Estimating trade network models without network econometrics*

Peter H. Egger†
ETH Zurich
CEPR, CESifo, GEP, WIFO

Kevin E. Staub‡
University of Melbourne
IZA

August 13, 2025

Abstract

This paper studies structural gravity network models à la Behrens, Ertur and Koch (2012). Such models provide an elegant linearization of nonlinear structural gravity models of international trade, and have a wide range of other applications to bilateral flow data such as investments and migration. In the context of trade, these models account for so-called multilateral resistance (or aggregate price index) terms in a theory-consistent way. Earlier research had proposed applying network-econometric techniques for estimating such models. We exploit the structure of the model to propose a simple alternative OLS estimator that does not require any specific network methods, making the structural network model amenable to broader use by practitioners. We show that all structural model parameters can be recovered from a linear regression that uses a properly-weighted average of the dependent variable in a control function. Our control-function approach can also be implemented with simple nonlinear estimators instead of OLS, such as Poisson pseudo maximum likelihood or other generalized linear model estimators.

Keywords: Gravity models; Multilateral resistance terms; Network models.

JEL-codes: F14; C23.

*The authors are grateful to two anonymous reviewers and the editor in charge for numerous helpful comments on an earlier version of the manuscript.

†ETH Zurich, Department of Management, Technology, and Economics, Weinbergstr. 35, 8092 Zurich, Switzerland; E-mail: egger@kof.ethz.ch; + 41 44 632 41 08.

‡Department of Economics, 111 Barry Street, The University of Melbourne, 3010 VIC, Australia; E-mail: kevin.staub@unimelb.edu.au; +61 3 903 53776.

1 Introduction

The gravity equation—describing aggregate demand for goods or services between any pair of countries—is among the most successful concepts in all of economics (see Leamer and Levinsohn, 1995). Its popularity derives from the fact that it nests a wide variety of isomorphic structural models of aggregate bilateral demand such as endowment-economy, Ricardian, and monopolistic-competition increasing-returns-to-scale models with a fixed markup (see Eaton and Kortum, 2002; Anderson and van Wincoop, 2003; Arkolakis, Costinot and Rodríguez-Clare, 2012; Bergstrand, Egger and Larch, 2013; Baltagi, Egger and Pfaffermayr, 2015).

The main reason why gravity equations are rarely estimated in their structural form is that they are nonlinear in parameters after taking logarithms. Rather, practitioners use either country fixed effects, which may be inefficient, or theory-based linear approximations in estimation. Behrens, Ertur and Koch (2012, henceforth BEK) introduced such a linearization in a version of a constant-elasticity-of-substitution model, where price indices can be expressed as implicit functions of trade flows. The linearized equilibrium system leads to an econometric specification in which trade flows between two countries depend on trade flows between all trading partners, thus exhibiting the characteristics of a network-weighted model. To deal with this interdependence among observations, BEK adapt methods from the literature on network econometrics that account for simultaneous, cross-sectionally autoregressive structures. While the model has since been used, an existing barrier to its further dissemination lies in the computationally demanding network-based estimation.

In this paper, we propose estimators that exploit the structural form of the same network model to recover the model's structural parameters. In contrast to the previous network-based approach to estimating the model, the proposed estimator is a simple OLS (or Poisson PML—pseudo maximum likelihood) estimator. We start by showing that the reduced form of the structural model exists, and then use the properties derived from this reduced form to show that the structural form of the model can be estimated by OLS to obtain estimates of the key economic parameters. In our proposed approach, the endogeneity resulting from the model's network structure is *fully* captured through the inclusion of a single appropriate regressor that serves as a control function, making it possible to estimate the structural parameters from the model's other regressors. As a result, the proposed estimator is fully efficient in the sense of there not being any efficiency loss due to estimating the linear approximation rather than the true non-linearized model. Constructing this control function is straightforward as it is just a weighted average of the data and does neither require instruments nor additional estimation steps. Therefore, the proposed alternative estimator is easily implementable in practice, providing a low-cost way for practitioners to estimate this popular structural gravity network model, either applied to international trade flows as in the original BEK paper or applied to other bilateral flow data such as immigration or financial investment.

We examine the properties of the structural gravity model and this linearization as well as the finite sample performance of the estimators in a series of numerical Monte Carlo experiments. The simulation data are obtained by randomly drawing exogenous variables from real-world data and then solving for endogenous variables according to the theory-based structural general equilibrium model. As we show theoretically and confirm in the simulations, the only source of bias in a reduced-form OLS estimation of this model stems from the original linearization of the model. Overall, the simulation results demonstrate that this bias is moderate and shrinks towards zero with increasing sample size. Moreover, the proposed new control-function OLS approach does not suffer from this linearization bias and is unbiased in all sample sizes.

We continue with presenting the generic structural gravity model and its linearization following BEK in Section 2. We show how this model can be estimated by linear (or generalized linear) methods in Section 3. A numerical assessment of the estimators is provided in Section 4, followed by our conclusions in Section 5.

2 A linearized structural gravity network model of trade

We consider a standard gravity equation derived from utility maximization subject to income constraints, which can be represented for exporter i and importer j as

$$Z_{ij} \equiv \frac{X_{ij}}{Y_i Y_j} = \frac{C_i^{\alpha-1} T_{ij}^{\alpha}}{\sum_{k=1}^{N} L_k C_k^{\alpha} T_{kj}^{\alpha}} \qquad C_i = L_i^{-1} \sum_{j=1}^{N} X_{ij} = \sum_{j=1}^{N} \frac{C_i^{\alpha} T_{ij}^{\alpha} C_j L_j}{\sum_{k=1}^{N} L_k C_k^{\alpha} T_{kj}^{\alpha}}. \tag{1}$$

Here, Z_{ij} are aggregate bilateral exports, X_{ij}, normalized by exporter and importer GDP, Y_i and Y_j, respectively. The variable C_i represents the costs per unit of a single or a bundle of factors L_i, T_{ij} are bilateral trade costs, and L_i is size (up to scale) of a market measured in terms of its factor supply (see Arkolakis et al., 2012). As mentioned, (1) is compatible with a variety of aggregate bilateral demand models such as endowment-economy, Ricardian, and monopolistic-competition-increasing returns-to-scale models (see the appendix for more details).

Equation (1) shows that the log of Z_{ij}, z_{ij}, is a log-nonlinear function of $\{C_i; T_{ij}; L_i\}$. The key structural parameter $\alpha = (-\infty, 0)$ reflects the partial response of trade with respect to changes in trade costs (see Dixit and Stiglitz, 1977, or Eaton and Kortum, 2002). Through (1), upon choice of a numéraire cost ($C_1 = 1$) and for a given α, the $N-1$ endogenous values of C_i are determined by $N-1$ equations for given (exogenous) values of L_i and T_{ij}^{α}.[1]

BEK take the logarithm of equation (1) and derive a first-order approximation around the point $\alpha = 0$. To provide a compact notation, let us generally use the convention that lower-case letters refer to variables in logarithms. and let us refer to N-size vectors and $N \times N$ square matrices by

[1] In a Dixit-Stiglitz-Krugman-type model the size of countries is parameterized by the endowment L_i. In an Eaton-Kortum Ricardian economy this would be productivity, and in an Armington economy it would be a preference shifter. For the purpose of the arguments in this paper, these differences are only semantic.

subscripts N and NN, respectively.[2] Letting the world endowment size be denoted by $L \equiv \sum_{k=1}^{N} L_k$, define the following vectors and matrices:

(i) the $N \times 1$ column vector ω_N, whose ith row element is $L_i/L \in (0,1)$,

(ii) the $N \times 1$ column vector of ones ι_N,

(iii) the $N \times N$ asymmetric matrix $W_{NN} = \iota_N \omega'_N$ which in every row of column j contains the same elements L_j/L,

(iv) the identity matrix $I_{NN} = diag(\iota_N)$.

Let us stack all observations on Z_{ij} across exporters i for a given importer j into the $N \times 1$ vector of log bilateral normalized exports $z_{jN} = (z_{ij})$, stack log world endowment into an $N \times 1$ vector $l_N = (l) = l\iota_N$ with identical row entries, stack log unit factor-bundle costs into the $N \times 1$ vector $c_N = (c_i)$, and stack log ad-valorem trade costs across all exporters i for a given importer j into the $N \times 1$ vector $t_{jN} = (t_{ij})$. Using this notion, BEK arrive at a log-transformed and linearized counterpart to (1):

$$z_{jN} = \alpha W_{NN} z_{jN} + (\alpha - 1)(l_N + c_N) + \alpha(I_{NN} - W_{NN})t_{jN} + \eta_{jN}, \qquad (2)$$

where η_{jN} is the approximation error due to linearization, which only varies between importers j but not between exporters i.[3,4] In its role as the coefficient on $W_{NN}z_{jN}$, BEK refer to α in equation (2) as the *autoregressive interaction coefficient*. They do so, because $W_{NN}z_{jN}$ captures the interdependence of bilateral exporters across countries and provides an intuitive network measure of "spatial interaction" or "competition". It is customary in empirical work to further specify $t_{ij} = \sum_{h=1}^{H} \gamma_h d_{h,ij}$ (see Anderson and van Wincoop, 2003, 2004, Eaton and Kortum, 2002, and many others), where $d_{h,ij}$ are H-many observable trade-cost variables in logs such as bilateral log bilateral distance. What is then estimated on $d_{h,ij}$ are the compound parameters $\alpha\gamma_h$.

[2] Recall that N parameterizes the number of countries, and N^2 the number of country pairs, including every domestic relation where $i = j$ is one such pair for every country i.

[3] This is equation (11) in BEK. The notation in BEK uses the parameterization $\alpha \equiv 1 - \sigma$ and stacks further all importers j. BEK linearize the model about an importer-specific term, the log-transformed ideal consumer-price term. As this term lacks variation across exporters i, the corresponding approximation error, too, is importer-specific.

[4] After defining $Y \equiv \sum_{i=1}^{N} Y_i$, at the approximation point $\alpha = 0$ of the model, we obtain $X_{ij} = \frac{L_i Y_j}{L}$ and $Y_i = \sum_{j=1}^{N} X_{ij} = \frac{L_i Y}{L}$, which implies factor-cost equalization, $C_i = C$. Choosing C as the numéraire, we obtain $X_{ij} = \frac{L_i L_j}{L}$. Then, trade costs are irrelevant, and the variance of log bilateral exports, $x_{ij} = \ln X_{ij}$, is fully determined by the variation in exporter- and importer-specific log factor endowments across countries or regions i and j.
Take logs of X_{ij} on both sides of equation (1) and see that αt_{ij} is the only ij-variant term in that equation. What remains is either linear in α – namely all i-specific terms – or j-specific and non-linear in α. The latter originates in the log of a sum across exporters of multiplicative terms for each importer in the denominator of (1). Hence, any linearization error with respect to α must stem from the j-specific term, which involves a sum of identically constructed ij-specific elements across i. Subtracting the log-linearized form in (2) from the GDP-normalized original nonlinear form in logs in (1) eliminates αt_{ij} from the difference. We will illustrate later in the Monte Carlo simulations that the difference between the structural and the reduced forms indeed lacks variation in the dimension i.

The reduced form that directly corresponds to (2) is

$$z_{jN} = (I_{NN} - \alpha W_{NN})^{-1}[(\alpha - 1)(l_N + c_N) + \alpha(I_{NN} - W_{NN})t_{jN} + \eta_{jN}]. \quad (3)$$

Existence and uniqueness of the latter requires that $(I_{NN} - \alpha W_{NN})$ is uniquely invertible. Finiteness of the model solution requires that $[(\alpha - 1)(l_N + c_N) + \alpha(I_{NN} - W_{NN})t_{jN} + \eta_{jN}]$ is an $N \times 1$ vector with finite elements. The latter property is trivially fulfilled in this setting. In contrast, the unique invertibility of the matrix $(I_{NN} - \alpha W_{NN})$ is not obvious in general. For the specific matrix at stake, the necessary assumptions for a unique inverse solution are based on insights that will be outlined in the next section. The network matrix and parameter space considered here is fundamentally different from what is typically assumed in spatial and network econometrics. E.g., the weighting matrix W_{NN} has nonzero diagonal elements. Hence, the network features self-loops (see Newman, 2018), which induce what Manski (1993) called the "reflection problem" (see also Bramoullé et al., 2009). It was the presence of these self-loops, which led BEK to reformulate equation (3) in a way that complicates estimation tremendously. Moreover, the parameter on the dependent variable weighted with W_{NN}, here α, is typically assumed to be smaller than unity in absolute value $|\alpha| < 1$ with row-normalized or maximum-row-sum-normalized network matrices. Here, W_{NN} is row-normalized for theoretical reasons, and $\alpha < -1$ throughout quantitative work in international economics. Nonetheless, we will be able to establish the unique analytical invertibility of $(I_{NN} - \alpha W_{NN})$ in the subsequent section.

3 Novel insights and econometric methods

3.1 Properties of the linearized gravity network model

Despite network matrices with self-loops falling outside the assumptions covered in the aforementioned literature in econometrics and statistics, such network matrices are quite common in economics. The so-called Leontief inverse—which is based on a selling-sector revenue-scaled inter-sector input-output-flow matrix—is one of the most prominent examples. Earlier work established results regarding the existence and uniqueness of such problems involving Leontief-type inverses with self-loops (see Woodbury, 1949, Lampert and Scholtes, 2023, Bellido and Prieto-Martínez, 2024), and we build on these results to show the existence of the reduced form for the linearized gravity network model in the following lemma.

Lemma 1 (Inverse of $(I_{NN} - \alpha W_{NN})$)**.** *Let each country exhibit an endowment that is positive and finite with $0 < \underline{L} \leq L_i \leq \overline{L} < \infty$, where $\{\underline{L}, \overline{L}\}$ are bounding constants for country endowments. Let the parameter α be bounded in the compact interval $-\infty < \underline{\alpha} \leq \alpha \leq \overline{\alpha} \leq 0$, where $\{\underline{\alpha}, \overline{\alpha}\}$ are bounding constants for α.*

Then, the inverse of the matrix $(I_{NN} - \alpha W_{NN})$ exists and is unique, and it is given by the Sherman-

Morrison-Woodbury formula as[5]

$$(I_{NN} - \alpha W_{NN})^{-1} = I_{NN} + \frac{\alpha}{1-\alpha} W_{NN}. \tag{4}$$

Proof. Let us use $a = -\alpha$ and the $M \times 1$ vector $a_N = a\iota_N$. Note that I_{NN} is a special case of an invertible square matrix with real-valued entries. And note that a_N and ω_N are special cases of column vectors with real-valued entries. In general, the Sherman-Morrison-Woodbury formula for a non-singular, real-valued $N \times N$ matrix I_{NN} and real-valued $N \times 1$ column vectors (a_N, ω_N) forming the $N \times N$ matrix $a_N \omega_N' = -\alpha W_{NN}$ states that (see Sherman and Morrison, 1950; Woodbury, 1950; Riedel, 1992; Hao and Simoncini, 2021)

$$(I_{NN} + a_N \omega_N')^{-1} = I_{NN}^{-1} - \frac{I_{NN}^{-1} a_N \omega_N' I_{NN}^{-1}}{1 + \omega_N' a_N}.$$

Clearly, $I_{NN}^{-1} = I_{NN}$ with an identity matrix. Moreover $I_{NN}^{-1} a_N \omega_N' I_{NN}^{-1} = a_N \omega_N' = -\alpha W_{NN}$ and $1 + \omega_N' a_N = 1 + a = 1 - \alpha$, because the row entries of ω_N' are all positive, smaller than unity, and sum up to one. Hence,

$$(I_{NN} - \alpha W_{NN})^{-1} = I_{NN} + \frac{\alpha}{1-\alpha} W_{NN}$$

for every real-valued scalar α in the admissible parameter region $[\underline{\alpha}, \overline{\alpha}]$ and any row-normalized $N \times N$ network matrix $W_{NN} = \iota_N \omega_N'$ based on the real-valued weight vector ω_N', whose elements sum up to unity. □

Our result on the existence of the inverse in Lemma 1 and therefore the existence of the reduced form (3) contributes to a recent literature in international economics which studies the graph stability and finiteness of responses in shocks of graphs in trade and migration models with self-loops (Allen et al., 2020; Kucheryavyy et al., 2023; Allen et al., 2024; Kucheryavyy et al., 2024; Bifulco et al., 2025). A key difference of our model to this literature is that here the model is linear in parameters. Thus, while this literature has to impose (strong) restrictions on parameters to establish the uniqueness of equilibrium responses to shocks (e.g., the exact or near symmetry or the relative magnitude of the

[5] In most of the work on spatial and network econometrics, one works with matrices W_{NN} (all or the maximum of) whose row sums are unity and a parameter α which obeys $|\alpha| < 1$. Then, the inverse of interest can be expressed as a Neumann series of the form $(I_{NN} - \alpha W_{NN})^{-1} = I_{NN} + \sum_{q=1}^{\infty} (\alpha W_{NN})^q$ with a finite sum. But note that $|\alpha| < 1$ is not only not required with Sherman-Morrison-Woodbury inverses but is also starkly contradicted by data and estimates in a vast quantitative trade literature. E.g., the results in Broda and Weinstein (2006), Simonovska and Waugh (2014), or Fontagné et al. (2018) suggest that for most products $\alpha \in [-2; -20]$. And it is fair to say that values of $\alpha \in [-3; -8]$ are most commonly found or used in quantitative work at the country level (see Eaton and Kortum, 2002; Anderson and van Wincoop, 2003, 2004; Egger et al., 2011; Arkolakis et al., 2012; Bergstrand et al., 2013).

elements t_{ij} of the trade-cost matrix), existence in the present case follows under relatively general conditions, as we have shown.[6]

Based on Lemma 1, we can further simplify the reduced form of the non-stochastic model in (3). To this end, note that any invariant vector v_N such as l_N or η_{jN} has the property that $I_{NN}v_N = W_{NN}v_N = v_N$. Moreover, note that W_{NN} is idempotent. To see this, recall that $W_{NN} = (\iota_N\omega'_N)$. Therefore, $W^2_{NN} = (\iota_N\omega'_N)(\iota_N\omega'_N) = \iota_N(\omega'_N\iota_N)\omega'_N$, where $(\omega'_N\iota_N) = 1$ by design. Using these results, we can state the right-hand-side terms of (3) as

$$\begin{aligned}(I_{NN} - \alpha W_{NN})^{-1}(\alpha - 1)(l_N + c_N) &= -(l_N + c_N) + \alpha(I_{NN} - W_{NN})c_N,\\ (I_{NN} - \alpha W_{NN})^{-1}\alpha(I_{NN} - W_{NN})t_{jN} &= \alpha(I_{NN} - W_{NN})t_{jN},\\ (I_{NN} - \alpha W_{NN})^{-1}\eta_{jN} &= \frac{1}{1-\alpha}\eta_{jN},\end{aligned}$$

so that, after rearranging terms, we can write the reduced form (3) as

$$z_{jN} = -(l_N + c_N) + \alpha(I_{NN} - W_{NN})(c_N + t_{jN}) + \frac{1}{1-\alpha}\eta_{jN}. \tag{5}$$

Further, by pre-multiplying both sides of (5) with W_{NN} we obtain

$$W_{NN}z_{jN} = -W_{NN}(l_N + c_N) + \frac{1}{1-\alpha}\eta_{jN} \tag{6}$$

where $-W_{NN}(l_N + c_N)$ is a constant. Solving this equation for η_{jN} gives

$$\eta_{jN} = (1-\alpha)[W_{NN}z_{jN} + W_{NN}(l_N + c_N)]. \tag{7}$$

Because $W_{NN}z_{jN}$ does not vary across i within j but only between j, the approximation error η_{jN} also varies between j only.

3.2 Novel estimators for the linearized gravity network model

Conventional network-econometric models that are linear in parameters in their structural form and involve a term $W_{NN}z_{jN}$ in an equation with z_{jN} as the dependent variable are nonlinear in parameters. For this case, previous work proposes instrumental-variable estimators to estimate the

[6] The structure and configuration of the network $W_{NN} = \iota_N\omega'_N$ is different from ones considered in most of the network-econometrics and social-interactions literature (see Kelejian and Prucha, 1999, Lee, 2003, 2004, Bramoullé et al., 2009, and many others), not only because of the presence of self-loops. It is also special in that it relies on the distribution of node weights in the overall network (here, a node being a country). The weight of node i in the network is the same for any node j, allowing the network-weighted averages of the characteristics of the nodes across all nodes N to be the same for every node i. To the best of our knowledge, the literature on network and social interactions does not provide insights into this particular situation. However, we are faced with such a network matrix and problem for theoretical economic reasons.

structural form of autoregressive network models that are linear in parameters as is the model in (2) (see Kelejian and Prucha, 1998, Lee, 2003, or Kelejian et al., 2004). A general result of our paper is that the reduced form of such models is not nonlinear but linear in parameters for linear structural network models where the weights matrix is constructed as $W_{NN} = \iota_N \omega'_N$ with ω'_N being a weights vector that will appear in each row of the idempotent matrix W_{NN}. The nature of $\{W_{NN}, l_N, c_N, \eta_{jN}\}$ and the parameter restrictions in the non-stochastic model imply that all of the variation in $W_{NN}z_{jN}$ is due to the approximation error. In other words, $W_{NN}z_{jN}$ is fully collinear with the approximation error. Including it absorbs the approximation error. Armed with these insights, we now discuss the estimation of the structural and reduced-form models by OLS.

Stochastic error term

For estimation, we introduce empirical counterparts to the structural equation (2) and the reduced-form equation (5) that in addition to the approximation error vector η_{jN} have a random error vector, ε_{jN}. The latter varies across ij in the data, and we interpret it as an additive measurement error of log bilateral exports (multiplicative measurement error in bilateral exports), in the same way as it was introduced and motivated by BEK. We assume the elements of ε_{jN} to be independently distributed and vary across all observations ij, so that $E(\varepsilon_{jN}\varepsilon'_{jN}) = diag_{N\times N}(\sigma^2_{\varepsilon,ij})$ in case of heteroskedasticity and $E(\varepsilon_{jN}\varepsilon'_{jN}) = \sigma^2_{\varepsilon} I_{NN}$ in case of homoskedasticity. We assume that the model's approximation error η_{jN} and the random error term ε_{jN} are independent of each other.

After adding $l_N + c_N$ on both sides of the equation and introducing the two-component error term $u_{jN} = \eta_{jN} + \varepsilon_{jN}$, we now have for the structural equation

$$\tilde{z}_{jN} \equiv z_{jN} + l_N + c_N \quad = \quad \alpha W_{NN} z_{jN} + \alpha(l_N + c_N) + \alpha(I_{NN} - W_{NN})t_{jN} + u_{jN}. \tag{8}$$

In equation (8), z_{jN} is determined by the structural model "up to" ε_{jN}; that is, $z_{jN} - \varepsilon_{jN}$ is the deterministic part of log normalized bilateral exports in vector form. Hence, the stochastic z_{jN} in the preceding equation (8) includes the stochastic error also in the z_{jN} which is part of the first term on the right-hand-side after the equality sign.

Similarly, we can write the stochastic counterpart to the reduced form (5) as

$$\tilde{z}_{jN} \quad = \quad \alpha(I_{NN} - W_{NN})(c_N + t_{jN}) + (I_{NN} + \frac{\alpha}{1-\alpha} W_{NN})u_{jN}, \tag{9}$$

where the error

$$(I_{NN} + \frac{\alpha}{1-\alpha} W_{NN})u_{jN} = \frac{1}{1-\alpha}\eta_{jN} + \frac{\alpha}{1-\alpha} W_{NN}\varepsilon_{jN} + \varepsilon_{jN},$$

and where $\frac{\alpha}{1-\alpha} W_{NN}\varepsilon_{jN}$ is a stochastic term that is invariant in i, similar to a j-specific random effect, and ε_{jN} is independently distributed by assumption.

Reduced-form OLS

We consider first the OLS estimation of the reduced form of the stochastic linearized model, that is, equation (9). The estimating equation corresponding to (9) relies exclusively on network-weighted exogenous variables but not on network-weighted lags of the dependent variable:

$$\tilde{z}_{jN} \quad = \quad \alpha^{RF}(I_{NN} - W_{NN})(c_N + t_{jN}) + e_{jN}^{RF}, \tag{10}$$

with the reduced-form regression error $e_{jN}^{RF} = (1-\alpha)^{-1}\eta_{jN} + \alpha(1-\alpha)^{-1}W_{NN}\varepsilon_{jN} + \varepsilon_{jN}$. Since the elements of e_{jN} feature equi-correlation across importers, inference after estimation of (10) should rely on cluster-robust standard errors clustered by importers. We denote this approach as RF-OLS, for reduced-form OLS.

In general, the approximation error η_{jN} that is part of e_{jN}^{RF} will be correlated with $(I_{NN} - W_{NN})c_N$ and $(I_{NN} - W_{NN})t_{jN}$, leading to bias in the RF-OLS estimation, $E(\hat{\alpha}^{RF}) \neq \alpha$. The reason is that η_{jN} is a function of z_{jN} (eq. 7) which in turn is a function of c_N and t_{jN}. The next section presents numerical experiments based on widely-used trade data to assess the extent and severity of this linearization bias and how it depends on sample size and the values of the structural model parameters.

Control-function OLS

Next we consider the OLS estimation of the model based directly on the stochastic structural equation (8). We write the corresponding estimating equation as

$$\tilde{z}_{jN} \quad = \quad \theta^{CF}\iota_N + \alpha^{CF}[c_N + (I_{NN} - W_{NN})t_{jN}] + \vartheta^{CF}W_{NN}z_{jN} + e_{jN}^{CF}, \tag{11}$$

where θ^{CF} is a regression constant which absorbs the invariant terms in u_{jN}. Importantly, the term $\vartheta^{CF}W_{NN}z_{jN}$ here works as a linear control function, which combines $\alpha W_{NN}z_{jN}$ from the main part of the structural model and the term $(1-\alpha)W_{NN}z_{jN}$ from η_{jN}. Through z_{jN}, the control function is also a function of ε_{jN}. Thus, altogether, the inclusion of the control function in the regression leaves the regression residual as being only a function of the stochastic error, $e_{jN}^{CF} = f(\varepsilon_{jN})$, but no longer a function of the linearization error, η_{jN}. Therefore, since the stochastic error ε_{jN} is independent of c_N and $(I_{NN} - W_{NN})t_{jN}$, the regression error e_{jN}^{CF} is uncorrelated with c_N and $(I_{NN} - W_{NN})t_{jN}$, so that $E(\hat{\alpha}^{CF}) = \alpha$. We denote this, our proposed approach, as CF-OLS, for control-function OLS.[7]

One can add a control function to the specification of RF-OLS to address the endogeneity in RF-OLS. Indeed, an appropriate control function for (10) is also $\vartheta W_{NN}z_{jN}$ since it controls for the

[7] Alternatively, one might call this approach structural-model OLS, but this could cause confusion by leading to the belief that $E(\hat{\vartheta}^{CF}) = \alpha$, which is clearly not the case. The coefficient on $W_{NN}z_{jN}$ absorbs the endogeneity from the structural-model error (as well as from the stochastic error) and has thus no structural interpretation.

exact form of the endogeneity induced by the linearization error. Remarkably, the result of adding this control function to (10) is precisely (11).

GLM implementation of reduced-form and control function approaches

The estimating equations for the reduced-form and and control-function approaches can also be implemented as generalized linear model (GLM) estimations. That is, instead of using OLS, equations (10) and (11) can be estimated via Poisson pseudo-likelihood estimation or other GLM procedures (see Santos Silva and Tenreyro, 2006) with appropriate importer-cluster-robust standard errors. For instance, the control-function approach could be based on the exponential of equation (11),

$$\tilde{Z}_{jN} \quad = \quad \exp(\theta^{CF} + \alpha^{CF}[c_N + (I_{NN} - W_{NN})t_{jN}] + \vartheta^{CF} W_{NN} z_{jN})\nu_{jN}^{CF}, \tag{12}$$

where $\tilde{Z}_{jN} = \exp(\tilde{z}_{jN})$ and $\nu_{jN}^{CF} = \exp(e_{jN}^{CF})$, and the parameters $\{\theta^{CF}, \alpha^{CF}, \vartheta^{CF}\}$ are estimated by Poisson regression of $\tilde{Z}_{jN}$ on a constant, $[c_N + (I_{NN} - W_{NN})t_{jN}]$ – or, instead, separately on c_N and on $(I_{NN} - W_{NN})t_{jN}$ – as well as on $W_{NN} z_{jN}$.

Whether OLS, Poisson, or some other GLM estimator is preferred depends on the higher-order properties of the stochastic error ε_{ij} (Santos Silva and Tenreyro, 2006). In this paper, we are more interested in how the approximation error η_{jN} affects estimation, and we further explore this numerically through simulation experiments below.

4 Monte Carlo experiments

4.1 Design of experiments

We construct *worlds* of countries and country pairs according to (1) where everything is known to the simulator, while the researcher does not know the parameters on the regressors. We consider two configurations regarding country numbers, $N \in \{30; 60\}$, leading to numbers of country pairs of $N^2 \in \{900; 3,600\}$. This corresponds to typical data situations found in empirical structural work on gravity models (see Eaton and Kortum, 2002; Anderson and van Wincoop, 2003; Balistreri and Hillberry, 2007; Behrens, Ertur and Koch, 2012). For each of these worlds, we consider three configurations $\alpha \in \{-2; -4; -9\}$, which are supported quantitatively by a sizeable body of work (see Arkolakis, Costinot and Rodríguez-Clare, 2012). Hence, there are six parameter configurations. For each of them, we randomly draw $1,000$ independent vectors of bilateral distances with typical element $DIST_{ij}$ and population sizes with typical element L_i from the empirical distribution of these variables as published by the Centre d'Études Prospectives et d'Informations Internationales for $DIST_{ij}$ and by the World Bank's World Development Indicators for L_i (using the year 2007). We parametrize bilateral trade costs as $t_{ij} = dist_{ij}^{\gamma_{dist}}$, where $dist_{ij}$ is the logarithm of $DIST_{ij}$. In line with the robust result of a coefficient on log distance of about $\beta_{dist} = \alpha\gamma_{dist} = -1$ in empirical gravity models, we assume that log distance is related to log trade costs t_{ij} by a parameter of $\gamma_{dist} = -1/\alpha$. Based on the draws for L_i and t_{ij}, the endogenous variables C_i and X_{ij} are solved by contraction-mapping based on (1).

4.2 Features of model variables and the approximation error

Before turning to estimation, it is useful to study some moments and the correlations of key variables in the model across all experiments. For this purpose, we report the averages of an analysis of variance of some key variables in Table 1 and average partial correlation coefficients in Table 2, each of them computed across all 1,000 draws within one of the six parameter configurations in $\{N; \alpha\}$. We first consider here a data generating process (DGP) without stochastic error component ($\varepsilon_{ij} = 0$ for all i, j), so that the model error $u_{ij} = \eta_{ij} + \varepsilon_{ij}$ consists entirely of the approximation error η_{ij}.

Table 1 reports on sums of squares of key model variables. It reports the total variation in each of the variables (row 'total'), as well as a decomposition of the total into variation across exporters (row 'i'), importers (row 'j') and across 'ij', i.e., the bilateral variation (row 'residual'). The table reveals that the (total) variation in the error, $u_{ij} = \eta_{ij}$, is large relative to normalized bilateral exports in logs, z_{ij}. Its size increases with the absolute level of α; i.e., with the distance to the approximation point used by BEK to linearize the model ($\alpha = 0$).

The approximation error varies to a greater degree than log factor costs, c_N, whose variance is the same as that of $(I_{NN} - W_{NN})c_N$. The relative magnitude of the sum of squares of u_{ij} relative to that of z_{ij} declines as N, the number of countries, rises. The variance of $(I_{NN} - W_{NN})t_{jN}$ is important relative to that of c_N. But its importance relative to u_{jN} depends on being closer to the approximation point for α. Clearly, the exporter- and importer-specific components in t_{ij} are symmetric by design (log-distance is symmetric). The pair-specific component of $(I_{NN} - W_{NN})t_{jN}$ naturally dominates the country-specific ones. Finally, as was clear from the theoretical derivations from the previous section, the variation in η_{ij} is purely importer-specific. This is because BEK's approximation is about an importer-specific term, the log ideal consumer-price index.

Table 2 shows that there is a perfect correlation between the elements of $W_{NN}z_{jN}$ and the ones of the error, u_{jN}, consistent with equation (6) and the fact that in this case $u_{jN} = \eta_{jN}$. There is some correlation between $(I_{NN} - W_{NN})t_{jN}$ and u_{jN}, although it is quite low in absolute value. Nevertheless, this means that in estimations where $(I_{NN} - W_{NN})t_{jN}$ is a regressor, the coefficient on it may exhibit some bias unless we condition on $W_{NN}z_{jN}$ (which means conditioning on η_{jN}, as mentioned before). This problem becomes more pertinent if the approximation error is larger, which is the case with a bigger absolute value of α.

Figure 1 visualizes the relationships in Table 2 based on one specific random draw for $N = 30$ and $\alpha = -4$. There are four general insights from an inspection of Figure 1 in conjunction with Table

2. First, the upper left panel of the figure documents that $W_{NN}z_{jN}$ is indeed perfectly correlated with $u_{jN} = \eta_{jN}$ as suggested by equation (6), and underscoring the rationale for using it as an ideal control function. Second, all of the panels in Figure 1 illustrate the block structure of u_{jN} which means it is not independently and identically distributed and which motivates the need to use clustered standard errors for inference. Third, while the correlation between u_{jN} and the other right-hand side model variables is weak on average, it may be stronger depending on the specific configuration of trade costs (t_{jN}) and population size (W_{NN}). From Table 2 we know that the risk of correlation between model variables and u_{jN} is higher for $(I_{NN} - W_{NN})t_{jN}$ than for c_N. Figure 1, for instance, illustrates a case where $(I_{NN} - W_{NN})t_{jN}$ is negatively and $(\alpha - 1)(l_N + c_N) + \alpha(I_{NN} - W_{NN})t_{jN}$ is positively correlated with u_{jN}. In such a case, we would expect the estimated parameter on $(I_{NN} - W_{NN})t_{jN}$ to be biased if we do not address the endogeneity with a control function. Altogether, we would expect a larger root-mean-squared error for the estimated parameter on this variable than on c_N or $(l_N + c_N)$, unless one controls for η_{jN}.

4.3 Parameter estimation

We compare the estimation of the linearized gravity network model via the two approaches, RF-OLS, based on (10), and CF-OLS, based on (11). To benchmark these estimators, we also compare them to a structural estimation of the original, non-linearized gravity model through an iterative least squares procedure. For this procedure, the dependent variable can also be defined as $\tilde{z}_{jN} = z_{jN} + l_N + c_N$, where z_{jN} are normalized trade flows, which is the same as in the RF-OLS and CF-OLS procedures. Taking the log of (1) and adding the stochastic error vector ε_{jN}, the structural model results in

$$\tilde{z}_{jN} = \theta^{SILS}\iota_N + \alpha^{SILS}(c_N + t_{jN}) - m_{jN} + \varepsilon_{jN}, \tag{13}$$

where the log-nonlinear term is defined as $m_{jN} \equiv \ln(\sum_{k=1}^{N} L_k C_k^{\alpha} T_{kj}^{\alpha})$. The model can be estimated by structural iterative least squares (cf. Anderson and van Wincoop, 2003), which we denote by SILS. In our implementation of SILS, for an initial guess of $\widehat{m}_{jN}$, the coefficients appearing in equation (13) can be estimated by an OLS regression of $\tilde{z}_{jN} + \widehat{m}_{jN}$ on a constant and $c_N + t_{jN}$. The estimated coefficients can be used to obtain an updated estimate of $\widehat{m}_{jN}$, which in turn can be used to perform an updated OLS regression. These steps are iterated until the values of the estimated coefficients converge.

Note that for (13) there is no approximation error term η_{jN}, since no approximation has been applied; this is the true non-linearized structural model. And since in our first DGP there is also no random error (that is, $\varepsilon_{jN} = 0$), SILS in this case is an algorithm to *solve* for the model parameters, rather than to estimate them.

For all three models—RF-OLS (10), CF-OLS (11), and SILS (13)—we only present unconstrained parameter estimations in which c_N and $(I_{NN} - W_{NN})t_{jN}$ (or t_{jN} for SILS) are treated as two separate regressors. Thus, this does not enforce that the coefficients on both variables are the same (i.e., equal to α) due to the chosen parametrization. We do so to mimic the situation of an empirical researcher who does not observe t_{ij} but only $dist_{ij}$. To distinguish the two parameters in our results, we refer to the coefficient on c_N as α_c and to that on t_{jN} or $(I_{NN} - W_{NN})t_{jN}$ as α_t.

The estimated parameters $\{\hat{\alpha}_c; \hat{\alpha}_t\}$ should be close to the true α, especially, when being based on models (13) or (11). While in the latter model there is an approximation error, we saw in Section 3 that the control function $W_{NN}z_{jN}$ accounts *fully* for it (that is, not just in expectation, as is typically the case with control function approaches; see, e.g., Wooldridge, 2015).

Apart from a data generating process (DGP) where the structural nonlinear model (13) is true, we consider a second DGP with the additional stochastic error term ε_{jN} introduced in Section 3. Specifically, we specify the random error as $\varepsilon_{ij} \overset{IID}{\sim} N(0, \sigma_\varepsilon^2)$. We calibrate σ_ε^2 so that, in each experiment, the explanatory power as measured by the R^2 is 80% $(= (1 - \sigma_\varepsilon^2/\sigma_{\tilde{z}^*}^2) \times 100\%)$, which is representative of a vast amount of empirical work on gravity models. The term ε_{ij} adds stochastics in a narrow sense which provides for a residual with Models (13) and (11), and one beyond the approximation (or linearization) error in Model (10).

We report on the average bias and root-mean-squared error (RMSE) in percent of the true α across all draws per configuration of $\{N; \alpha\}$ in Tables 3 and 4. Both tables are organized in three by two blocks. Each horizontal block contains estimates for the models SILS, CF-OLS and RF-OLS for the cases $N = \{30; 60\}$. Vertically, we have three blocks corresponding to $\alpha = \{-2; -4; -9\}$. For each of the six blocks, we report on the estimated structural parameters $\hat{\alpha}_c$ and $\hat{\alpha}_t$; as well as $\hat{\vartheta}$ for CF-OLS, the coefficient on the control function.

Table 3 reports the results for the DGP without ε_{ij}; that is, only with the approximation error η_{ij} present: $u_{jN} = \eta_{jN}$. In the absence of ε_{ij}, both models SILS and CF-OLS correspond to the true one so that both their biases and the RMSEs for $\{\hat{\alpha}_c; \hat{\alpha}_c\}$ in percent are zero. That is, the CF-OLS estimator of the linearized gravity model is optimal in the sense of suffering zero efficiency loss due to the linearization. Recall that conditioning on $W_{NN}z_{jN}$ means conditioning on η_{jN}, according to (6). Thus, all the linearization bias is picked up by the coefficient $\hat{\vartheta}$. Only the RF-OLS approach has an estimation residual and therefore a non-zero variance under this DGP. The RF-OLS estimates of α exhibit some bias, but both bias and RMSE are relatively small. For instance, the largest bias in absolute value over all DGP settings and parameters is only -1.21 percent. Further, as expected, both the bias and the RMSE of $\{\hat{\alpha}_c; \hat{\alpha}_t\}$ in percent decline as the number of countries rises. As the number of countries increases, it has been shown that the need for controlling for general equilibrium effects and nonlinear trade-cost effects as captured by m_{jN} in (13) also declines (see Egger and Staub, 2016).

Table 4 depicts results for the DGP with both approximation error and stochastic error: $u_{jN} = \eta_{jN} + \varepsilon_{jN}$. The biases for all three estimators are very small, with the maximum bias in absolute value across all parameter configurations and estimators being 1 percent (for RF-OLS in the DGP with $\alpha = -2$ and $N = 30$). The simple CF-OLS estimator achieves RMSEs only slightly higher than those of the more demanding iterative SILS procedure. Even the simpler RF-OLS, while having the highest RMSE of the three estimators, still performs well by this measure.

Finally, in Table 5, we present results where all estimators are based on exponentiated equations and a Poisson (pseudo)-likelihood, as detailed for the control-function approach in equation (12). The estimation quality deteriorates somewhat across all approaches in that case. For SIP and CF-Pois (the Poisson versions of SILS and CF-OLS), the biases continue to be small, with the maximum absolute bias for CF-Pois being about 2 percent. RMSEs are also only moderately higher. With ε_{ij} being homoscedastic normal, OLS estimation in logs is efficient, so these results are expected. The reduced-form approach is dependent on both ε_{ij} and η_{ij}. The latter is not only non-normal, but also not fully mean-independent of the regressors. The results in the table show that, when exponentiated, the endogeneity is aggravated and RF-Pois suffers from larger absolute biases of up to 9.28 percent. From the perspective of the linearization error η_{ij}, this speaks for using the reduced-form approach in logs (RF-OLS) rather than in exponentiated form (RF-Pois). In contrast, for the control-function approach the properties of the linearization error are of little relevance, and the choice of CF-OLS or CF-Pois can be based entirely on other considerations, such as the statistical properties of the stochastic error.

5 Conclusions

This paper shed light on the nature of structural linearized gravity models involving an endogenous network-weighted lag – other countries' population-share – of bilateral trade flows as developed in Behrens, Ertur and Koch (2012). We demonstrated that the properties of the network model are such that it can be estimated without any use of network-econometric tools. Exporter-population-share-weighted log bilateral exports on the right-hand side of the model serve as a control function for the approximation error of the linearization, and this variable can be included without specific treatment (i.e., ignoring its endogeneity). These results should please the applied researcher, since estimation of such linearized models only involves OLS (on log-transformed trade flows) with clustered standard errors at the level of importers.

References

Allen T, Arkolakis C, Li X. 2024. On the equilibrium properties of spatial models. *American Economic Review: Insights* **6**: 472–489.

Allen T, Arkolakis C, Takahashi Y. 2020. Universal gravity. *Journal of Political Economy* **128**: 393–433.

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

Anderson JE, van Wincoop E. 2004. Trade costs. *Journal of Economic Literature* **42**: 691–751.

Arkolakis C, Costinot A, Rodríguez-Clare A. 2012. New trade models, same old gains? *American Economic Review* **102**: 94–130.

Balistreri EJ, Hillberry RH. 2007. Structural estimation and the border puzzle. *Journal of International Economics* **72**: 451–463.

Baltagi BH, Egger PH, Pfaffermayr M. 2015. Panel data gravity models of international trade. In Baltagi BH (ed.) *The Oxford Handbook of Panel Data*. Oxford, UK: Oxford University Press, 608–641.

Behrens K, Ertur C, Koch W. 2012. 'Dual' gravity: Using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics* **27**: 773–794.

Bellido JC, Prieto-Martínez LF. 2024. Existence and uniqueness of solutions for leontief's input–output model, graph theory and sensitivity analysis. *Linear and Multilinear Algebra* : 1–22.

Bergstrand JH, Egger PH, Larch M. 2013. Gravity redux: estimation of gravity-equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs. *Journal of International Economics* **89**: 110–121.

Bifulco P, Glück J, Krebs O, Kukharskyy B. 2025. Single and attractive: Uniqueness and stability of economic equilibria under monotonicity assumptions. *CESifo Working Paper No. 11623* .

Bramoullé Y, Djebbari H, Fortin B. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**: 41–55.

Broda C, Weinstein DE. 2006. Globalization and the gains from variety. *The Quarterly journal of economics* **121**: 541–585.

Dixit AK, Stiglitz JE. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* **67**: 297–308.

Eaton J, Kortum SS. 2002. Technology, geography, and trade. *Econometrica* **70**: 1741–1779.

Egger PH, Larch M, Staub KE, Winkelmann R. 2011. The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy* **3**: 113–143.

Egger PH, Staub KE. 2016. GLM estimation of trade gravity models with fixed effects. *Empirical Economics* **50**: 137–175.

Fontagné L, Martin P, Orefice G. 2018. The international elasticity puzzle is worse than you think. *Journal of International Economics* **115**: 115–129.

Hao Y, Simoncini V. 2021. The Sherman-Morrison-Woodbury formula for generalized linear matrix equations and applications. *Numerical Linear Algebra with Applications* **28**: e2384.

Kelejian HH, Prucha IR. 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* **17**: 99–121.

Kelejian HH, Prucha IR. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **40**: 509–533.

Kelejian HH, Prucha IR, Yuzefovich Y. 2004. Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample result. In LeSage J, Pace RK (eds.) *Advances in Econometrics*. New York: Elsevier, 163–198.

Kucheryavyy K, Lyn G, Rodríguez-Clare A. 2023. Grounded by gravity: A well-behaved trade model with industry-level economies of scale. *American Economic Journal: Macroeconomics* **15**: 372–412.

Kucheryavyy K, Lyn G, Rodríguez-Clare A. 2024. Spatial equilibria: The case of two regions. *Journal of International Economics* **152**: 104008.

Lampert M, Scholtes I. 2023. The self-loop paradox: Investigating the impact of self-loops on graph neural networks. In *The Second Learning on Graphs Conference*.

Leamer EE, Levinsohn J. 1995. International trade theory: the evidence. In Grossman GM, Rogoff K (eds.) *Handbook of International Economics Vol. 3*. North Holland, 1339–1394.

Lee LF. 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* **22**: 307–335.

Lee LF. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72**: 1899–1925.

Manski CF. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* **60**: 531–542.

Newman M. 2018. *Networks*. Oxford, UK: Oxford University Press, 1st edition.

Riedel KS. 1992. A Sherman–Morrison–Woodbury identity for rank augmenting matrices with application to centering. *SIAM Journal on Matrix Analysis and Applications* **13**: 659–662.

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

Sherman J, Morrison WJ. 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics* **21**: 124–127.

Simonovska I, Waugh ME. 2014. The elasticity of trade: Estimates and evidenceâ˜†. *Journal of International Economics* **92**: 34–50.

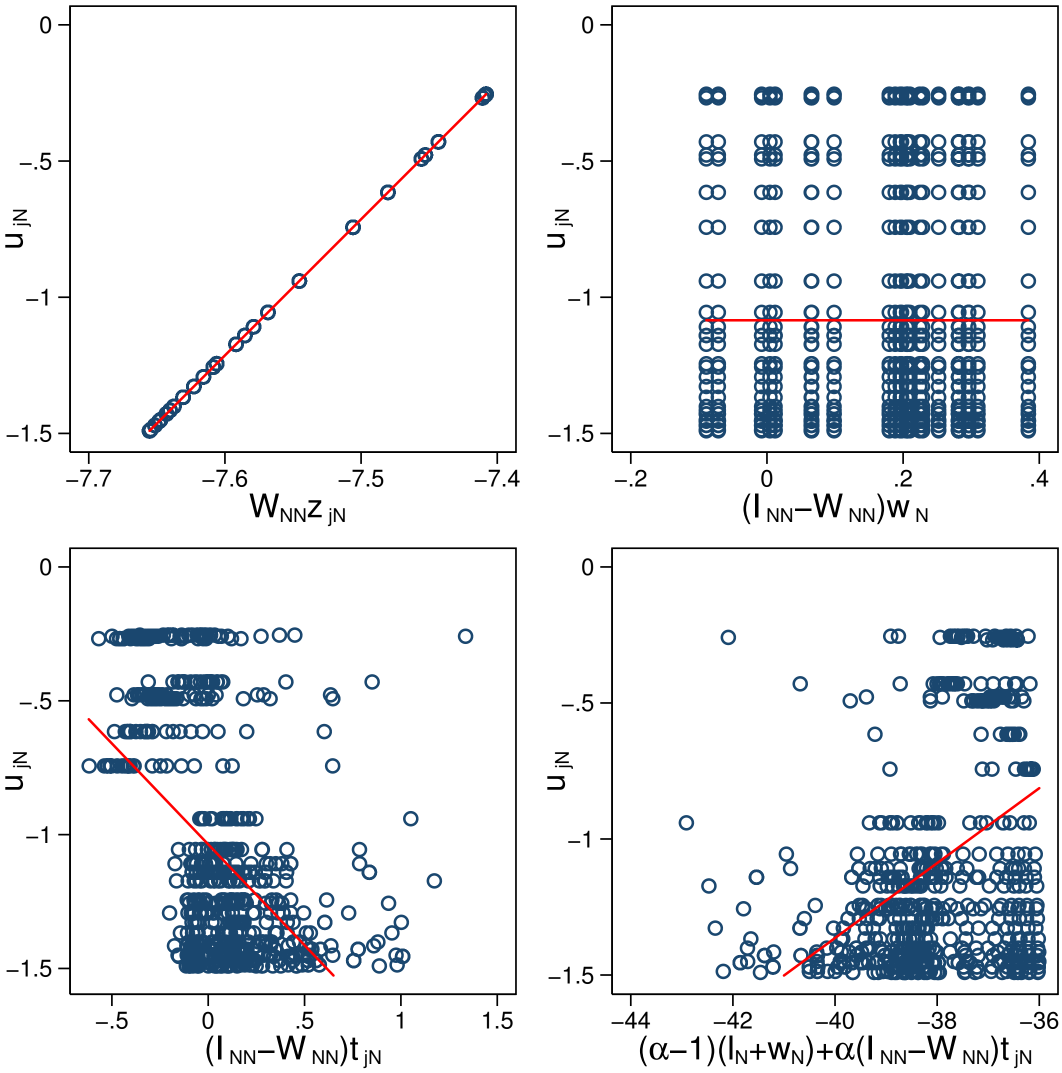
Woodbury MA. 1949. The stability of out-input matrices. *Chicago, IL* .

Woodbury MA. 1950. Inverting modified matrices. *Memorandum Rept. 42, Statistical Research Group, Princeton University, Princeton, NJ* .

Wooldridge JM. 2015. Control function methods in applied econometrics. *Journal of Human Resources* **50**: 420–445.

Tables and figures

Figure 1: Scatterplot and linear fit of approximation error ($u_{jN} = \eta_{jN}$) and right-hand side variables from a random draw of the DGP with $\alpha = -4$ for $N = 30$



***Notes*:** The four panels of the figure display scatterplots of data obtained from one random draw of the DGP with $\alpha = -4$ for 30 countries (900 observations). The red line represents the fit from a linear regression.

Table 1: Analysis of variance for key model variables (mean sums of squares over 1,000 replications)

	$N = 30$				$N = 60$			
SS	z_{ij}	u_{ij}	c_i	$\ddot{t}_{ij}$	z_{ij}	u_{ij}	c_i	$\ddot{t}_{ij}$
				$\alpha = -2$				
i (exporter)	112.26	0.00	14.99	20.47	370.46	0.00	52.21	77.74
j (importer)	112.26	68.40	0.00	29.39	370.46	171.00	0.00	95.89
residual	738.85	0.00	0.00	184.71	2398.63	0.00	0.00	599.66
total	963.38	68.40	14.99	234.58	3139.55	171.00	52.21	773.29
				$\alpha = -4$				
i (exporter)	127.93	0.00	6.01	5.11	416.30	0.00	20.81	19.47
j (importer)	127.93	205.19	0.00	7.47	416.30	519.76	0.00	24.03
residual	737.95	0.00	0.00	46.12	2400.35	0.00	0.00	150.02
total	993.82	205.19	6.01	58.70	3232.95	519.76	20.81	193.52
				$\alpha = -9$				
i (exporter)	139.89	0.00	1.63	1.01	458.94	0.00	5.64	3.84
j (importer)	139.89	949.64	0.00	1.48	458.94	2481.73	0.00	4.78
residual	737.95	0.00	0.00	9.11	2398.50	0.00	0.00	29.61
total	1017.72	949.64	1.63	11.60	3316.38	2481.73	5.64	38.23

Notes: SS refers to sum of squares. $\ddot{t}_{ij} \equiv t_{ij} - \sum_i \frac{L_i}{L} t_{ij}$ is a typical element of $(I_{NN} - W_{NN})t_{jN}$.

Table 2: Partial correlation coefficients of model variables with approximation error $u_{jN} = \eta_{jN}$ (mean and standard deviations over 1,000 replications)

	$N = 30$		$N = 60$	
	Mean	SD	Mean	SD
	$\alpha = -2$			
$W_{NN} z_{jN}$	1.00	0.00	1.00	0.00
$(l_N + c_N)$	-0.00	0.00	0.00	0.00
$(I_{NN} - W_{NN})t_{jN}$	0.06	0.14	0.06	0.10
	$\alpha = -4$			
$W_{NN} z_{jN}$	1.00	0.00	1.00	0.00
$(l_N + c_N)$	-0.00	0.00	-0.00	0.00
$(I_{NN} - W_{NN})t_{jN}$	-0.01	0.19	-0.02	0.14
	$\alpha = -9$			
$W_{NN} z_{jN}$	1.00	0.00	1.00	0.00
$(l_N + c_N)$	0.00	0.00	-0.00	0.00
$(I_{NN} - W_{NN})t_{jN}$	-0.06	0.21	-0.08	0.17

Table 3: Average bias and root mean squared error of estimated model parameters (1,000 replications): DGP with approximation error only: $u_{jN} = \eta_{jN}$

		$N = 30$			$N = 60$		
		SILS	CF-OLS	RF-OLS	SILS	CF-OLS	RF-OLS
			$\alpha = -2$				
$\hat{\alpha}_w$	Bias	-0.00	0.00	-1.21	-0.00	-0.00	-0.76
	RMSE	0.00	0.00	2.05	0.00	0.00	1.28
$\hat{\alpha}_t$	Bias	-0.00	-0.00	-0.79	-0.00	-0.00	-0.60
	RMSE	0.00	0.00	1.56	0.00	0.00	1.05
$\hat{\vartheta}$	Bias	–	-150.00	–	–	-150.00	–
	RMSE	–	150.00	–	–	150.00	–
			$\alpha = -4$				
$\hat{\alpha}_w$	Bias	-0.00	0.00	-0.81	-0.00	-0.00	-0.36
	RMSE	0.00	0.00	1.67	0.00	0.00	0.99
$\hat{\alpha}_t$	Bias	-0.00	-0.00	-0.13	-0.00	-0.00	0.08
	RMSE	0.00	0.00	1.80	0.00	0.00	1.24
$\hat{\vartheta}$	Bias	–	-125.00	–	–	-125.00	–
	RMSE	–	125.00	–	–	125.00	–
			$\alpha = -9$				
$\hat{\alpha}_w$	Bias	-0.00	-0.00	-0.60	-0.00	0.00	-0.16
	RMSE	0.00	0.00	1.46	0.00	0.00	0.91
$\hat{\alpha}_t$	Bias	-0.00	0.00	0.40	-0.00	0.00	0.64
	RMSE	0.00	0.00	2.25	0.00	0.00	1.74
$\hat{\vartheta}$	Bias	–	-111.11	–	–	-111.11	–
	RMSE	–	111.11	–	–	111.11	–

Notes: Columns SILS, CF-OLS, and RF-OLS refer to structural iterative least squares, control-function OLS and reduced-form OLS estimates of models (13), (11) and (10) in Section 4.3. Table entries are average biases and root mean squared errors in percent of the true α.

Table 4: Average bias and root mean squared error of estimated model parameters (1,000 replications): DGP with approximation error and stochastic error: $u_{jN} = \eta_{jN} + \varepsilon_{jN}$

		$N = 30$			$N = 60$		
		SILS	CF-OLS	RF-OLS	SILS	CF-OLS	RF-OLS
				$\alpha = -2$			
$\hat{\alpha}_w$	Bias	0.20	-0.00	-1.00	-0.03	-0.18	-0.80
	RMSE	4.13	4.11	4.49	1.70	1.70	2.10
$\hat{\alpha}_t$	Bias	0.01	-0.07	-0.79	0.01	-0.12	-0.59
	RMSE	1.00	1.11	1.76	0.43	0.59	1.13
$\hat{\vartheta}$	Bias	–	-128.50	–	–	-128.95	–
	RMSE	–	129.33	–	–	129.51	–
				$\alpha = -4$			
$\hat{\alpha}_w$	Bias	0.20	0.12	-0.61	-0.05	-0.04	-0.41
	RMSE	3.32	3.30	3.62	1.38	1.38	1.68
$\hat{\alpha}_t$	Bias	0.00	0.33	-0.14	0.00	0.24	0.09
	RMSE	1.04	1.37	1.99	0.45	0.77	1.31
$\hat{\vartheta}$	Bias	–	-114.20	–	–	-114.57	–
	RMSE	–	114.42	–	–	114.72	–
				$\alpha = -9$			
$\hat{\alpha}_w$	Bias	0.15	0.12	-0.45	-0.05	0.03	-0.20
	RMSE	2.94	2.94	3.22	1.22	1.22	1.51
$\hat{\alpha}_t$	Bias	0.00	0.63	0.39	0.00	0.50	0.65
	RMSE	1.06	1.67	2.41	0.46	1.01	1.80
$\hat{\vartheta}$	Bias	–	-106.46	–	–	-106.70	–
	RMSE	–	106.51	–	–	106.73	–

Notes: Columns SILS, CF-OLS, and RF-OLS refer to structural iterative least squares, control-function OLS and reduced-form OLS estimates of models (13), (11) and (10) in Section 4.3. Table entries are average biases and root mean squared errors in percent of the true α.

Table 5: Poisson PML estimators — Average bias and root mean squared error of estimated model parameters (1,000 replications): DGP with approximation error and stochastic error ($u_{jN} = \eta_{jN} + \varepsilon_{jN}$)

		$N = 30$			$N = 60$		
		SIP	CF-Pois	RF-Pois	SIP	CF-Pois	RF-Pois
			$\alpha = -2$				
$\hat{\alpha}_w$	Bias	0.31	-2.15	-9.28	0.02	-1.94	-7.08
	RMSE	6.69	5.70	13.62	2.65	3.00	9.18
$\hat{\alpha}_t$	Bias	0.07	-1.46	-6.00	0.03	-1.32	-4.49
	RMSE	3.36	2.78	7.90	1.13	1.82	5.27
$\hat{\vartheta}$	Bias	–	-129.90	–	–	-130.49	–
	RMSE	–	130.60	–	–	130.95	–
			$\alpha = -4$				
$\hat{\alpha}_w$	Bias	0.28	-1.78	-7.92	-0.01	-1.60	-5.99
	RMSE	5.65	4.61	11.61	2.21	2.46	7.91
$\hat{\alpha}_t$	Bias	0.09	-0.37	-4.05	0.02	-0.46	-2.75
	RMSE	3.83	2.86	7.59	1.24	1.49	4.55
$\hat{\vartheta}$	Bias	–	-115.05	–	–	-115.28	–
	RMSE	–	115.23	–	–	115.41	–
			$\alpha = -9$				
$\hat{\alpha}_w$	Bias	0.24	-1.64	-7.41	-0.02	-1.40	-5.56
	RMSE	5.28	4.12	10.66	1.97	2.17	7.32
$\hat{\alpha}_t$	Bias	0.11	0.39	-2.58	0.02	0.19	-1.36
	RMSE	4.44	3.48	8.08	1.34	1.73	4.80
$\hat{\vartheta}$	Bias	–	-106.92	–	–	-107.02	–
	RMSE	–	106.95	–	–	107.05	–

Notes: Columns SIP, CF-Pois, and RF-Pois refer to structural iterative Poisson PML, control-function Poisson PML and reduced-form Poisson PML estimates of (exponentiated versions of) models (13), (11) and (10) in Section 4.3. An explicit example of such an exponentiated model is given for CF-Pois in (12). Table entries are average biases and root mean squared errors in percent of the true α.

Appendix. Generic gravity models with constant markups

Framework

Arkolakis et al. (2012) demonstrate the generic structure of a host of gravity-model types featuring constant markups, where output prices change in response to trade costs exclusively due to endogenous adjustments in costs but not markups. Models which fall in this class are Dixit-Stiglitz-Krugman type models of monopolistically competitive firms (see, e.g., Bergstrand et al., 2013 for a multi-country gravity version of this type), Ricardian models with perfectly competitive firms (see Eaton and Kortum, 2002), and Armington endowment-economy models (see Anderson and van Wincoop, 2003). All of these models fundamentally adhere to the structure in (1), and what they differ by are only the interpretations of L_i, C_i, and α (the so-called trade elasticity).

While α is directly related to the elasticity of produced varieties—of firms in Dixit-Stiglitz-Krugman models and of countries in Anderson-van-Wincoop models—it measures the production-cost (or productivity) dispersion among the potential producers any country.

L_i is a measure of factor endowments (and firm numbers) in Dixit-Stiglitz-Krugman models, a measure of preference-scaled factor or goods endowments in Anderson-van-Wincoop models, and a measure of average country-level productivity in Eaton-Kortum models. It can generally be obtained when normalizing (dividing) country i's aggregate sales value (in most models GDP) Y_i by C_i.

C_i are the costs per unit of L_i. It can be wages in a Dixit-Stiglitz-Krugman model, where L_i is a country's labor endowment. If L_i is a factor bundle, C_i measures the unit costs of the bundle (e.g., a Cobb-Douglas aggregate of observed factor costs). This is the same in an endowment-economy model, if L_i measures the labor endowment. If L_i is a preference-scaled endowment with goods or a scaled Armington parameter as in Anderson and van Wincoop (2003), C_i measures the unit value or price (per exported unit of good). in Eaton-Kortum models, C_i also measures the variable factor costs per unit of output, as in a Dixit-Stiglitz-Krugman framework.

For all these models, the numerator in the expression of Z_{ij} in equation (1) is log-additive, while the denominator, to be interpreted as the $-\alpha$-scaled log of the ideal consumer price in j, is not log-linear. The latter is in the focus of BEK's linearization. Because all of the mentioned models feature a price index of this form, BEK's linearization is a powerful tool for everyone of them.

All of the mentioned gravity models—with different interpretations of $\{L_i, C_i, \alpha\}$—have the original form for nominal bilateral exports of

$$X_{ij} = \frac{L_i C_i^{\alpha} T_{ij}^{\alpha}}{\sum_{k=1} L_k C_k^{\alpha} T_{kj}^{\alpha}} Y_j. \tag{14}$$

What this equation says is that bilateral purchases or expenditures are proportional to the aggregate level of expenditures in a country. Note that the ratio term appearing in (14) allocates these expenditures across all sources, including the domestic market. After dividing both sides of (14) by Y_j, we arrive at equation (10) in Eaton and Kortum (2002). Moreover, we see that (14) has exactly the structural form given in equations (6) and (7) in Anderson and van Wincoop (2003). And the form is the same as in equation (9) in Bergstrand et al. (2013).[8]

Using $Z_{ij} = X_{ij}/(L_i C_i Y_j)$, we arrive at equation (1) and the corresponding right-hand side of the model given there (we have used the short hand of $L_i C_i = Y_i$, there). This expression is used by Behrens et al. (2012). Log-transforming (1) and linearizing it in the point where $\alpha = 0$, Behrens et al. (2012) arrive at equation (2) in the main text above.

[8] As said, what is called $\{L_i, C_i\}$ here would be $\{T_i c_i\}$ in Eaton and Kortum (2002), it would be $\{\beta_i^{\alpha}, p_i\}$ in Anderson and van Wincoop (2003), and it would be $\{L_i, w_i\}$ with $w_i = Y_i/L_i$ in Bergstrand et al. (2013).