculation shows that the Hájek projection (van der Vaart, 2000, Section 11.3) of $s(\psi_0)$, conditional on the covariates, equals

$$p_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(\varepsilon_{ij}-1).$$

Note that $E[\,p_n\,|\,x_{11},\ldots,x_{nn}] = 0$ and that

$$V_* = n^2\,\mathrm{var}(p_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E[w_{ij}w_{ij}'\,\sigma_{ij}^2].$$

To show that $s(\psi_0)$ is asymptotically equivalent to p_n in the sense of (2.2) it suffices to show that

$$\text{(c)} \qquad n^2\,E[(p_n - s(\psi_0))(p_n - s(\psi_0))'] \to 0$$

as $n\to\infty$ (see, e.g., van der Vaart 2000, Chapter 12).

The main step needed to establish (c) is the calculation of the variance of the moment conditions $s(\psi_0)$. Use (b) to see that $\mathrm{var}(s(\psi_0)) = E[s(\psi_0)s(\psi_0)']$ equals the expectation of the matrix

$$\text{(d)}\quad \begin{aligned} &\frac{\varrho^{-1}}{4}\sum_{i_1=1}^{n}\sum_{i_2\neq i_1}\sum_{j_1=1}^{n}\sum_{j_2\neq j_1}\phi(x_{i_1j_1},x_{i_1j_2},x_{i_2j_1},x_{i_2j_2};\psi_0)\,(u_{i_1j_1}\,u_{i_2j_2}-u_{i_1j_2}\,u_{i_2j_1}) \\ \times&\frac{\varrho^{-1}}{4}\sum_{i_3=1}^{n}\sum_{i_4\neq i_3}\sum_{j_3=1}^{n}\sum_{j_4\neq j_3}\phi(x_{i_3j_3},x_{i_3j_4},x_{i_4j_3},x_{i_4j_4};\psi_0)'(u_{i_3j_3}\,u_{i_4j_4}-u_{i_3j_4}\,u_{i_4j_3}). \end{aligned}$$

Because $u_{ij}=\alpha_i\gamma_j\,\varepsilon_{ij}$ and the ε_{ij} are independent across both i and j, we have that

$$\text{(e)}\qquad E[(u_{i_1j_1}\,u_{i_2j_2}-u_{i_1j_2}\,u_{i_2j_1})(u_{i_3j_3}\,u_{i_4j_4}-u_{i_3j_4}\,u_{i_4j_3})|\,x_{11},\ldots,x_{nn}]$$

equals zero unless any of the dyads in $\{(i_1,j_1),(i_2,j_2),(i_1,j_2),(i_2,j_1)\}$ co-incides with any of the dyads in $\{(i_3,j_3),(i_4,j_4),(i_3,j_4),(i_4,j_3)\}$. Of the $O(n^8)$ terms in $\text{var}(s(\psi_0))$, $O(n^6)$ have at least dyad in common. Moreover, the number of terms with two or more dyads in common is $O(n^4)$. Because

$$\text{var}(s(\psi_0))=\frac{O(n^6)}{\varrho^2}=\frac{O(n^6)}{O(n^8)}=O(n^{-2}),$$

only terms with at least one dyad in common provide a non-zero contribution to the asymptotic variance. By symmetry of (d), all the expressions are permutation invariant and so we are free to choose a dyad that is common across terms in our calculations and multiply through the resulting expression by 4^2, thereby accounting for all possible choices. With $(i_3,j_3)=(i_1,j_1)$, the expectation in (e) equals

$$\text{(f)}\qquad \alpha_{i_1}^2\gamma_{j_1}^2\alpha_{i_2}\gamma_{j_2}\alpha_{i_4}\gamma_{j_4}\,\sigma_{i_1j_1}^2,$$

Setting $(i_3,j_3)=(i_1,j_1)$ in (d), and using (f) and the definition of w_{ij} given in the text we find

$$n^2\,\text{var}(s(\psi_0))=V_*+o(1).$$

The same argument can be used to show that $n^2\, E[s(\psi_0)\, p_n'] = V_* + o(1)$. Hence,

$$E[(p_n - s(\psi_0))(p_n - s(\psi_0))'] = o(n^{-2})$$

and (c) has been shown.

(ii) Uniform convergence of the Jacobian matrix. Differentiating $s(\psi)$ gives the Jacobian as

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

Convergence of the second term to its expectation follows as in the proof of Theorem 1, with ϕ' replacing ϕ, by Assumptions 4 and 5. For the first term, observe that

$$\frac{\partial (u_{ij}(\psi) u_{i'j'}(\psi) - u_{i'j}(\psi) u_{ij'}(\psi))}{\partial \psi'}$$

equals

$$u_{i'j}(\psi) u_{ij'}(\psi) (\tau(x_{i'j}; \psi)' + \tau(x_{ij'}; \psi)') - u_{ij}(\psi) u_{i'j'}(\psi) (\tau(x_{ij}; \psi)' + \tau(x_{i'j'}; \psi)').$$

Assumptions 4 and 5 imply that

$$E[\| \phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi)\, u_{ij}(\psi) u_{i'j'}(\psi)\, \tau(x_{ij}; \psi)' \|^2] < C$$

for some finite constant C. Therefore, again by the same argument as in the proof of Theorem 1, this term converges uniformly to its expectation which, as will be verified below, equals Σ. By Theorem 1, $\|\psi_n - \psi_0\| = o_p(1)$ as $n \to \infty$. Therefore,

$$\| S(\psi) - \Sigma \| \xrightarrow{p} 0$$

for any ψ that lies in between ψ_n and ψ_0. This conclusion, together with the asymptotic equivalence of $s(\psi_0)$ and p_n, can be combined with a mean-value expansion of $s(\psi_n)$ around ψ_0 to obtain the sampling-error representation for $n\,(\psi_n - \psi_0)$ in (a).

To see that the limit of the expectation of $S(\psi_0)$ equals Σ, first note that the term involving ϕ' drops out because

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

Therefore, up to $o_p(1)$, $S(\psi_0)$ equals

$$\varrho^{-1}\sum_{i=1}^{n}\sum_{i<i'}\sum_{j=1}^{n}\sum_{j<j'} E[\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)\,((\tau_{i'j}+\tau_{ij'})-(\tau_{ij}+\tau_{i'j'}))'\,\alpha_i\alpha_{i'}\gamma_j\gamma_{j'}],$$

where we let $\tau_{ij} = \tau(x_{ij}; \psi_0)$. Exploit symmetry and expand the sum on the right-hand side to see that

$$\begin{aligned} S(\psi_0) = {} & \frac{\varrho^{-1}}{4}\sum_{i=1}^{n}\sum_{i'\neq i}\sum_{j=1}^{n}\sum_{j'\neq j} E[\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)\,(\alpha_i\alpha_{i'}\gamma_j\gamma_{j'})\,\tau'_{i'j}] \\ & + \frac{\varrho^{-1}}{4}\sum_{i=1}^{n}\sum_{i'\neq i}\sum_{j=1}^{n}\sum_{j'\neq j} E[\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)\,(\alpha_i\alpha_{i'}\gamma_j\gamma_{j'})\tau'_{ij'}] \\ & - \frac{\varrho^{-1}}{4}\sum_{i=1}^{n}\sum_{i'\neq i}\sum_{j=1}^{n}\sum_{j'\neq j} E[\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)\,(\alpha_i\alpha_{i'}\gamma_j\gamma_{j'})\tau'_{ij}] \\ & - \frac{\varrho^{-1}}{4}\sum_{i=1}^{n}\sum_{i'\neq i}\sum_{j=1}^{n}\sum_{j'\neq j} E[\phi(x_{ij}, x_{ij'}, x_{i'j}, x_{i'j'}; \psi_0)\,(\alpha_i\alpha_{i'}\gamma_j\gamma_{j'})\tau'_{i'j'}] + o_p(1). \end{aligned}$$

By permutation invariance, the fourth right-hand side term is identical to the third while the first and second right-hand side terms are identical to the third up to sign. Collapsing the four expressions on the right-hand side and using the definition of w_{ij} we therefore find that

$$S(\psi_0) = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} E[w_{ij}\,\tau'_{ij}] + o_p(1) \to \Sigma$$

as $n \to \infty$.

(iii) Central limit theorem. Steps (i) and (ii) validate the linear approximation stated in (a). Theorem 2 will then follow by showing that

$$\text{(g)} \qquad n\, V^{-1/2}\, p_n \xrightarrow{d} N(0, I),$$

where I denotes the identity matrix of conformable dimension. To do so note that, conditional on $x_{11}, \ldots, x_{nn}$, the Hájek projection p_n is an average of independent heterogeneously-distributed zero-mean random variables with variance $n^{-2}\, W_*$, where

$$W_* = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} w_{ij}'\, \sigma_{ij}^2.$$

By virtue of Assumption 4, p_n satisfies Lyapunov's condition and so, applying a conditional version of the central limit theorem (see, e.g., Prakasa Rao 2009), we have

$$n\, W_*^{-1/2}\, p_n \xrightarrow{d} N(0, I),$$

conditional on $x_{11}, \ldots, x_{nn}$. Now, $\|W_* - V\| \leq \|W_* - V_*\| + \|V_* - V\|$ by the triangle inequality, and each of these right-hand side terms converges to zero in probability as $n \to \infty$. Therefore, the conditional limit distribution of $n\, p_n$ is normal with zero mean and constant covariance V. Because this distribution is independent of the covariate values $x_{11}, \ldots, x_{nn}$ it equals the unconditional limit distribution. This yields (g). Therefore, the theorem has been shown. □

References

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

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.

Charbonneau, K. B. (2013). *Multiple fixed effects in theoretical and applied econometrics.* Ph. D. thesis, Princeton University.

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.

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Elsevier.

Prakasa Rao, B. L. S. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics 61*, 441–460.

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

Santos Silva, J. M. C. and S. Tenreyro (2011). Further simulation evidence on the performance of the Poisson-PML estimator. *Economics Letters 112*, 220–222.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

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