Empir Econ (2016) 50:137–175
DOI 10.1007/s00181-015-0935-x

GLM estimation of trade gravity models with fixed effects

Peter H. Egger · Kevin E. Staub

Received: 2 June 2014 / Accepted: 6 January 2015 / Published online: 26 February 2015

Abstract Many empirical gravity models are now based on generalized linear models (GLM), of which the poisson pseudo-maximum likelihood estimator is a prominent example and the most frequently used estimator. Previous literature on the performance of these estimators has primarily focussed on the role of the variance function for the estimators' behavior. We add to this literature by studying the small sample performance of estimators in a data-generating process that is fully consistent with general equilibrium economic models of international trade. Economic theory suggests that (1) importer- and exporter-specific effects need to be accounted for in estimation, and (2) that they are correlated with bilateral trade costs through general equilibrium (or balance-of-payments) restrictions. We compare the performance of structural estimators, fixed effects estimators, and quasi-differences estimators in such settings, using the GLM approach as a unifying framework.

Keywords Gravity models · Generalized linear models · Fixed effects

The authors gratefully acknowledge helpful comments by Badi Baltagi, Michael Pfaffermayr, João Santos Silva, and two anonymous reviewers on an earlier version of the manuscript. Egger acknowledges funding by Czech Science Fund (GA ČR) through Grant Number P402/12/0982. The authors acknowledge financial support from the Faculty of Business and Economics, University of Melbourne.

P. H. Egger (✉)
Department of Management, Technology, and Economics, ETH Zurich,
Weinbergstr. 35, 8092 Zurich, Switzerland
e-mail: egger@kof.ethz.ch

K. E. Staub
Department of Economics, The University of Melbourne,
111 Barry Street, The University of Melbourne, VIC 3010, Australia
e-mail: kevin.staub@unimelb.edu.au

JEL Classification F14 · C23

1 Introduction

Models of international trade often imply a gravity equation for bilateral international exports of country i to country j (X_{ij}),

$$X_{ij} = E_i M_j T_{ij},$$

where E_i and M_j are exporter-specific factors (a function of goods prices and mass of suppliers) and importer-specific factors (a function of the price index and total expenditures on goods), and T_{ij} are bilateral, pair-specific factors (a function of country-pair-specific consumer preferences and ad-valorem trade costs). Examples include Eaton and Kortum (2002), Anderson and van Wincoop (2003), Baier and Bergstrand (2009), Waugh (2010), Anderson and Yotov (2012), Arkolakis et al (2012), and Bergstrand et al (2013). Baltagi et al (2014) and Head and Mayer (2014) provide recent surveys of this literature.

Writing $e_i = \ln E_i$, $m_j = \ln M_j$, $\mathrm{t}_{ij} = \ln T_{ij}$, and using the parametrization $\mathrm{t}_{ij} = d'_{ij}\beta$, where d_{ij} is a vector of observable bilateral variables and β a conformable vector of parameters, empirical models of the gravity equation follow the general structure

$$\mathrm{E}(X_{ij}|e_i, m_j, d_{ij}) = \exp(e_i + m_j + d'_{ij}\beta). \tag{1}$$

Thus, there are two departures from the original equation. The first relates to the parametrization of the unknown t_{ij} in terms of observables. The linear index structure coupled with the exponential function allows the inclusion of variables in a flexible way, while giving the elements of β the convenient interpretation of direct (semi-) elasticities of exports with respect to the variables in d_{ij} and restricting the domain of exports to be positive. The second departure from the original is that the relationship is stochastic and assumed to hold (only) in expectation. The stochastic formulation implies an error term which makes the relationship between X_{ij} and its specified expectation $\mathrm{E}(X_{ij}|e_i, m_j, d_{ij})$ exact. Thus, the conditional expectation (1) implies two equivalent representations with stochastic errors, either additive or multiplicative:

$$X_{ij} = \exp(e_i + m_j + d'_{ij}\beta) + \varepsilon_{ij} = \exp(e_i + m_j + d'_{ij}\beta)\eta_{ij}, \tag{2}$$

where $\eta_{ij} = 1 + \varepsilon_{ij}\exp(-e_i - m_j - d'_{ij}\beta)$. Together with (1), this implies that $E(\varepsilon_{ij}|e_i, m_j, d_{ij}) = 0$ for the additive error ε_{ij}, and similarly $\mathrm{E}(\eta_{ij}|e_i, m_j, d_{ij}) = 1$ for the multiplicative error η_{ij}, i.e., these errors are mean-independent of the covariates. The goal of this paper is to discuss a number of estimators which are consistent, if model (1) is correctly specified and explore their small-sample properties under varying distributions of η_{ij}. This is a subject which has attracted considerable attention recently, and the body of work studying the performance of estimators of gravity models for trade includes Santos Silva and Tenreyro (2006, 2011), Martin and Pham

(2008), Gómez Herrera (2013), Martínez Zarzoso (2013), and Head and Mayer (2014), among others.

This paper offers three additions to this literature. First, it uses a data-generating process (DGP) which corresponds to a general equilibrium model of international trade as the basis for Monte Carlo simulations. The model of international trade is calibrated to match some features from real-world data. Thus, unlike the previous literature, the performance of estimators is analyzed in a setting with exporter- and importer-specific effects which are correlated with d_{ij}, as suggested by economic theory and as imposed in all applied structural work on gravity models (e.g., in Anderson and van Wincoop 2003, or Bergstrand et al 2013).[1] Second, apart from estimators used by the previous literature, the comparison also includes (1) iterative-structural estimators which fully exploit information from the underlying economic general equilibrium model of international trade, and (2) estimators based on quasi-differenced moment conditions recently proposed by Charbonneau (2012). Third, we provide a unified discussion of the available estimators within the framework of generalized linear models (GLM). As exemplified in an application to a cross section of 94 countries in the year 2008, the GLM framework can be useful for choosing between competing estimators.

Because our focus is on consistent estimators of the gravity Eq. (1), one common estimator which we will ignore is the OLS estimator of the regression with dependent variable $\ln X_{ij}$. The log-linearized OLS estimator will generally deliver inconsistent estimates for β, even if the true model is (1). A lucid exposition of this issue is given in Santos Silva and Tenreyro (2006). Following their illustration, note that the logarithmized equation for bilateral exports is

$$\ln X_{ij} = e_i + m_j + d_{ij}'\beta + \ln \eta_{ij},$$

and OLS estimation will only be consistent if $\mathrm{E}(\ln \eta_{ij}|e_i, m_j, d_{ij}) = 0$. But, as (2) makes clear, $\ln \eta_{ij}$ is a function of the covariates and ε_{ij}, and so in general the condition for consistency of OLS will be violated if the conditional expectation function is exponential as in (1).

In this paper, we examine economic models where the conditional distribution of bilateral trade does not have any mass point at zero. While zero trade flows are common in disaggregated data and in whole-world country-level datasets, an important area of research focusses on trade within more integrated country-blocks where zero trade flows are less prevalent and the number of observations is small to moderate. Our study speaks primarily to this literature.[2] While the estimators discussed in our paper can handle zero trade flows, one may want to rely on alternative models, e.g two-part

[1] In some of the aforementioned work, e.g., in Eaton and Kortum (2002) or Waugh (2010) structural constraints are assumed by the underlying theory but not imposed in estimation. As shown by Fally (2014), estimation with fixed effects is consistent with such structural constraints if adding-up constraints are imposed on trade flows, or the model is estimated by a Poisson regression.

[2] Examples include Eaton and Kortum (2002), Anderson and van Wincoop (2003), Aviat and Coeurdacier (2007), Dekle et al (2007), Baier and Bergstrand (2009), Novy (2013), and Bergstrand et al (2013), among many others. In the applications cited, the number of countries (or other geographical units such as provinces or states) ranges from about 20 to 40 and zero trade flows are rare.

models, for an analysis with mass points in bilateral exports or imports (see Egger et al 2011; Santos Silva et al 2015).

Approaches that estimate Eq. (1) consistently are discussed below in Sect. 2. The Monte Carlo setup and results from the simulation study are presented in Sect. 3, followed by an application with real-world data in Sect. 4. We summarize and discuss our findings in Sect. 5.

2 Estimation of gravity models of bilateral trade

In gravity equations derived from economic models of trade, the exporter- and importer-specific effects e_i and m_j are functions of the bilateral-specific trade costs t_{ij}. This implies that the exporter- and importer-specific terms are, in general, correlated with the bilateral variables d_{ij}. Thus, gravity model estimates omitting these terms—as had been done in most gravity applications in the previous millennium—will be biased in general.

Approaches which do allow for dependence between exporter- and importer-specific effects are structural estimation and fixed effects estimation. Structural estimation explicitly specifies the exporter- and importer-specific terms as a function of the economic model's variables, accruing to adding-up or resource constraints. In contrast, the fixed effects approach is to leave the relationship between d_{ij} and the exporter- and importer-specific terms –the fixed effects– unspecified and is thus consistent regardless of the nature of their dependence. Within this approach, there are two ways of handling the fixed effects. The first way is to estimate β from moment conditions which are derived from (1) but which do not depend on the fixed effects. The second way treats the fixed effects as parameters to be estimated. Although both fall into the fixed effects domain, for convenience we will from now on refer to estimators under the first approach as *quasi-differences estimators* and to the ones under the latter approach as *fixed effects estimators*.

Structural estimators—i.e., ones that impose the nonlinear gravity model structure in estimation rather than utilizing fixed country effects—fully exploit the information on the data-generating process, estimate fewer parameters, and are thus potentially more efficient than other approaches such as fixed country effects estimators. For instance, Eaton and Kortum (2002), Waugh (2010), Egger et al (2011), or Egger et al (2012) do not impose the model structure in estimation but employ fixed effects. In contrast to that, Anderson and van Wincoop (2003), Bergstrand et al (2013), or Egger and Nigai (2014) estimate structural models where non-linear price (general equilibrium) terms are solved iteratively in estimation. To be consistent, iterative-structural approaches require the underlying economic model to be correctly specified. On the other hand, quasi-differences estimators and fixed effects estimators are consistent under much weaker assumptions. The quasi-differences approach has the feature of avoiding estimating a potentially large number of parameters. This can be an advantage because estimation of the set of fixed effects can be both computationally difficult and lead to less precise estimates. On the other hand, some objects of interest (such as predictions and average marginal effects) are functions not only of β, but also of e_i and m_j. Without having estimates of these fixed effects, it may be impossible to recover estimates for such objects of interest. Moreover, since fixed effects may capture structural model variables but also factors outside of the model, the associated results are harder to interpret than iterated structural model estimates.

An important drawback of fixed effects estimators of nonlinear models is that they suffer from the incidental parameters problem. In particular, the trade gravity equation is subject to the "large-N large-T"-version of the incidental parameters problem that invalidates inference based on the asymptotic distribution (see Hahn and Newey 2004; Fernández-Val and Weidner 2013). Our Monte Carlo results of Sect. 3 point in this direction, showing that t-statistics obtained from fixed effects estimators are not centered around zero even when β is virtually unbiased.[3]

We begin by discussing the GLM approach (Sect. 2.1), which can be used to estimate a variety of fixed effects and structural estimators, and in principle even serve as a "wrapper" for some quasi-differenced moment conditions. To fix ideas, we explain the GLM approach in the context of fixed effects estimation. Then, we discuss structural and quasi-differences estimation in some more detail in Sects. 2.2 and 2.3.

2.1 GLM estimation of gravity models with fixed effects

Nonlinear models with a linear index, such as (1), can be estimated consistently by a host of so-called generalized linear model (GLM) estimators. These models have a common structure, as is well-known and outlined briefly in this section. The GLM framework (Nelder and Wedderburn 1972; McCullagh and Nelder 1989) is a likelihood-based approach that remains consistent for the parameters of interest as long as the conditional expectation function is correctly specified. It is widely used to estimate nonlinear models in statistics (see for instance Adams et al 2004, and references therein), and many econometric applications can be cast in the GLM framework (for examples in international economics, see Santos Silva and Tenreyro 2006; Bosquet and Boulhol 2009; Anderson and Yotov 2010, 2012; or Mélitz and Toubal 2014).

GLM is based on likelihood estimation of potentially misspecified densities of the linear exponential family (LEF) of distributions, which includes the Poisson, Negative Binomial, Gamma, Normal, and other distributions. A GLM estimator is determined by (1) the specification of a *link* function—which governs the relationship between the conditional expectation function and the linear index of covariates—together with (2) the specification of a *family*, a density from the LEF.

Let us view e_i and m_j under the fixed effects approach, where the fixed effects are parameters to be estimated. The distribution of X_{ij} is part of the LEF class if its density can be written as

[3] This is different from the better-known "fixed-T large-N"-version of the incidental parameters problem which constitutes an inability to estimate the so-called nuisance parameters (the fixed effects in this context) consistently—an inconsistency that passes over to the "common parameters" (β in this context). In the gravity equation setup, both dimensions (numbers of exporter and importers) increase as the number of countries C increases. For every additional country, there are $2 \times (C - 1)$ additional observations but only 2 additional parameters (e_C and m_C). Thus, all parameters are estimated with less bias as C increases. The source of the incidental parameters problem here is the fact that the rate of convergence for the "common parameters" e_i and m_j is slower.

$$f(X_{ij}) = \exp\left\{\frac{X_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(\phi, X_{ij})\right\},$$

with parameters $\theta_{ij} = (\beta', e_i, m_j)'$, ϕ (which we will later show to be related to the variance of X_{ij}, denoted as $V(X_{ij})$), and arbitrary functions $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$. Such a random variable will have the log-likelihood

$$l(\theta_{ij}) = \frac{X_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(\phi, X_{ij}),$$

and score

$$s(\theta_{ij}) = \frac{\partial l(\theta_{ij})}{\partial \theta_{ij}} = \frac{X_{ij} - b'(\theta_{ij})}{a(\phi)}.$$

A fundamental result of likelihood theory is that the expected score evaluated at the true value of the parameter is zero. For members of the LEF, this implies

$$\mathrm{E}[s(\theta_{ij})] = \frac{\mathrm{E}(X_{ij}) - b'(\theta_{ij})}{a(\phi)} = 0,$$

and, therefore, $\mathrm{E}(X_{ij}) = b'(\theta_{ij}) \equiv \mu_{ij}$. In GLM estimation, the mean of X_{ij}, μ_{ij}, is specified as a function of a linear index, in our context $d'_{ij}\beta + e_i + m_j$, so that $\mu_{ij} = g^{-1}(d'_{ij}\beta + e_i + m_j)$, or $g(\mu_{ij}) = d'_{ij}\beta + e_i + m_j$. This function, $g(\cdot)$, is called the link function. The gravity model (1), consequently, implies a *log-link*.

A second result from likelihood theory, the information matrix equality, allows to derive the variance function. The Hessian of an LEF random variable is

$$H(\theta_{ij}) = \frac{\partial^2 l(\theta_{ij})}{\partial \theta_{ij}^2} = \frac{-b''(\theta_{ij})}{a(\phi)}.$$

By the information matrix equality, $V[s(\theta)] = -\mathrm{E}[H(\theta)]$, and therefore

$$\frac{\mathrm{E}[X_{ij} - b'(\theta_{ij})]^2}{a(\phi)^2} = \frac{b''(\theta_{ij})}{a(\phi)},$$

which is equivalent to $V(X_{ij}) = b''(\theta_{ij})a(\phi)$.

The parameter vector θ is estimated by maximizing the sample (quasi-)likelihood function

$$l_C(\beta) = \sum_{i=1}^{C}\sum_{j=1}^{C} \frac{X_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(\phi, X_{ij}),$$

with corresponding score for β

Table 1 Variance functions $V(X_{ij})$ and F.O.C. weights w_{ij} of some GLM gravity estimators

Estimator	$V(X_{ij})$	w_{ij}
Poisson	μ_{ij}	1
Gamma	μ_{ij}^2	μ_{ij}^{-1}
Negative Binomial	$\mu_{ij} + \mu_{ij}^2$	$(1 + \mu_{ij})^{-1}$
Gaussian	1	μ_{ij}
Inverse Gaussian	μ_{ij}^3	μ_{ij}^{-2}

$$s_C(\beta) = \sum_{i=1}^{C}\sum_{j=1}^{C} \frac{X_{ij} - b'(\theta_{ij})}{a(\phi)} \frac{\partial \theta_{ij}}{\partial \beta} = \frac{X_{ij} - b'(\theta_{ij})}{a(\phi)} \frac{\partial \theta_{ij}}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \beta} = \frac{X_{ij} - b'(\theta_{ij})}{b''(\theta_{ij})a(\phi)} \frac{\partial \mu_{ij}}{\partial \beta}.$$

Analogous equations hold for e_i and m_j. The last equality follows from $\mu_{ij} = b'(\theta_{ij})$. Thus, the first-order conditions $s_N(\beta) = 0$ can be written as

$$\sum_{i=1}^{C}\sum_{j=1}^{C} \frac{\left[X_{ij} - \mathrm{E}(X_{ij})\right]}{V(X_{ij})} \frac{\partial \mathrm{E}(X_{ij})}{\partial \beta} = 0. \tag{3}$$

Equation (3) shows that GLM estimators are moment estimators based on the conditional expectation residual (the expression in square brackets). As long as the expectation function is correctly specified, they will be consistent for β. Equation (3) also shows that GLM estimators weight the residuals in the first-order conditions by the inverse of the variance, $V(X_{ij})$. Therefore, relative efficiency gains – but not consistency of β – depend on the correct specification of the variance function. Different choices of the LEF distribution (i.e., of the *family*) lead to different variance functions.

As mentioned above, gravity models use the log-link function, so that in general the first-order conditions for β in models based on (1) have the form

$$\sum_{i=1}^{C}\sum_{j=1}^{C} \frac{\left[X_{ij} - \exp(e_i + m_j + d_{ij}'\beta)\right]}{V(X_{ij})} \exp(e_i + m_j + d_{ij}'\beta) d_{ij} = 0. \tag{4}$$

The first-order conditions for the fixed effects are given in the Appendix.

Various distributions are suitable candidates for gravity models in the sense that they are members of the LEF and that they imply variance functions $V(X_{ij})$ that are potentially compatible with bilateral export flows. Poisson and Gamma estimators have been discussed in Santos Silva and Tenreyro (2006) and others; estimators based on Gaussian, inverse Gaussian, and Negative Binomial quasi-likelihood are also possible and have been used in the literature, though much less so than Poisson and Gamma estimators. The implied variance functions, in terms of the conditional expectation function μ_{ij}, are given in Table 1. Because of the exponential function, the derivative $\partial \mathrm{E}(X_{ij})/\partial \beta$ is $\mu_{ij} d_{ij}$. The column on the right of the table reports the weight given in the first-order conditions to the residual of country-pair ij, i.e., $w_{ij} = \mu_{ij} d_{ij} / V(X_{ij})$.

Thus, as pointed out by Santos Silva and Tenreyro (2006), the Poisson estimator weights every observation equally, which might be optimal when wanting to remain agnostic with respect to higher-order specification. The other estimators in Table 1, with the exception of the Gaussian one, down-weight observations with larger means. If these observations are also noisier in the sense of having a larger variance, this would result in efficiency gains. In contrast, the Gaussian estimator gives more weight to country-pairs with larger expected exports, a weighting scheme which might be beneficial, for instance, if there are trustworthiness issues with small trade flows.[4] The Gaussian-GLM gravity estimator is equivalent to the nonlinear least squares (NLS) estimator of (1).[5]

2.2 Iterative-structural estimation

A structural approach to estimating the gravity equation makes use of economic theory. Most modern gravity models assume resource constraints of the form

$$E\left[\sum_{j=1}^{C} X_{ij}\right] = E\left[\sum_{j=1}^{C} \exp(e_i + m_j + d_{ij}'\beta)\right] = E\left[\exp(e_i)\sum_{j=1}^{C} \exp(m_j + d_{ij}'\beta)\right] \tag{5}$$

and

$$E\left[\sum_{i=1}^{C} X_{ij}\right] = E\left[\sum_{i=1}^{C} \exp(e_i + m_j + d_{ij}'\beta)\right] = E\left[\exp(m_j)\sum_{i=1}^{C} \exp(e_i + d_{ij}'\beta)\right]. \tag{6}$$

Hence, (5) and (6) imply that the country-specific effects e_i and m_j are implicitly fully determined given data on X_{ij} and d_{ij} together with estimates of β. Clearly, this implies that $E[e_i d_{ij}] \neq 0$, $E[m_j d_{ij}] \neq 0$, and $E[e_i m_j] \neq 0$ according to economic theory.[6]

In our Monte Carlo study, we will employ data-generating processes that respect this feature. Specifically, we will consider an endowment economy. Many data-generating processes considered in the literature are isomorphic to the one of endowment economies (see Anderson and van Wincoop 2003, for such an approach; and Eaton and Kortum 2002; Arkolakis et al 2012; Bergstrand et al 2013, for isomorphic processes).

[4] Such trade flows may emerge particularly with or among less developed countries, where reporting standards are poor. For instance, Egger and Nigai (2014) illustrate that structural parameterized GLM gravity models tend to predict large bilateral trade flows with much smaller error than small trade flows.

[5] The Negative Binomial estimator is only an LEF member *for a given value* of the overdispersion parameter, say α in $\mu_{ij} + \alpha\mu_{ij}^2$. In Table 1 and in what follows, we consider the Negative Binomial GLM estimator with overdispersion parameter fixed at 1.

[6] As a consequence, omitting $e_i + m_j$ from the specification in (2) or simply replacing it by a log-additive function of GDP and/or GDP per capita of countries i and j, as had been done for decades in a-theoretical empirical gravity models, will generally lead to inconsistent estimates of β, the semi-elasticity of trade with respect to d_{ij}.

Suppose an economy i is endowed with a production volume of H_i which it sells at a mill price of P_i, earning it an aggregate income (GDP) of $Y_i = P_i H_i$. In general, sales to market j are impeded by some trade costs, for which we use the notation T_{ij} as in the introduction. Then, in an endowment economy with Armington differentiation (between the goods from different countries) at an elasticity of $1-\alpha$, nominal aggregate bilateral trade flows in a world akin to the one in Anderson and van Wincoop (2003) are determined as[7]

$$\mathrm{E}(X_{ij}) = \frac{P_i^{\alpha} T_{ij} Y_j}{\sum_{k=1}^{C} P_k^{\alpha} T_{kj}}, \tag{7}$$

where, at given H_i and T_{ij}, P_i is implicitly determined from the resource constraint as generically introduced in (5) and (6), which in the present context becomes

$$Y_i = P_i H_i = \sum_{j=1}^{C} \frac{P_i^{\alpha} T_{ij} P_j H_j}{\sum_{k=1}^{C} P_k^{\alpha} T_{kj}}, \tag{8}$$

$$\Rightarrow P_i^{1-\alpha} = \frac{1}{H_i} \sum_{j=1}^{C} \frac{T_{ij} P_j H_j}{\sum_{k=1}^{C} P_k^{\alpha} T_{kj}}. \tag{9}$$

This is exactly the data-generating process which we will use for the endogenous variables in the model, $\{\mathrm{E}(X_{ij}), P_i, Y_i\}$, before adding a stochastic term to (7). In terms of the gravity Eq. (1), $\ln T_{ij} = d_{ij}'\beta$, $\ln P_i^{\alpha} = e_i$ and $\ln \frac{P_j H_j}{\sum_{k=1}^{C} P_k^{\alpha} T_{kj}} = m_j$.[8]

Assuming that data on GDP of each country i (Y_i) are available, a structural estimator for β, e_i, m_j ($i = 1, \ldots, C$, $j = 1, \ldots, C$) can be defined as the solution obtained by iterating between the following Step 1 and Step 2 until convergence:

Step 0 **Initial values:**
Set initial values for e_i, m_j ($i = 1, \ldots, C$, $j = 1, \ldots, C$).
Call vectors of these values $\{\hat{e}^{(n)}, \hat{m}^{(n)}\}$.

Step 1 **Update $\hat{\boldsymbol{\beta}}$:**
Given current values $\{\hat{e}^{(n)}, \hat{m}^{(n)}\}$, generate the auxiliary variable $\widehat{em}_{ij}^{(n)} = \hat{e}_i^{(n)} + \hat{m}_j^{(n)}$.
Estimate β by GLM regression of X_{ij} on d_{ij} and $\widehat{em}_{ij}^{(n)}$, under the constraint that the coefficient corresponding to $\widehat{em}_{ij}^{(n)}$ is set to 1.
The resulting estimator of β is called $\hat{\beta}^{(n)}$.

[7] In the interest of brevity, we skip the exporter-specific preference parameter introduced in Anderson and van Wincoop (2003) without loss of generality.

[8] For DGPs of the type (5)–(9), fixed effects estimates of m_j and e_i could be used to obtain estimates of $P^{1-\alpha}$ and Y_i (see Fally 2014, Lemma 1A).

Step 2 **Update** $\{\hat{e}, \hat{m}\}$:

Given current estimates $\hat{\beta}^{(n)}$, $\{\hat{e}^{(n)}, \hat{m}^{(n)}\}$, GDP $\{Y_i\}_{i=1}^{C}$, and α, obtain $T_{ij}^{(n)} = \exp\left\{d_{ij}'\hat{\beta}^{(n)}(1-\alpha)\right\}$, $\hat{e}_i^{(n+1)} = \ln E_i^{(n+1)}$ and $\hat{m}_j^{(n+1)} = \ln M_j^{(n+1)}$, where $E_i^{(n+1)} = \frac{Y_i}{\sum_j \frac{T_{ij}^{(n)} Y_j}{\sum_k E_k^{(n)} T_{kj}^{(n)}}}$ and $M_j^{(n+1)} = \frac{Y_j}{\sum_i E_i^{(n)} T_{ij}^{(n)}}$.

After Step 2, $(n+1)$ is redefined as (n) and Step 1 is repeated. Convergence criteria can be defined by norms on the vector of differences $\left(\hat{\beta}^{(n+1)'} - \hat{\beta}^{(n)'}, \hat{e}^{(n+1)'} - \hat{e}^{(n)'}, \hat{m}^{(n+1)'} - \hat{m}^{(n)'}\right)$. As in the fixed effects approach, different iterative-structural estimators are defined by the choice of the respective family of distributions for the GLM regression in Step 1.[9] Except for unrealistic trade cost configurations, trade models of the form analyzed here have a unique solution, which can be represented by a contraction mapping (see, e.g., Allen et al 2014).

2.3 Quasi-differences estimation

The approaches in the previous sections estimate the gravity model by either viewing the fixed effects as additional parameters to be jointly estimated with β or by obtaining them from the solution to an economic system of resource constraints. An advantage of these approaches is that the estimated fixed effects can be used to calculate general equilibrium comparative statics (see, e.g., Fally 2014, who uses a Poisson estimation with fixed effects to recover parameters of an Anderson and van Wincoop 2003, model). Compared to the structural estimation, the approach of estimating the fixed effects allows to account for additional unobserved importer- and exporter-specific trade costs not part of the structural importer- and exporter-specific terms (see, e.g., Egger et al 2012, for an example of such an empirical approach).

An alternative approach is to transform the model to obtain a moment condition that does not depend on the fixed effects at all. This has the advantage that the number of parameters to be estimated is reduced substantially which can lead to considerable efficiency gains in the estimation of β. Since β_k (the kth element of β) has the interpretation of the ceteris paribus or partial equilibrium semi-elasticity of exports with respect to the bilateral trade cost $d_{ij,k}$, the vector β is often a first-order object of interest in gravity estimation.

In linear models, differencing observations with respect to i and j yields a fixed-effects-free expression (see for instance Baltagi 2013, or Cameron and Trivedi 2005). In nonlinear models with an exponential mean and one-way fixed effects, a quasi-differencing analog exists: fixed effects can be eliminated by using i-specific ratios of observations. Charbonneau (2012) introduced a quasi-differencing approach for exponential mean models with two-way fixed effects, deriving a moment condition from (1) which does not depend on either e_i or m_j.

[9] Structural estimators based on log-linearization, such as structurally iterated least squares (Head and Mayer 2014) suffer from the same problems as OLS: for instance, in general they are inconsistent if the errors are heteroskedastic. In contrast to the algorithm presented here, structurally iterated least squares also requires data on a country's trade and trade costs with itself.

In the context of the gravity model (1), this moment condition requires sets s of 4 country-pairs involving two exporters (i, l) and two importers (k, j). The two products of exports,

$$X_{ik}X_{lj} = \exp\left[(e_i + e_l) + (m_k + m_j) + (d_{ik} + d_{lj})'\beta\right] + (\varepsilon_{ik} + \varepsilon_{lj}), \quad (10)$$
$$X_{lk}X_{ij} = \exp\left[(e_i + e_l) + (m_k + m_j) + (d_{lk} + d_{ij})'\beta\right] + (\varepsilon_{lk} + \varepsilon_{ij}), \quad (11)$$

have the same sums of exporter- and importer-specific terms. Dividing by the bilateral-specific part and taking conditional expectations yields

$$E\left\{X_{ik}X_{lj}\exp\left[-(d_{ik} + d_{lj})'\beta\right]|\cdot\right\} = \exp\left[(e_i + e_l) + (m_k + m_j)\right], \quad (12)$$
$$E\left\{X_{lk}X_{ij}\exp\left[-(d_{lk} + d_{ij})'\beta\right]|\cdot\right\} = \exp\left[(e_i + e_l) + (m_k + m_j)\right], \quad (13)$$

where the conditioning variables have been suppressed to avoid cluttered notation.[10] Since the expectations of these terms are the same, a method of moments estimator can be based, for instance, on the difference (12) minus (13). Specifically, Charbonneau (2012) proposes to use the just-identified set of unconditional moment conditions

$$E\left\{\left[X_{ik}X_{lj} - X_{lk}X_{ij}\exp(d_{ik} + d_{lj} - d_{lk} - d_{ij})'\beta\right](d_{ik} + d_{lj} - d_{lk} - d_{ij})\right\} = 0 \quad (14)$$

as a basis for GMM estimation.

This approach is related to ratio-of-ratios estimators of trade (cf. Head and Mayer 2014, pp. 152–153 and references therein). To see this, define $X_s \equiv X_{ik}X_{lj}/X_{lk}X_{ij}$ and $d_s = (d_{ik} + d_{lj}) - (d_{lk} + d_{ij})$ to obtain

$$E(X_s|d_s) = \exp(d_s'\beta). \quad (15)$$

The quasi-differenced model (15) lends itself to estimation by any of the GLM estimators discussed before.[11] In comparison with (15), estimation based on (14) has the advantages that it is applicable even in case the data contain observations where bilateral export flows are zero, and, more generally, that it does not rely on ratios.

Even if the country-pair-specific errors ε_{ij} are independent, the corresponding errors $\varepsilon_s = X_s - E(X_s|d_s)$ will exhibit a mechanical dependence structure as a result of overlaps in exporter and importer countries in two observations s and s'. Therefore, an appropriate cluster-robust variance estimator should be adopted when using estimators based on (14) or (15).

The number of possible sets of countries s is very large, so much that for the number of countries typically used in applications, it might cause computational difficulties. Sacrificing efficiency, one might only select a subset of all possible sets s. We only use sets for which $l = i+1$ and $j < k \leq i+C-2$. This still increases the observations by an order of magnitude. To get a perspective of the number of sets, for 10 countries (90

[10] The conditioning variables are $d_{ik}, d_{li}, e_i, e_l, m_k, m_j$ in the first equation, and the corresponding variables in the second.

[11] As with the standard gravity equation, the previous literature proceeded by log-linearization and OLS estimation, which is subject to the same problems as discussed in Sect. 1.

country-pairs) this results in 245 sets; for 50 countries (2,450 country-pairs) it results in 55,225 sets.

3 Monte Carlo experiments

If the data are generated from the gravity Eq. (1) and according to an underlying economic model satisfying Eqs. (5), (6), (7), and (9), all estimators discussed in Sect. 2 are consistent. To explore the performance of the estimators in finite samples, we set up a Monte Carlo experiment. The questions that we seek to explore are how the relative performance of the estimators is affected by features of the economic model, by the distribution of the stochastic errors, and by an increase in the number of countries.

3.1 Data-generating process

The data-generating process (DGP) consists of a *structural part*, in which the bilateral-, exporter-, and importer-specific determinants of exports are drawn; and a *stochastic part*, in which random errors are drawn and joined to the structural part of exports. In terms of the gravity equation, the structural part is the conditional expectation function given in (1) (denoted by μ_{ij}), and the stochastic part is ε_{ij}.

3.1.1 Structural economic model

The structural part of the DGP follows the economic model (5)–(9). Given

(1) a value of the substitution elasticity α,
(2) endowments H_i for every country ($i = 1, \ldots, C$), and
(3) bilateral trade cost factors T_{ij} for every country-pair ij,

the model implicitly determines GDP Y_i, prices P_i, and structural exports $\mathrm{E}(X_{ij}) = \mu_{ij}$ through general equilibrium constraints. Thus, these variables are determined by α and the joint distribution of H_i and T_{ij}.

Broadly in line with Anderson and van Wincoop (2003) and a host of other estimates in the literature, the parameter α is set at the value -4 as a baseline. The sensitivity of the gravity estimators to this parameter is explored by varying α to -9 in an alternative scenario. Ceteris paribus, a higher elasticity of substitution between goods (higher absolute value of α) leads to a higher variability of exports, while endowments remain constant.

The parameters H_i and T_{ij} are drawn in a way which is consistent with economic theory. In particular, we set intra-national trade frictions to zero (see Eaton and Kortum 2002; or Anderson and van Wincoop 2003), whereby $T_{ii} = 1$ for all i. For $j \neq i$, $T_{ij} \in (0, 1]$. First, two auxiliary bivariate normal variables are drawn:[12]

[12] The first and second moments of these variables are broadly in line with the data used in the application in Sect. 4 below.

$$z_{ij}^{H}, z_{ij}^{T} \sim N(\mu_z, \Sigma_z), \quad \mu_z = (3, -2)',$$
$$\sigma_z = \upsilon_{HT} \begin{pmatrix} (3\sqrt{C}/4)^2 & 0.95 \times 15\sqrt{C}/4 \\ 0.95 \times 15\sqrt{C}/4 & 25 \end{pmatrix}.$$

Then, the parameters of interest are obtained as

$$H_i = \exp\left(\sum_j z_{ij}^{H}/C\right) C > 0 \quad \text{and} \quad T_{ij} = \left(\frac{\exp(z_{ij}^{T})}{1 + \exp(z_{ij}^{T})}\right)^{-\alpha/4} \in (0, 1).$$

The number of countries C is incorporated in H_i (and z^H) such that the share of a country's endowment in world endowment remains constant as C increases. With a baseline of 10 countries, the correlation between H_i and T_{ij} is about 25 % and weakens to 12 % as C increases to 50. Endowments H_i have a mean of about 200.

The parameter υ_{HT} is a scaling constant for the variance of H_i and T_{ij} which is set equal to 0.1 at baseline. This produces a variance of about 2,200. In an alternative scenario, we increase the cross-country inequality in endowments by setting $\upsilon_{HT} = 0.3$ which increases the variance to about 7,800. Similar to the increase in $|\alpha|$ as discussed above, more inequality in endowments (higher υ_{HT}) leads to more variability in exports as well. However, it also leads to higher average exports. The resulting coefficient of variation of exports is even slightly lower (1.37) than in the baseline (1.5), whereas the increase in $|\alpha|$ is associated with a significantly higher coefficient of variation (2.0). Thus, the challenge of the experiment of increasing $|\alpha|$ lies in the increased dispersion of exports, whereas the challenge of increasing υ_{HT} lies in the greater variance of the fixed effects.

Finally, we parameterize T_{ij} by a single observable trade cost parameter d_{ij} for each ij as

$$\ln T_{ij} = \beta_0 + \beta_1 d_{ij}, \tag{16}$$

with $\beta_0 = 0$ and $\beta_1 = 1$.

3.1.2 Stochastic shocks

Observed exports are obtained by joining stochastic shocks with the mean of exports, μ_{ij}, as determined by the structural model: $X_{ij} = \mu_{ij}\eta_{ij}$. Provided the errors η_{ij} are mean-independent, $\mathrm{E}(\eta_{ij}) = 1$, all considered estimators are consistent. Their asymptotic efficiency depends on the variance of η_{ij}. We consider six different variance functions for η_{ij}. Five of these imply asymptotic efficiency of the Poisson, Gamma, Negative Binomial, Gaussian, and Inverse Gaussian estimators, respectively. The sixth variance function of η_{ij} is not optimal for any of these five estimators.

The errors η_{ij} are drawn from a heteroskedastic log-normal distribution

$$\eta_{ij} = \exp z_{ij}^{\eta}, \quad z_{ij}^{\eta} \sim N(-0.5\sigma_\eta^2, \sigma_\eta^2).$$

Table 2 Variance functions in Monte Carlo simulations

Number	σ^2_η	$V(\eta_{ij})$	$V(X_{ij})$	Optimal estimator
1	$\ln(2)$	1	μ^2_{ij}	Gamma
2	$\ln(1+\mu^{-1}_{ij})$	μ^{-1}_{ij}	μ_{ij}	Poisson
3	$\ln(2+\mu^{-1}_{ij})$	$1+\mu^{-1}_{ij}$	$\mu_{ij}+\mu^2_{ij}$	NegBin
4	$\ln(1+\mu^{-2}_{ij})$	μ^{-2}_{ij}	μ^{-2}_{ij}	Gaussian
5	$\ln(1+\mu_{ij})$	μ_{ij}	μ^3_{ij}	Inv. Gaussian
6	$\ln(1+\mu^{-1.5}_{ij})$	$\mu^{-1.5}_{ij}$	$\sqrt{\mu_{ij}}$	Other

Hence, $E(\eta_{ij}) = 1$ and $V(\eta_{ij}) = \exp(\sigma^2_\eta) - 1$. Since $V(X_{ij}|d_{ij}, e_i, m_j) = \mu^2_{ij}\sigma^2_\eta$, we determine σ^2_η as detailed in Table 2. The term "Optimal estimator" in the table refers to asymptotic efficiency. This does not exclude the possibility of poor small-sample performance. While the sixth (dubbed "Other") variance function, we consider does not correspond to anyone of the discussed estimators, we note that among the estimators of Table 2 it is 'closest' to the one of Poisson for large values of μ_{ij}.

In every case, we rescale the errors η_{ij} to achieve a constant pseudo-$R^2 = V(\mu_{ij})/[V(\mu_{ij}) + V(\varepsilon_{ij})]$ of about 50 % for all scenarios and variance functions. The errors ε_{ij} in this pseudo R^2 are the implicit additive errors $X_{ij} - \mu_{ij}$.

3.1.3 Specifications and estimators

For estimation, we consider four alternative specifications:

$$\text{S1: } X_{ij} = \exp(\beta_{10} + \beta_{11}d_{ij} + e_i + m_j) + \varepsilon_{1ij}, \tag{17}$$

$$\text{S2: } X_{ij} = \exp\left(\beta_{20} + \beta_{21}d_{ij} + \sum_{k=2}^{C} e_{2k}D_{ki} + \sum_{k=2}^{C} m_{2k}D_{kj}\right) + \varepsilon_{2ij}, \tag{18}$$

$$\text{S3: } X_{ij} = \exp(\beta_{30} + \beta_{31}d_{ij} + \beta_{32}y_i + \beta_{33}y_j) + \varepsilon_{3ij}, \tag{19}$$

$$\text{S4: } X_{ij} = \exp(\beta_{40} + \beta_{41}d_{ij}) + \varepsilon_{4ij}. \tag{20}$$

All specifications include a constant which is denoted by $\{\beta_{10}, \beta_{20}, \beta_{30}, \beta_{40}\}$. The object of interest is β_1 from (16) which here is estimated by the coefficients on d_{ij}, $\{\beta_{11}, \beta_{21}, \beta_{31}, \beta_{41}\}$. *S1* is a structural model which includes the true terms of e_i and m_j with constrained unitary parameters on them, in line with economic theory. The true unknown parameter β_{11} is unity but allowed to be estimated differently from that. *S2* is a two-way fixed effects model which estimates e_{2i} and m_{2j} by country-specific constants through exporter and importer indicator variables $D_{ki} = \mathbf{1}(i = k)$ and $D_{kj} = \mathbf{1}(j = k)$, and the true parameter β_{21} is unity. Hence, *S2* estimates $e_i + m_j$ by $2(C - 1)$ constants rather than by structural constraints as in *S1* which is less efficient than *S1*. *S3* is an old-fashioned gravity model—akin to the ones that had been estimated prior to Eaton and Kortum (2002) and Anderson and van Wincoop (2003)—which replaces $\{e_i, m_j\}$ by log exporter and importer GDP, $\{y_i, y_j\}$, and estimates parameters $\{\beta_{32}, \beta_{33}\}$ on them—about which we do not have priors. Clearly, this model is misspecified, as $\{y_i, t_{ij}, y_j\}$ are correlated with ε_{3ij}. However, given its vast application in the past, it is interesting to see how this model fares in a laboratory experiment relative to *S1* and *S2*. Finally, *S4* is an ad hoc gravity model which only includes t_{ij} apart from a constant. For the same reason as *S3*, this model is misspecified, as t_{ij} is correlated with ε_{4ij}. Notice that while both *S3* and *S4* are misspecified, the problem that $E[e_i d_{ij}] \neq 0$ and $E[m_j d_{ij}] \neq 0$ vanishes as the number of countries C grows. Since the number of countries is large empirically, the bias of β due to omitting $e_i + m_j$ should be small in theory (even though countries vary a lot in terms of their size).[13] This issue is commonly disregarded. Finally, we also estimate two quasi-differences models, based on Eqs. (14) and (15).

For every specification $S1$–$S4$, we use the five GLM estimators based on the Gamma, Poisson, NegBin, Gaussian, and Inverse Gaussian LEF family. In our baseline scenario $\alpha = -4$, $v_{HT} = 0.1$, and $C = 10$. Since we disregard intra-national trade in estimation (as is commonly the case in applied work), $C = 10$ results in 90 observations. We consider three alternative scenarios: a higher value of $|\alpha|$ with $\alpha = -9$, a higher value of v_{HT} with $v_{HT} = 0.3$, and a larger number of countries with $C = 50$ (2,450 observations). In each alternative scenario, we leave the two remaining parameters of $\{\alpha, v_{HT}, C\}$ at baseline values. Results of a fourth alternative with $\alpha = -2$ are displayed in Table 10 in the Appendix.

3.2 Simulation results

All statistics presented throughout this subsection are based on 1,000 replications of the DGP.

3.2.1 Baseline scenario $\alpha = -4$, $v_{HT} = 0.1$, $C = 10$

Table 3 presents our main results corresponding to iterative-structural (IS) (dubbed *S1* above), fixed effects (FE), and quasi-differences (QD) estimation (FE and QD were dubbed *S2* above). Estimators are grouped in columns. Six panels present statistics for the estimated β_1 under the six different variance functions of the errors in the DGP: the number of convergences achieved out of 1,000 runs (CR), the mean of $\hat{\beta}_1$ over converged estimations (Mean), the standard deviation (SD), median (Med), and the 5th and 95th percentile of the distribution of estimates.

A glance at the results for the IS estimators reveals that despite the small sample size of only 90 country-pairs, exploiting all available economic structure of the DGP results in extremely precise estimates. All estimators are virtually unbiased (mean equal to 1) and display standard deviations which more often than not are smaller than the two decimal places we report in the table. The excellent performance of the

[13] To see this, note that as the number of countries grows to infinity countries' consumer and producer prices will be entirely determined by the rest of the world. Hence, for any arbitrary pair of countries, the impact of bilateral trade costs on price indices, producer prices, or factor costs will be infinitesimally small. The correlation between bilateral trade costs and any multilateral variable (such as gross domestic product, wages, or prices) approaches zero as the number of countries grows large.

Table 3 Baseline Monte Carlo results, 1,000 replications

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
Variance function 1 (Gamma)												
CR	1,000	1,000	1,000	909	988	1,000	1,000	1,000	992	999	1,000	1,000
Mean	1.00	1.00	1.00	1.01	1.00	1.00	1.00	0.99	1.55	1.26	0.99	0.93
SD	0.00	0.01	0.01	0.03	0.02	0.07	0.12	0.08	1.03	0.23	0.11	0.22
Med	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	1.31	1.21	0.99	0.89
5th	1.00	1.00	1.00	1.00	1.00	0.88	0.82	0.85	0.85	0.99	0.81	0.66
95th	1.00	1.00	1.00	1.08	1.00	1.11	1.21	1.13	2.98	1.67	1.18	1.33
Variance function 2 (Poisson)												
CR	1,000	1,000	1,000	998	818	1,000	1,000	1,000	1,000	969	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00	1.46	1.00	1.19
SD	0.00	0.00	0.00	0.01	0.02	0.05	0.01	0.01	0.01	0.40	0.05	0.36
Med	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00	1.36	1.01	1.11
5th	1.00	1.00	1.00	1.00	1.00	0.95	0.98	0.96	0.98	1.04	0.93	0.84
95th	1.00	1.00	1.00	1.02	1.01	1.12	1.02	1.03	1.02	2.24	1.05	1.78
Variance function 3 (Negative Binomial)												
CR	1,000	1,000	1,000	911	816	1,000	1,000	1,000	993	920	1,000	1,000
Mean	1.00	1.00	1.00	1.02	1.00	1.05	1.00	0.99	1.63	1.88	1.00	1.09
SD	0.00	0.01	0.01	0.04	0.02	0.10	0.12	0.09	1.87	0.56	0.14	0.43
Med	1.00	1.00	1.00	1.00	1.00	1.06	0.99	0.99	1.30	1.74	1.01	1.00
5th	1.00	1.00	1.00	1.00	1.00	0.89	0.82	0.84	0.87	1.21	0.78	0.66
95th	1.00	1.00	1.00	1.09	1.00	1.23	1.23	1.12	3.01	3.01	1.23	1.82
Variance function 4 (Gaussian)												
CR	1,000	1,000	1,000	1,000	594	1,000	1,000	1,000	1,000	799	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.06	1.00	1.00	1.00	1.86	1.00	1.36
SD	0.00	0.00	0.00	0.00	0.03	0.06	0.01	0.02	0.00	0.72	0.04	0.88
Med	1.00	1.00	1.00	1.00	1.00	1.06	1.00	1.01	1.00	1.68	1.01	1.18
5th	1.00	1.00	1.00	1.00	1.00	0.95	0.98	0.97	1.00	1.02	0.95	0.76
95th	1.00	1.00	1.01	1.00	1.01	1.16	1.01	1.03	1.00	3.25	1.05	2.39
Variance function 5 (Inverse Gaussian)												
CR	996	999	999	982	996	996	1,000	1,000	993	996	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	0.98	0.97	0.97	1.09	0.99	0.98	0.97
SD	0.00	0.00	0.00	0.01	0.00	0.02	0.08	0.04	0.50	0.02	0.07	0.07
Med	1.00	1.00	1.00	1.00	1.00	0.98	0.97	0.97	0.98	0.99	0.98	0.97
5th	1.00	1.00	1.00	1.00	1.00	0.95	0.90	0.93	0.90	0.96	0.92	0.90
95th	1.00	1.00	1.00	1.02	1.00	1.01	1.04	1.02	1.66	1.02	1.03	1.03
Variance function 6 (Other)												
CR	1,000	1,000	1,000	1,000	691	1,000	1,000	1,000	1,000	910	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.05	1.00	1.00	1.00	1.65	1.00	1.25

Table 3 continued

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
SD	0.01	0.00	0.00	0.00	0.02	0.06	0.01	0.02	0.00	0.57	0.04	0.63
Med	1.00	1.00	1.00	1.00	1.00	1.04	1.00	1.00	1.00	1.51	1.01	1.12
5th	1.00	1.00	1.00	1.00	1.00	0.95	0.98	0.97	1.00	1.01	0.94	0.78
95th	1.00	1.00	1.00	1.00	1.00	1.14	1.02	1.03	1.01	2.75	1.05	1.97

The table shows descriptive statistics for the distribution of estimates of $\beta_1 = 1$ in Eq. (16) for 1,000 replications of the DGP of Sect. 3.1 for each of six different variance functions of Table 2. The statistics are as follows: number of converged replications (CR), as well as mean, standard deviation (SD), median (Med), 5th and 95th percentiles, all of these computed over converged replications. Gam, P, NB, Gau, and IG stand for Gamma, Poisson Negative Binomial, Gaussian and Inverse Gaussian-GLM estimators. QD stands for Quasi-differences estimators, and GMM and P correspond to Eqs. (14) estimated by two-step GMM and (15) estimated by Poisson-GLM. Where statistics were larger than 10 they have been replaced by the symbol †

IS estimators in this case stems from the fact that they specify the model absolutely correctly and that the DGP with a single well-behaved explanatory variable d_{ij} is very simple. In this case, the IS estimators may serve as a benchmark against which the other estimators can be measured. The only exception to the outstanding performance of the IS estimators lies in the inverse Gaussian (IS-IG) estimator's serious convergence difficulties. With a convergence failure rate as high as 40 %, this issue points to major numerical difficulties of this estimator in handling the optimization process reliably for DGPs other than its optimal variance function.

In the group of FE estimators, the performances of Poisson (FE-P) and Negative Binomial (FE-NB) come closest to the one of the IS estimators. They display no bias and present small standard deviations for FE estimators, lagging only behind the estimator that specifies the variance process correctly.[14] While the Gamma FE estimator (FE-Gam), too, has similar dispersion, it is slightly biased upward in most DGPs. In contrast, the Gaussian -i.e., nonlinear least squares- (FE-Gau) and inverse Gaussian estimators (FE-IG) display more serious problems. FE-Gau's mean is seriously biased upward under the Gamma and Negative Binomial variance function DGPs. The problem is not just one of a few outliers, as can be seen from FE-Gau's median which does not seem much better-behaved. FE-IG's performance is dismal. Its mean is substantially distorted and so is its median. Only if the variance function of exports is cubic (i.e., with *Variance function 5* in the table) does FE-IG produce good results (indeed,

[14] The NB-GLM estimators have their overdispersion parameter fixed at 1. Theoretically, one could improve on this by estimating an optimal overdispersion parameter in a two-step NB Quasi-Generalized Pseudo-Maximum Likelihood procedure, as proposed by Bosquet and Boulhol (2014). Table 12 in the Appendix displays results for this estimator. We found no or negligible efficiency gains in our DGPs from doing so. Moreover, the estimator had severe difficulties in many of our DGPs. The reason is that NB QGPML estimates the coefficients a and b in the variance function $a\mu_{ij} + b\mu_{ij}^2$ – which are zero in some of our DGPs – to build an overdispersion parameter b/a. Note that naïvely estimating the overdispersion jointly with β in a quasi-Likelihood approach falls outside of GLM estimation and its property of consistency.

Table 4 Baseline scenario: inconsistent approaches, 1,000 replications

	Est. with GDP as proxy for FE					Est. without add. regressors				
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG
Variance function 1 (Gamma)										
CR	1,000	1,000	1,000	998	989	1,000	1,000	1,000	1,000	1,000
Mean	0.92	0.84	0.87	0.98	1.15	0.89	0.81	0.85	0.88	1.09
SD	0.08	0.12	0.10	1.10	0.33	0.09	0.12	0.10	2.30	0.42
Med	0.93	0.83	0.87	0.79	1.08	0.90	0.80	0.85	0.74	1.02
5th	0.78	0.65	0.72	0.54	0.83	0.74	0.62	0.69	0.51	0.76
95th	1.05	1.04	1.02	1.58	1.66	1.03	1.03	1.01	1.23	1.60
Variance function 2 (Poisson)										
CR	1,000	1,000	1,000	1,000	872	1,000	1,000	1,000	1,000	991
Mean	0.93	0.84	0.88	0.79	1.83	0.90	0.81	0.85	0.77	2.08
SD	0.07	0.06	0.05	0.07	1.54	0.07	0.07	0.06	0.09	2.65
Med	0.94	0.85	0.89	0.79	1.28	0.91	0.82	0.86	0.77	1.11
5th	0.82	0.74	0.78	0.66	0.84	0.78	0.70	0.75	0.62	0.75
95th	1.03	0.93	0.95	0.90	4.43	1.01	0.91	0.94	0.90	6.93
Variance function 3 (Negative Binomial)										
CR	1,000	1,000	1,000	996	817	1,000	1,000	1,000	1,000	977
Mean	0.94	0.84	0.88	1.01	2.10	0.91	0.82	0.85	0.85	1.99
SD	0.11	0.12	0.10	1.53	1.90	0.12	0.12	0.11	0.92	2.23
Med	0.94	0.84	0.87	0.80	1.48	0.91	0.81	0.85	0.75	1.16
5th	0.76	0.65	0.72	0.55	0.77	0.70	0.62	0.68	0.52	0.67
95th	1.12	1.07	1.06	1.54	5.24	1.09	1.04	1.04	1.22	6.96
Variance function 4 (Gaussian)										
CR	1,000	1,000	1,000	1,000	630	1,000	1,000	1,000	1,000	885
Mean	0.95	0.84	0.88	0.79	2.79	0.91	0.82	0.85	0.77	3.35
SD	0.08	0.06	0.05	0.07	2.91	0.08	0.07	0.06	0.09	4.59
Med	0.95	0.85	0.89	0.79	1.46	0.92	0.82	0.86	0.77	1.21
5th	0.80	0.74	0.78	0.66	0.80	0.76	0.70	0.75	0.62	0.74
95th	1.06	0.93	0.95	0.91	8.58	1.04	0.92	0.94	0.90	12.56
Variance function 5 (Inverse Gaussian)										
CR	996	1,000	1,000	988	996	997	1,000	1,000	999	996
Mean	0.90	0.82	0.86	†	1.02	0.88	0.80	0.83	†	1.00
SD	0.07	0.13	0.09	†	0.09	0.09	0.15	0.10	†	0.15
Med	0.91	0.82	0.86	0.77	1.01	0.88	0.79	0.84	0.74	0.98
5th	0.81	0.70	0.75	0.61	0.88	0.76	0.66	0.71	0.57	0.83
95th	0.98	0.92	0.95	0.92	1.18	0.97	0.91	0.93	0.91	1.21
Variance function 6 (Other)										
CR	1,000	1,000	1,000	1,000	781	1,000	1,000	1,000	1,000	944
Mean	0.94	0.84	0.88	0.79	2.36	0.91	0.82	0.85	0.77	2.71

Table 4 continued

	Est. with GDP as proxy for FE					Est. without add. regressors				
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG
SD	0.07	0.06	0.05	0.07	2.29	0.08	0.07	0.06	0.09	3.99
Med	0.94	0.85	0.89	0.79	1.39	0.91	0.82	0.86	0.77	1.14
5th	0.81	0.74	0.78	0.66	0.81	0.77	0.70	0.75	0.62	0.73
95th	1.05	0.92	0.95	0.90	6.11	1.02	0.91	0.94	0.90	8.81

See Table 3

in this case FE-IG provides the best performance among FE estimators, as one would expect).

The last two columns of Table 3 give descriptive statistics of two quasi-differences (QD) estimators' small-sample distribution in these DGPs. Using 245 observations of country-tetrads, a two-step GMM estimator (QD-GMM) and a simple Poisson estimator (QD-P) are used. The performance of QD-GMM is clearly superior to QD-P. Its mean and median are both centered at the true value of unity. Its standard deviation tends to be larger than that of the better FE estimators, though. On the other hand, QD-P is visibly biased in some DGPs.

As the results from the alternative scenarios will show, these results are quite robust and to a large degree they remain unchanged throughout the alterations that the DGP is subjected to.

Finally, results of the baseline DGP for these IS, FE, and QD estimators can be compared to the traditional approaches, which ignore general equilibrium effects through the inclusion of e_i and m_j as fixed effects or structural nonlinear terms. The results are displayed in Table 4. The first five columns contain results for GLM estimators using log-GDP (y_i and y_j) as proxies for e_i and m_i (dubbed *S3* above); the last five columns contain results from GLM regressions of X_{ij} on d_{ij} without any additional regressors (dubbed *S4* above). Table 4 illustrates that large biases can arise from the omission and incorrect treatment of fixed effects. We will not consider these estimators any further in detail. It will suffice to say that the biases in the alternative scenarios with increased α or v_{HT} are substantially larger than those in this baseline. On the other hand, the biases decrease with a larger sample size; a consequence of the fact that in our structural model the correlation between endowments and bilateral trade costs decreases in a growing world (i.e., with more trade partners becoming available).

3.2.2 *Higher elasticity of substitution $\alpha = -9$, and higher variance of endowments, $v_{HT} = 0.3$*

Results for the DGP with $\alpha = 9$ are displayed in Table 5. As discussed above, this experiment increases the variability of trade flows by making countries reacting more strongly to price differences of traded goods, all the while the world endowment is held constant. In general, this is a more challenging DGP for the estimators. The IS estimators maintain their good performance, but convergence failures becomes more prevalent.

Table 5 Alternative 1 ($\alpha = -9$), 1,000 replications

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
Variance function 1 (Gamma)												
CR	1,000	971	999	757	799	1,000	1,000	1,000	965	876	1,000	1,000
Mean	1.00	1.01	1.00	1.02	1.01	1.00	1.00	0.99	1.60	1.24	0.99	0.96
SD	0.00	0.02	0.01	0.04	0.03	0.03	0.11	0.07	0.93	0.32	0.06	0.46
Med	1.00	1.01	1.00	1.00	1.00	1.00	0.99	0.98	1.32	1.17	0.99	0.86
5th	1.00	0.99	1.00	0.97	1.00	0.95	0.84	0.88	0.89	0.99	0.89	0.59
95th	1.00	1.03	1.01	1.08	1.01	1.05	1.20	1.10	3.13	1.77	1.10	1.57
Variance function 2 (Poisson)												
CR	990	999	1,000	957	33	937	1,000	1,000	1,000	17	1,000	1,000
Mean	1.00	1.00	1.00	1.01	1.00	1.34	1.00	1.00	1.00	†	1.03	4.02
SD	0.01	0.01	0.00	0.02	0.01	0.12	0.02	0.03	0.01	†	0.10	22.00
Med	1.00	1.00	1.00	1.00	1.00	1.33	1.00	1.01	1.00	†	1.03	0.92
5th	1.00	1.00	1.00	1.00	1.00	1.15	0.97	0.95	0.99	2.20	0.87	0.33
95th	1.00	1.02	1.01	1.05	1.03	1.54	1.03	1.05	1.02	†	1.17	10.34
Variance function 3 (Negative Binomial)												
CR	974	978	999	721	34	914	1,000	1,000	960	14	1,000	1,000
Mean	1.00	1.01	1.00	1.02	1.01	1.36	1.01	1.00	1.66	†	1.06	3.60
SD	0.01	0.02	0.01	0.04	0.04	0.13	0.11	0.08	2.11	†	0.15	18.16
Med	1.00	1.01	1.00	1.01	1.00	1.36	1.00	1.00	1.32	†	1.07	0.88
5th	1.00	0.99	1.00	0.97	1.00	1.16	0.84	0.87	0.89	0.20	0.81	0.34
95th	1.00	1.03	1.01	1.08	1.12	1.57	1.22	1.13	3.03	†	1.29	11.07
Variance function 4 (Gaussian)												
CR	883	998	1,000	1,000	0	602	1,000	1,000	1,000	44	1,000	1,000
Mean	1.00	1.01	1.00	1.00	–	1.62	1.01	1.02	1.00	†	1.04	3.59
SD	0.01	0.01	0.00	0.00	–	0.17	0.03	0.04	0.00	†	0.12	26.88
Med	1.00	1.00	1.00	1.00	–	1.61	1.01	1.02	1.00	†	1.04	0.69
5th	1.00	1.00	1.00	1.00	–	1.35	0.97	0.96	1.00	†	0.83	0.24
95th	1.00	1.02	1.01	1.00	–	1.91	1.03	1.05	1.00	†	1.23	6.80
Variance function 5 (Inverse Gaussian)												
CR	998	988	998	912	996	998	1,000	1,000	992	942	1,000	1,000
Mean	1.00	1.00	1.00	1.01	1.00	0.99	0.98	0.98	1.03	1.00	0.98	0.99
SD	0.00	0.01	0.01	0.03	0.00	0.01	0.04	0.02	0.27	0.01	0.02	0.14
Med	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.98	0.99	1.00	0.99	0.99
5th	0.99	0.99	1.00	0.99	1.00	0.98	0.94	0.95	0.91	0.99	0.97	0.92
95th	1.00	1.02	1.01	1.06	1.00	1.00	1.01	1.00	1.23	1.01	1.00	1.04
Variance function 6 (Other)												
CR	950	999	1,000	1,000	7	780	1,000	1,000	1,000	47	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.08	1.50	1.00	1.01	1.00	†	1.04	2.85

Table 5 continued

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
SD	0.01	0.01	0.00	0.00	0.24	0.15	0.02	0.03	0.00	†	0.10	16.98
Med	1.00	1.00	1.00	1.00	1.00	1.49	1.01	1.02	1.00	†	1.04	0.72
5th	1.00	1.00	1.00	1.00	0.93	1.26	0.98	0.96	1.00	†	0.86	0.29
95th	1.00	1.02	1.01	1.00	1.62	1.77	1.03	1.05	1.00	†	1.22	9.55

See Table 3

The IG estimator, both in its IS and its FE variant, cannot be used. Its convergence rates are close to 0 %, and even the instances recorded as converged are likely to be convergence failures as the estimates were extremely large numbers. We have marked such cases where the statistic was not credible with the symbol "†".

Among the FE estimators, the performances broadly echo the baseline. FE-P and FE-NB continue to show favorable performances, and likewise FE-Gau continues having difficulties in the Gamma and Negative Binomial DGP. The most striking result is the stark deterioration of FE-Gam, which now only shows acceptable properties in the Gamma and Inverse Gaussian DGPs, while having up to 30 % bias in the median in the other DGPs.

In contrast, QD-GMM is only slightly negatively affected by this DGP relative to the baseline. Similarly, there is not much difference in QD-P's medians. Its means, however, seem quite distorted by outliers.

By increasing $v_{HT} = 0.3$, the variability of the endowments (and, hence, of the fixed effects e_i and m_j) is increased. The simulation results in Table 6 suggest that this kind of change can be handled better by the estimators than the increase in the substitution elasticity; the IS estimators, for instance, have a visibly better convergence rate. The IS-IG estimator works quite reliably; however, its FE-IG counterpart shows the same poor performance as before.

3.2.3 Higher number of countries $C = 50$

Finally, a much larger sample is considered. With $C = 50$, one obtains 2,450 observations on country-pairs that can be used in estimation. This change in the DGP can help in assessing to what extent the problems described above are small-sample difficulties that can be resolved with more observations. The results in Table 7 indicate that by and large most of the issues indeed vanish when data on more countries are available. FE-Gam, FE-P and FE-NB all have little to no bias and standard deviations which are often not far from IS. The average bias of FE-Gau is substantially smaller, and even more so is its median bias. However, the bias is still visible, and this suggests that, while vanishing asymptotically, it can be quite persistent. At $C = 50$, QD-GMM has essentially zero remaining small-sample bias. Its standard deviation is also small, although it is most often larger than that of FE-Gam, FE-P, and FE-NB. It seems unlikely that the two-step QD-GMM will catch up and overtake this group of

Table 6 Alternative 2 ($v_{HT} = 0.3$), 1,000 replications

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
Variance function 1 (Gamma)												
CR	1,000	998	1,000	820	929	1,000	1,000	1,000	985	965	1,000	1,000
Mean	1.00	1.00	1.00	1.02	1.00	1.00	1.00	0.99	1.66	1.26	0.99	0.92
SD	0.00	0.01	0.01	0.05	0.01	0.04	0.11	0.07	1.23	0.29	0.08	0.29
Med	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	1.33	1.20	0.99	0.86
5th	1.00	1.00	1.00	1.00	1.00	0.93	0.84	0.87	0.83	1.00	0.86	0.62
95th	1.00	1.01	1.00	1.09	1.00	1.07	1.20	1.11	3.65	1.67	1.13	1.40
Variance function 2 (Poisson)												
CR	1,000	1,000	1,000	976	262	1,000	1,000	1,000	1,000	200	1,000	1,000
Mean	1.00	1.00	1.00	1.01	1.00	1.17	1.00	1.00	1.00	2.67	1.01	1.97
SD	0.00	0.00	0.00	0.02	0.01	0.09	0.02	0.03	0.01	0.82	0.07	3.88
Med	1.00	1.00	1.00	1.00	1.00	1.16	1.00	1.01	1.00	2.53	1.02	1.13
5th	1.00	1.00	1.00	1.00	1.00	1.02	0.97	0.95	0.98	1.74	0.90	0.50
95th	1.00	1.01	1.00	1.05	1.00	1.33	1.03	1.05	1.02	4.24	1.11	5.17
Variance function 3 (Negative Binomial)												
CR	1,000	999	1,000	834	313	1,000	1,000	1,000	986	186	1,000	1,000
Mean	1.00	1.00	1.00	1.02	1.00	1.20	1.01	1.00	1.67	2.67	1.04	2.05
SD	0.00	0.01	0.01	0.05	0.03	0.11	0.12	0.08	1.25	0.86	0.14	4.55
Med	1.00	1.00	1.00	1.00	1.00	1.19	1.00	0.99	1.35	2.54	1.05	1.07
5th	1.00	1.00	1.00	1.00	1.00	1.02	0.83	0.86	0.88	1.64	0.82	0.46
95th	1.00	1.01	1.00	1.09	1.01	1.39	1.24	1.14	3.40	4.01	1.26	5.49
Variance function 4 (Gaussian)												
CR	1,000	1,000	1,000	1,000	31	998	1,000	1,000	1,000	32	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.33	1.00	1.01	1.00	†	1.03	3.57
SD	0.00	0.00	0.00	0.00	0.01	0.14	0.02	0.03	0.00	†	0.07	15.01
Med	1.00	1.00	1.00	1.00	1.00	1.32	1.01	1.02	1.00	†	1.03	1.03
5th	1.00	1.00	1.00	1.00	1.00	1.12	0.97	0.95	1.00	3.16	0.92	0.37
95th	1.00	1.01	1.01	1.00	1.00	1.56	1.03	1.05	1.00	†	1.13	11.11
Variance function 5 (Inverse Gaussian)												
CR	997	996	1,000	959	992	997	1,000	1,000	992	994	1,000	1,000
Mean	1.00	1.00	1.00	1.01	1.00	0.99	0.98	0.98	1.05	1.00	0.98	0.98
SD	0.00	0.01	0.00	0.02	0.00	0.01	0.06	0.03	0.41	0.01	0.04	0.07
Med	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.98	0.99	0.98	0.98
5th	0.99	1.00	1.00	1.00	1.00	0.98	0.92	0.94	0.90	0.98	0.95	0.92
95th	1.00	1.00	1.00	1.05	1.00	1.00	1.04	1.01	1.46	1.01	1.01	1.04
Variance function 6 (Other)												
CR	1,000	1,000	1,000	998	96	1,000	1,000	1,000	1,000	44	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.01	1.26	1.00	1.01	1.00	†	1.03	2.75

Table 6 continued

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
SD	0.00	0.00	0.00	0.00	0.04	0.12	0.02	0.03	0.00	†	0.07	7.27
Med	1.00	1.00	1.00	1.00	1.00	1.25	1.01	1.01	1.00	2.70	1.03	1.10
5th	1.00	1.00	1.00	1.00	1.00	1.07	0.97	0.95	1.00	1.46	0.92	0.44
95th	1.00	1.01	1.01	1.00	1.01	1.46	1.03	1.05	1.01	†	1.13	9.07

See Table 3

Table 7 Alternative 3 ($C = 50$), 1,000 replications

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
Variance function 1 (Gamma)												
CR	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.15	1.13	1.00	0.98
SD	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.02	0.12	0.04	0.02	0.07
Med	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.13	1.13	1.00	0.97
5th	1.00	1.00	1.00	1.00	1.00	0.98	0.96	0.97	1.01	1.07	0.97	0.89
95th	1.00	1.00	1.00	1.00	1.00	1.02	1.04	1.02	1.36	1.22	1.03	1.12
Variance function 2 (Poisson)												
CR	1,000	1,000	1,000	1,000	844	1,000	1,000	1,000	1,000	640	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	3.03	1.00	1.60
SD	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.62	0.01	0.51
Med	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	2.95	1.00	1.49
5th	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	2.17	0.99	1.13
95th	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.01	1.00	4.08	1.01	2.35
Variance function 3 (Negative Binomial)												
CR	1,000	1,000	1,000	1,000	824	1,000	1,000	1,000	1,000	432	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.16	3.60	1.00	1.53
SD	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.02	0.12	0.66	0.03	0.45
Med	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.14	3.52	1.00	1.41
5th	1.00	1.00	1.00	1.00	1.00	0.99	0.96	0.97	1.00	2.76	0.96	1.09
95th	1.00	1.00	1.00	1.00	1.00	1.05	1.04	1.03	1.38	4.78	1.04	2.37
Variance function 4 (Gaussian)												
CR	1,000	1,000	1,000	1,000	273	1,000	1,000	1,000	1,000	3	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00	5.99	1.00	2.42
SD	0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.01	0.00	0.49	0.01	3.97
Med	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.00	1.00	5.94	1.00	1.59
5th	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	5.53	0.99	1.03
95th	1.00	1.00	1.00	1.00	1.00	1.05	1.00	1.01	1.00	6.51	1.01	6.06

Table 7 continued

	Iterative-structural estimators					Fixed Effects Estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
Variance function 5 (Inverse Gaussian)												
CR	997	1,000	1,000	988	997	997	1,000	1,000	987	997	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.98	1.48	0.99	0.98	0.98
SD	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.01	1.50	0.00	0.04	0.04
Med	1.00	1.00	1.00	1.00	1.00	0.98	0.97	0.97	1.00	0.99	0.98	0.98
5th	1.00	1.00	1.00	1.00	1.00	0.97	0.94	0.96	0.93	0.99	0.95	0.95
95th	1.00	1.00	1.00	1.00	1.00	0.99	1.06	1.00	3.32	1.00	1.03	1.02
Variance function 6 (Other)												
CR	1,000	1,000	1,000	1,000	572	1,000	1,000	1,000	1,000	60	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.00	4.19	1.00	2.19
SD	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.01	0.00	0.82	0.01	6.98
Med	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.00	1.00	4.26	1.00	1.55
5th	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00	3.16	0.99	1.10
95th	1.00	1.00	1.00	1.00	1.00	1.04	1.00	1.01	1.00	5.60	1.01	3.77

See Table 3

FE estimators in terms of efficiency with further increases in sample size. However, we only used a small fraction of the available sets of country-tetrads s, and further efficiency gains might be achieved by increasing the number of sets s included in estimation. The estimator FE-IG cannot be recommended, in general. It is the only FE estimator exhibiting tremendous biases in both mean and median. QD-P cannot be recommended either. Worryingly, some of the biases are larger than in the case of $C = 10$, which suggests that QD-P and estimators like it might not have moments even in larger samples (at least not for the DGPs we considered), or that the rate of convergence is very slow.

3.2.4 t-statistics from fixed effects estimators

We conclude this section by investigating the extent of the incidental parameters problem for FE estimators. As the previous results illustrated, the problem does not bias the estimate of β, which was essentially unity on average for all FE estimators except FE-IG. Rather, the problem manifests itself in the estimation of the asymptotic variance. Table 8 displays descriptive statistics for the estimated t-statistics corresponding to the (true) null hypothesis $H_0 : \beta_{21} = 1$ for the coefficient on d_{ij}; i.e., $t = (\hat{\beta}_{21} - 1)/s.e.(\hat{\beta}_{21})$, where $s.e.(\hat{\beta}_{21})$ is the heteroskedasticity-robust asymptotic standard error for $\hat{\beta}_{21}$. Table 8 is divided into two panels. The left panel gives results for the baseline specification with 10 countries ("C=10"). The mean of the t-statistics over the 1,000 Monte Carlo replications is almost always quite different from the theoretical mean of zero for all five FE estimators. A look at the rows for the medians reveals that they are close to the means, indicating that it is not a few outliers that are

Table 8 t-statistics for fixed effects estimators (baseline specification), 1,000 replications

	$C = 10$					$C = 50$				
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG
Variance function 1 (Gamma)										
CR	1,000	1,000	1,000	992	999	1,000	1,000	1,000	1,000	1,000
Mean	0.02	−0.13	−0.26	1.20	2.13	0.01	−0.10	−0.32	1.75	4.17
SD	1.29	1.56	1.30	1.57	1.66	1.00	1.14	1.11	1.10	1.47
Med	0.04	−0.12	−0.27	1.07	1.86	0.01	−0.06	−0.31	1.72	4.07
5th	−2.21	−2.65	−2.29	−1.11	−0.15	−1.59	−2.06	−2.19	0.09	1.95
95th	2.13	2.46	1.87	3.90	5.09	1.66	1.58	1.36	3.69	7.00
Variance function 2 (Poisson)										
CR	1,000	1,000	1,000	1,000	969	1,000	1,000	841	1,000	625
Mean	0.86	0.12	0.22	0.04	7.10	1.37	0.10	0.10	0.04	†
SD	1.30	1.10	1.16	1.19	5.42	1.38	1.01	1.00	1.05	8.45
Med	1.00	0.02	0.16	0.02	6.03	1.37	0.06	0.06	0.04	†
5th	−1.39	−1.68	−1.59	−1.93	0.82	−0.90	−1.51	−1.48	−1.75	3.73
95th	2.76	2.01	2.27	2.05	†	3.61	1.91	1.87	1.71	†
Variance function 3 (Negative Binomial)										
CR	1,000	1,000	1,000	993	920	1,000	1,000	1,000	1,000	434
Mean	0.74	−0.05	−0.14	1.24	5.80	1.24	−0.08	−0.30	1.80	†
SD	1.40	1.48	1.27	1.63	4.59	1.25	1.14	1.10	1.13	5.71
Med	0.76	−0.11	−0.18	1.03	4.69	1.26	−0.04	−0.24	1.78	†
5th	−1.62	−2.39	−2.21	−1.01	0.90	−0.75	−2.06	−2.14	0.03	2.81
95th	3.09	2.52	2.04	3.98	†	3.30	1.75	1.48	3.65	†
Variance function 4 (Gaussian)										
CR	1,000	1,000	1,000	1,000	799	1,000	1,000	845	1,000	3
Mean	1.06	0.42	0.51	0.06	†	2.75	0.48	0.45	−0.01	†
SD	1.18	1.11	1.18	1.02	†	1.64	1.18	1.18	1.06	8.16
Med	1.29	0.38	0.48	0.02	8.61	2.76	0.36	0.34	−0.02	†
5th	−1.44	−1.22	−1.27	−1.58	0.78	0.03	−1.16	−1.18	−1.71	†
95th	2.60	2.31	2.53	1.82	†	5.47	2.69	2.66	1.77	†
Variance function 5 (Inverse Gamma)										
CR	996	1,000	1,000	993	996	997	999	909	987	997
Mean	−2.85	−2.06	−2.33	−1.08	−2.10	−6.02	−3.09	−3.23	1.05	−2.20
SD	1.88	1.64	1.58	2.13	2.30	3.13	2.76	2.53	4.36	1.79
Med	−2.92	−2.32	−2.53	−1.70	−1.77	−5.68	−3.02	−3.17	−0.00	−2.02
5th	−5.81	−4.29	−4.59	−3.42	−6.25	†	−7.55	−7.39	−4.61	−5.47
95th	0.32	1.13	0.56	3.10	0.93	−1.46	1.12	0.68	9.33	0.51
Variance function 6 (Other)										
CR	1,000	1,000	1,000	1,000	910	1,000	1,000	821	1,000	66
Mean	0.96	0.28	0.37	0.07	9.04	2.05	0.25	0.25	0.00	†

Table 8 continued

	$C = 10$					$C = 50$				
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG
SD	1.29	1.10	1.18	1.10	8.45	1.62	1.10	1.11	1.02	†
Med	1.14	0.22	0.28	0.05	7.18	2.12	0.19	0.18	0.04	†
5th	−1.51	−1.38	−1.37	−1.71	0.39	−0.71	−1.34	−1.37	−1.67	2.85
95th	2.73	2.21	2.44	1.95	†	4.69	2.14	2.15	1.59	†

Displayed summary statistics are for t-statistics from GLM fixed effects estimators, based on robust asymptotic standard errors. The t-statistics are based on the (true) null hypothesis $H_0 : \beta_{21} = 1$. See notes of Table 3 for more information

causing the problem, but rather that the whole distributions are centered away from zero. Glancing over at the right-hand-side panel with the results for $C = 50$ countries exposes the persistency of these biases, which more often than not show little to no improvement compared to the $C = 10$ case. Table 11 in the Appendix shows t-statistics ($C = 10$) for the iterative-structural estimators, which do not suffer from the incidental parameters problem. Like the FE estimators, the IS estimators showed little bias for β in the simulations. Table 11 shows that unlike the FE estimators, however, the same holds for the IS estimators' t-statistics, which are all very close to zero in both mean and median.

4 Application

The purpose of this section is to apply the models discussed in the previous section to data. For this, we employ cross-sectional bilateral export data in nominal US dollars from the United Nations' Comtrade Database among 94 countries in the year 2008 and combine them with data on bilateral trade costs from the Centre d'Études Prospectives et d'Informations Internationales' Geographical Database (2013). In particular, we use two variables from the latter database: bilateral distance (entitled *dist* in the variable list) and common language (entitled *comlang_off* in the variable list). Rather than using (log) distance in the original format, we generate five indicator variables based on the quintiles of log distance (=1 if a pair *ij* exhibits distance in that quintile, =0 else). This strategy is akin to, e.g., Eaton and Kortum (2002). Letting the linear index be more flexible (by allowing for non-linearities and interactions) aims at mitigating concerns that there might be misspecification in the conditional expectation function. Then, we use these five indicators (we suppress the fifth quintile as the norm in order to estimate a parameter on the constant) and additionally interact them with the language indicator. Altogether the model then includes nine arguments in the trade cost function: four distance quintile main indicator variables and five interacted distance quintile with language indicator variables:

$$T_{ij} = \exp\left(\sum_{d=1}^{4} \beta_d^D \text{Dist}_{d,ij} + \sum_{d=1}^{5} \beta_d^L \text{Dist}_{d,ij}\text{Lang}_{ij}\right) \tag{21}$$

While this model seems parsimonious (it excludes other candidates from the trade cost function such as cultural, economic, historical, and institutional similarity indicators), it captures many of those aspects due to their collinearity with geographical proximity. In any case, the analysis here is meant to be illustrative for applied researchers, and the trade cost function could be altered at discretion.

Beyond those variables, all considered GLM models include a constant and some include exporter and importer country fixed effects to estimate e_i and m_j (referred to as *Fixed effects Models*), while others include the iteratively determined structural counterparts to e_i and m_j as functions of the estimates $\hat{T}_{ij}$ (referred to as *iterative-structural models*). Similarly, we tried to estimate two versions of the quasi-differenced estimator with differenced-out fixed effects, one relying on GMM (*QD-GMM*) and one on Poisson-type estimation (*QD-P*). *QD-GMM* failed to converge and we thus only report results for *QD-P*.

Table 9 summarizes the parameter estimates and robust standard errors (in parentheses) for all considered models[15] and, at the bottom of the table, some information on goodness-of-fit beginning with the Akaike information criterion for the GLM models. The latter is a particularly meaningful statistic with gravity models as considered here, since the iterative-structural (IS) models are nested in (are constrained versions of) the fixed effects (FE) models.

The parameter results may be summarized as follows. First, we observe a relatively stark difference among the parameter estimates across the considered estimators. Only a rough pattern is common to all estimators: Trade between countries in the first quintile of distance is higher by orders of magnitude relative to the more distant countries. Distant country-pairs that share a common language tend to trade more than those with different languages; but country-pairs in the first distance quintile that share the same language trade only about half as much as those with different languages.

From our Monte Carlo simulations in the previous section, we would conclude that the differences across estimators can hardly be due to a misspecification of the variance process alone, since there was virtually no bias in much smaller samples considered before. In any case, it turns out that the Gamma-GLM and the NegBin-GLM estimators obtain very similar parameter estimates whereas those for Gaussian-GLM and Poisson-GLM are relatively different. The Akaike Information Criterion (AIC) suggests that the Gamma-GLM performs best among both the FE and the IS GLMs each. But NegBin-GLM is very close to Gamma-GLM in that regard. It is interesting to note that the AIC for IS estimators is only marginally higher, which suggests some support for the structural restrictions. The row labeled R^2 contains the squared correlation between predicted and actual trade. Here, the Gaussian-FE estimator has the best value, followed by all IS estimators and the Poisson FE estimator. The last two rows of the table address prediction accuracy by looking at squared prediction residuals obtained by leave-one-out cross-validation on a random subsample of 1,000 observations. Both the mean and median squared prediction residual are considered. Poisson-GLM does well in terms of the mean squared prediction residuals in both FE and IS estimators,

[15] The standard errors for the QD-P estimator are derived from resampling the data 100 times for subsamples of one-quarter of the number of countries

Table 9 Estimation results: Gravity model of trade, $C = 94$, $N = 8,836$

	Fixed effects estimators				Iterative-structural estimators				QD estimator
	Gamma	Poisson	NegBin	Gaussian	Gamma	Poisson	NegBin	Gaussian	Poisson
dist1	10.10***	7.10***	9.82***	8.25***	8.89***	7.08***	8.88***	6.05***	7.04***
	(0.16)	(0.26)	(0.15)	(0.85)	(0.20)	(0.29)	(0.20)	(0.51)	(0.25)
dist2	2.23***	1.05***	2.08***	0.29	2.26***	1.03***	2.24***	−0.08	0.18
	(0.10)	(0.28)	(0.09)	(1.32)	(0.18)	(0.28)	(0.18)	(0.52)	(0.27)
dist3	1.24***	1.20***	1.16***	0.48	1.36***	1.21***	1.36***	0.56	0.50*
	(0.08)	(0.28)	(0.07)	(0.50)	(0.09)	(0.28)	(0.09)	(0.51)	(0.27)
dist4	0.80***	0.81***	0.75***	0.09	0.97***	0.82***	0.96***	0.28	0.31
	(0.08)	(0.30)	(0.07)	(0.95)	(0.09)	(0.29)	(0.09)	(0.54)	(0.34)
lang×dist1	−5.62***	−4.54***	−5.55***	−4.97***	−5.58***	−4.52***	−5.58***	−3.90***	−4.79***
	(0.19)	(0.24)	(0.19)	(0.43)	(0.24)	(0.21)	(0.24)	(0.20)	(0.22)
lang×dist2	0.26**	0.60**	0.28**	0.81	-0.61***	0.64**	−0.60***	1.35***	1.53***
	(0.13)	(0.30)	(0.12)	(0.99)	(0.21)	(0.28)	(0.21)	(0.36)	(0.36)
lang×dist3	0.88***	−0.024	0.86***	−0.45	0.26	−0.03	0.25	−0.29	0.31
	(0.18)	(0.31)	(0.17)	(0.57)	(0.18)	(0.26)	(0.18)	(0.32)	(0.25)
lang×dist4	0.35***	0.030	0.38***	−0.42	0.42***	0.06	0.41***	−0.40	0.54*
	(0.13)	(0.29)	(0.12)	(0.56)	(0.14)	(0.19)	(0.14)	(0.29)	(0.294)

Table 9 continued

	Fixed effects estimators				Iterative-structural estimators				QD estimator
	Gamma	Poisson	NegBin	Gaussian	Gamma	Poisson	NegBin	Gaussian	Poisson
lang×dist5	0.58***	0.72*	0.57***	−0.28	1.07***	0.77**	1.07***	−0.05	1.14***
	(0.16)	(0.38)	(0.15)	(0.54)	(0.18)	(0.32)	(0.17)	(0.52)	(0.33)
$AIC \times 10^{-5}$	0.119	975.0	0.120	0.246	0.134	975.7	0.134	0.250	0.006
R^2	0.251	0.439	0.253	0.630	0.436	0.436	0.436	0.435	–
Mean SPR $\times 10^{-9}$	5,333	177	3,049	1,163,211	206	104	206	103	–
Median SPR	4,879	208,644	4,355	5,986	92,824	217,454	92,029	702,867	–

The table contains estimates of a gravity model with specification (21). Robust standard errors in parenthesis. Bootstrapped standard errors for QD-Poisson in parenthesis. The variables dist1,...,dist4 are indicator variables =1 if the country-pair is in the 1,...,4 distance quintile. The variables lang×dist1,...,lang×dist5 are interactions of the distance quintile indicators with an indicator variable =1 if the country-pair shares the same language.

*, ** and *** indicate statistical significance at the 10, 5 and 1 % level. AIC is the Akaike Information Criterion. R^2 is the squared correlation between outcome and fitted values. "Mean SPR" and "Median SPR" are the average and median squared prediction residuals based on leave-one-out cross-validation on a randomly drawn subsample of 1,000 observations

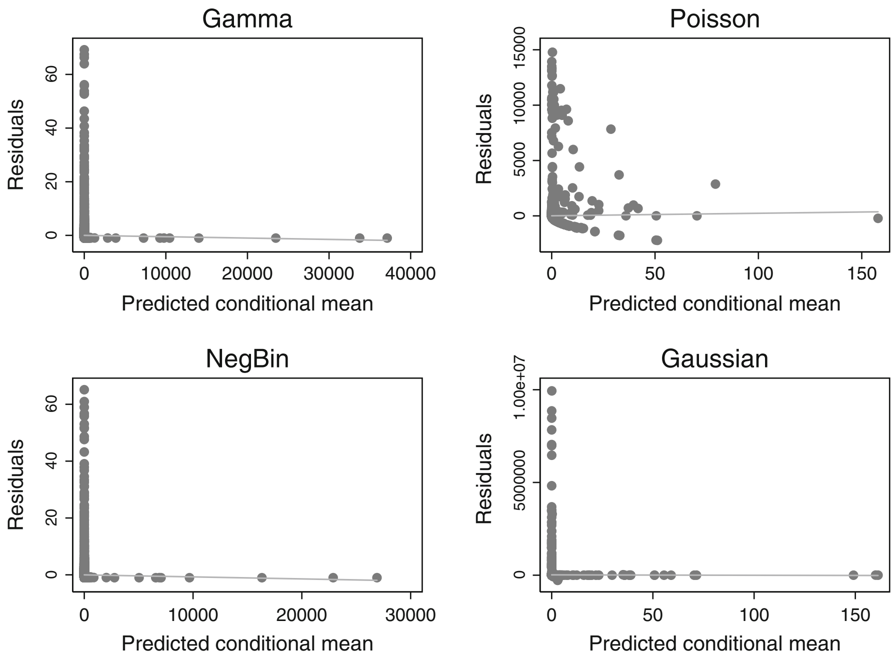

Fig. 1 Fixed Effects estimators: Predictions and Pearson residuals. *Notes*: The figure plots predictions of trade flows against Pearson residuals for fixed effects GLM estimates of the trade gravity model (21), corresponding to columns 2–5 of Table 9

but not so well in terms of the median squared prediction residual, where Gamma and NegBin again display the lowest values within the classes of FE and IS estimators.

Hence, in view of the similarity in the parameter estimates and the comparably low values of the Akaike Information Criterion, Gamma-GLM and NegBin-GLM seem to be the preferred estimators with the data and specification at hand.

Let us define the Pearson residuals of a GLM model of the family type ℓ as

$$\varepsilon_{\ell,ij}^{\text{Pearson}} = \frac{X_{ij} - \hat{\mu}_{ij}}{\hat{\sigma}_{\ell X,ij}}, \tag{22}$$

where $\hat{\sigma}_{\ell X,ij}$ is the square root of the variance function evaluated at the estimated value of the conditional mean, $\hat{\mu}_{ij}$. With a proper specification of the variance function, it should be the case that $\varepsilon_{\ell,ij}^{Pearson}$ is independent of $\hat{\mu}_{ij}$. This is illustrated for the four FE GLMs by way of scatterplots in Fig. 1. In the scatterplots, we include a linear regression line to visualize the relative mean-dependence of the Pearson residuals. If the variance process was correctly specified, we would expect a horizontal line, i.e.. (mean-)independence of the Pearson residuals of $\hat{\mu}_{ij}$. This is largely the case with Gamma-GLM and NegBin-GLM and less so with Poisson-GLM and Gaussian-GLM. Hence, the Pearson residual plots provide further evidence for the first two models.

Furthermore, we may scrutinize the question of the proper specification of the variance function by way of so-called deviance residuals:

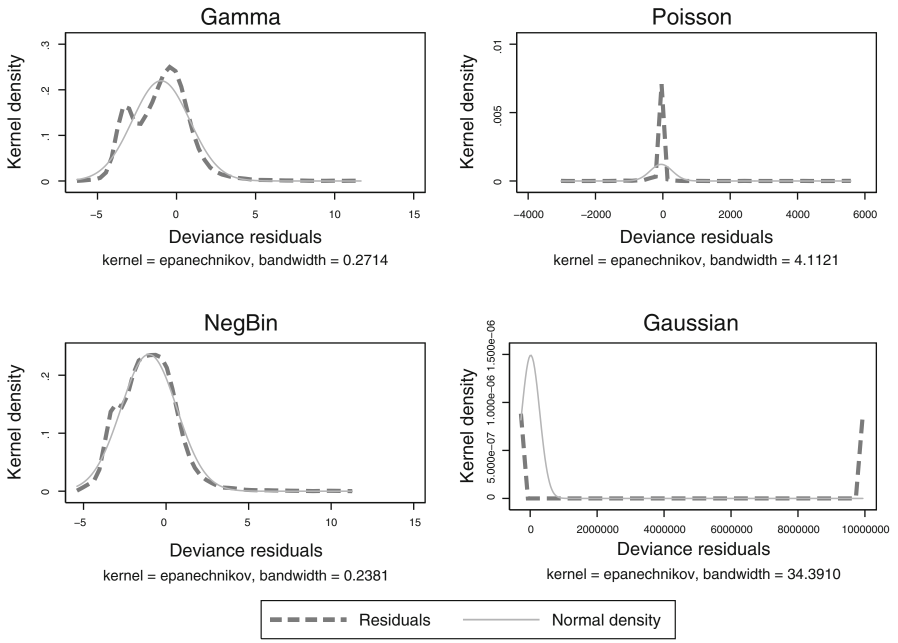

Fig. 2 Fixed effects estimators: Density of deviance residuals. *Notes*: The figure plots kernel density of deviance residuals for fixed effects GLM estimates of the trade gravity model (21), corresponding to columns 2–5 of Table 9

$$\varepsilon_{\ell,ij}^{\text{deviance}} = \frac{\text{sign}(X_{ij} - \hat{\mu}_{ij})}{\sqrt{2}} \left[f_\ell(X_{ij}) - f_\ell(\hat{\mu}_{ij}) \right] \phi, \tag{23}$$

where $f_\ell(\cdot)$ measures the linear exponential family-ℓ conditional density evaluated at the argument. The statistic $\varepsilon_{\ell,ij}^{deviance}$ should have mean zero and be approximately normally distributed for any GLM of family type ℓ (cf. Pierce and Schafer 1986). We shed light on this issue for the FE GLMs in Fig. 2.[16] To facilitate the readability, we present kernel density plots of $\varepsilon_{\ell,ij}^{deviance}$ illustrated by a black dashed curve and add a normal density plot based on the same variance as model ℓ. Figure 2 suggests that, among the considered models, NegBin-GLM performs best, followed by Poisson-GLM. Hence, there appears to be some indication of a larger degree of misspecification of the variance function for Gamma-GLM than for NegBin-GLM. Overall, Fig. 2 in conjunction with Fig. 1 and the goodness-of-fit statistics from Table 9 lead us to classify NegBin-GLM as the preferred model for the data and specification at hand.

5 Conclusions

This paper alludes to issues in the application of generalized linear models for the estimation of (structural) gravity equations of bilateral international trade. The cur-

[16] Corresponding figures for the IS estimators can be found in the Appendix (Figs. 3, 4).

rent status of research on the matter is the following. First, in a host of theoretical models, bilateral trade flows are structurally determined by an exponential function based on a log-linear index (see Arkolakis et al 2012). Second, it has been remarked and well received that exponential-family, generalized linear models rather than log-linearized models should be employed for consistent estimation (see Santos Silva and Tenreyro 2006). The literature appears to favor Poisson-, Gamma- and, to a lesser extent, Gaussian-type GLMs in both analysis and application. Other approaches such as the inverse Gaussian or the Negative Binomial model tend to be ignored. Yet other approaches such as quasi-differencing and generalized method of moments estimation have not been considered at all. As to the model selection, researchers are recommended to resort to goodness-of-fit measures (see Santos Silva and Tenreyro 2006) or not given strong guidance at all. In terms of small-to-medium sample analysis provided in earlier work, the data-generating process has never been selected in accordance with structural models of bilateral trade, and little is known as to how fixed-country-effects estimators fare relative to structural-iterative models.

This paper takes the generic gravity model of bilateral trade literally and focuses on data-generating processes that are fully aligned with general equilibrium or resource constraints present in such models. The paper presents a rich set of Monte Carlo results for various sizes of the world economy (in terms of the number of countries) and various assumptions about the error process. Moreover, the paper takes those insights to cross-sectional data for the year 2008 and illustrates issues with the model selection. The main insights of the paper are the following. First, we find that the Poisson and Negative Binomial quasi-maximum likelihood estimators as well as the quasi-differenced GMM estimators appear to be the best all-round estimators for small as well as larger sample cases and for various stochastic processes. However, we encountered difficulties with the GMM estimator in our application with year 2008 data. For the chosen specification, the Negative Binomial model was the preferred model. A further insight is that the iterative-structural GLM estimators perform better in the simulations than fixed-country-effects estimators due to their greater parsimony. We also illustrated the potentially seriously misleading inference which can result from using asymptotic standard errors with the fixed effects approach. This is due to the incidental parameters problem, which both iterative-structural and quasi-differenced estimators avoid.

The Monte Carlo simulation revealed severe problems with the quasi-differenced Poisson model which is based on ratios of bilateral exports.[17] Likewise, the poor performance of the inverse Gaussian model emerged both in the Monte Carlo simulations with non-inverse-Gaussian data and in the application with real data. It is possible that the problems associated with the Inverse Gaussian estimator are to some degree numerical, though; further research is needed to determine whether using better starting values and optimizing algorithms might improve this estimator's performance.

[17] Some of these issues might carry over to the tetradic-differenced estimators as discussed in Head and Mayer (2014) which are also based on ratios of exports.

Appendix

First-order conditions for fixed effects

The fixed effects e_i and m_j are estimated as the coefficients on a set of exporter and importer dummy variables, respectively:

$$e_i = \sum_{k=2}^{C} e_k D_{ki}, \quad m_j = \sum_{k=2}^{C} m_k D_{kj},$$

where $D_{ki} = \mathbf{1}(i = k)$, $k = 2, \ldots, C$ are the exporter indicator variables; and $D_{kj} = \mathbf{1}(j = k)$, $k = 2, \ldots, C$, the importer indicator variables. The constant β_0 absorbs e_1 and m_1. Then the $2 \times (C - 1)$ first-order conditions for the parameters e_k and m_k, for $k = 2, \ldots, C$, are

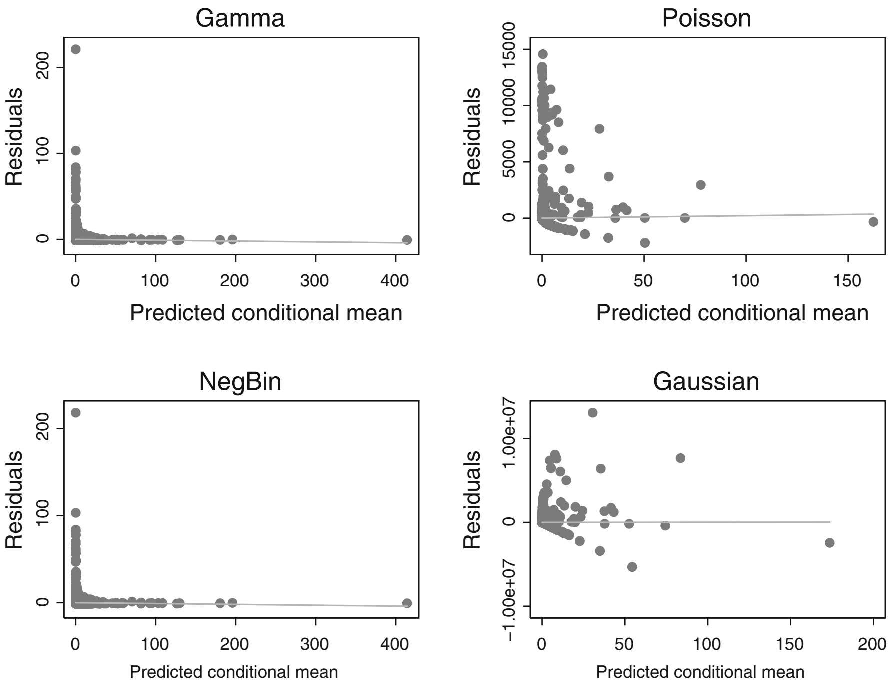

Fig. 3 Iterative structural estimators: Predictions and Pearson residuals. *Notes*: The figure plots predictions of trade flows against Pearson residuals for iterative structural GLM estimates of the trade gravity model (21), corresponding to columns 6–9 of Table 9

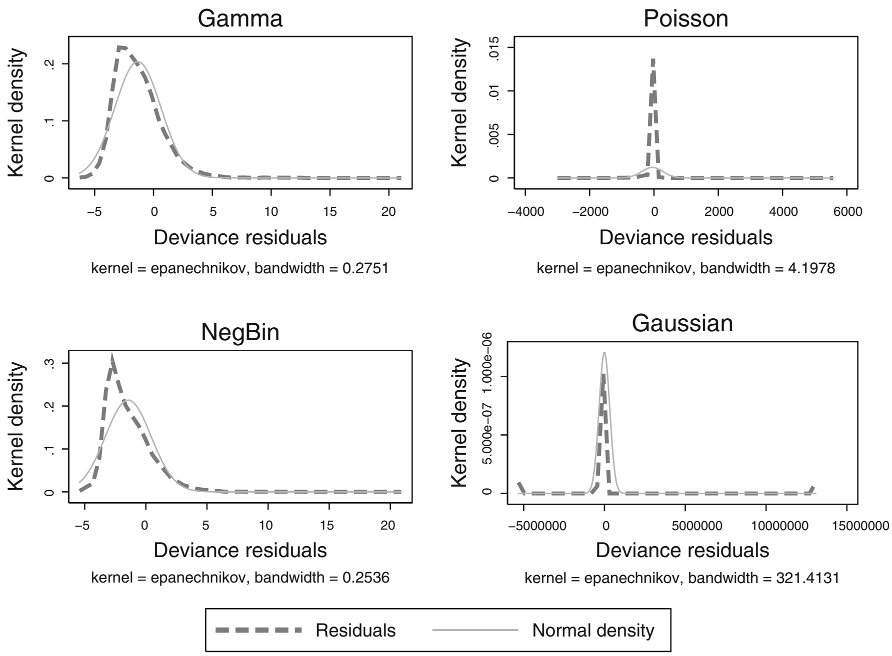

Fig. 4 Iterative-structural estimators: Density of deviance residuals. *Notes*: The figure plots kernel density of deviance residuals for iterative-structural GLM estimates of the trade gravity model (21), corresponding to columns 6–9 of Table 9

$$\sum_{i=1}^{C}\sum_{j=1}^{C}\frac{\left[X_{ij}-\exp\left(\sum_{k=2}^{C}e_k D_{ki}+\sum_{k=2}^{C}m_k D_{kj}+d'_{ij}\beta\right)\right]}{V(X_{ij})}$$
$$\times\exp\left(\sum_{k=2}^{C}e_k D_{ki}+\sum_{k=2}^{C}m_k D_{kj}+d'_{ij}\beta\right)D_{ki}=0,$$
$$\sum_{i=1}^{C}\sum_{j=1}^{C}\frac{\left[X_{ij}-\exp\left(\sum_{k=2}^{C}e_k D_{ki}+\sum_{k=2}^{C}m_k D_{kj}+d'_{ij}\beta\right)\right]}{V(X_{ij})}$$
$$\times\exp\left(\sum_{k=2}^{C}e_k D_{ki}+\sum_{k=2}^{C}m_k D_{kj}+d'_{ij}\beta\right)D_{kj}=0.$$

Table 10 Alternative scenario $\alpha = -2$ (1,000 replications)

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
Variance function 1 (Gamma)												
CR	1,000	1,000	1,000	989	1,000	1,000	1,000	1,000	997	1,000	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.43	1.25	0.99	0.95
SD	0.00	0.01	0.00	0.02	0.00	0.14	0.17	0.14	0.85	0.25	0.20	0.24
Med	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	1.26	1.23	0.98	0.92
5th	1.00	1.00	1.00	1.00	1.00	0.77	0.73	0.75	0.75	0.87	0.66	0.61
95th	1.00	1.00	1.00	1.02	1.00	1.21	1.30	1.21	2.54	1.71	1.34	1.38
Variance function 2 (Poisson)												
CR	1,000	1,000	1,000	1,000	999	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.03	1.00	1.02
SD	0.01	0.00	0.00	0.00	0.01	0.03	0.01	0.02	0.01	0.06	0.03	0.07
Med	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.02	1.00	1.01
5th	1.00	1.00	1.00	1.00	1.00	0.95	0.97	0.97	0.98	0.94	0.95	0.93
95th	1.00	1.00	1.01	1.00	1.00	1.05	1.02	1.03	1.02	1.14	1.04	1.13
Variance function 3 (Negative Binomial)												
CR	1,000	1,000	1,000	990	997	1,000	1,000	1,000	1,000	999	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.02	1.00	0.99	1.53	1.41	0.99	0.96
SD	0.00	0.00	0.00	0.02	0.00	0.15	0.18	0.15	1.79	0.35	0.22	0.27
Med	1.00	1.00	1.00	1.00	1.00	1.02	0.99	0.99	1.26	1.36	1.00	0.93
5th	1.00	1.00	1.00	1.00	1.00	0.77	0.72	0.74	0.73	0.93	0.65	0.61
95th	1.00	1.00	1.00	1.02	1.00	1.27	1.33	1.25	2.91	2.03	1.34	1.42
Variance function 4 (Gaussian)												
CR	1,000	1,000	1,000	1,000	998	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.01
SD	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.06	0.01	0.06
Med	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5th	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99	1.00	0.97	0.99	0.97
95th	1.00	1.00	1.00	1.00	1.00	1.02	1.01	1.01	1.00	1.06	1.01	1.06
Variance function 5 (Inverse Gaussian)												
CR	993	999	999	982	993	993	1,000	1,000	992	994	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.97	1.16	0.98	0.96	0.96
SD	0.00	0.01	0.00	0.06	0.06	0.06	0.19	0.12	0.90	0.07	0.18	0.12
Med	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.97	0.98	0.98	0.97	0.97
5th	1.00	1.00	1.00	1.00	1.00	0.90	0.82	0.87	0.81	0.93	0.79	0.84
95th	1.00	1.00	1.00	1.00	1.00	1.04	1.12	1.05	2.36	1.05	1.10	1.05
Variance function 6 (Other)												
CR	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.01

Table 10 continued

	Iterative-structural estimators					Fixed effects estimators					QD	
	Gam	P	NB	Gau	IG	Gam	P	NB	Gau	IG	GMM	P
SD	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.01	0.00	0.04	0.01	0.04
Med	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00
5th	1.00	1.00	1.00	1.00	1.00	0.97	0.99	0.98	0.99	0.96	0.98	0.95
95th	1.00	1.00	1.00	1.00	1.00	1.03	1.01	1.02	1.01	1.08	1.02	1.08

See Table 3

Table 11 t-statistics for iterative-structural estimators (baseline specification, $C = 10$), 1,000 replications

	Gam	P	NB	Gau	IG
Variance function 1 (Gamma)					
CR	1,000	1,000	1,000	909	988
Mean	0.00	0.02	0.02	0.11	0.01
SD	0.04	0.08	0.11	0.27	0.14
Med	0.00	0.01	0.01	0.01	0.00
5th	−0.01	−0.01	−0.01	−0.02	−0.01
95th	0.01	0.05	0.03	0.66	0.02
Variance function 2 (Poisson)					
CR	1,000	1,000	1,000	998	818
Mean	0.01	0.04	0.04	0.09	0.12
SD	0.12	0.17	0.23	0.27	1.80
Med	0.00	0.04	0.02	0.10	0.00
5th	−0.01	−0.23	−0.13	−0.42	−0.01
95th	0.04	0.31	0.25	0.53	0.12
Variance function 3 (Negative Binomial)					
CR	1,000	1,000	1,000	911	816
Mean	0.01	0.02	0.02	0.12	0.07
SD	0.06	0.08	0.12	0.29	1.14
Med	0.00	0.01	0.00	0.01	0.00
5th	−0.00	−0.01	−0.01	−0.02	−0.01
95th	0.01	0.05	0.03	0.69	0.05
Variance function 4 (Gaussian)					
CR	1,000	1,000	1,000	1,000	594
Mean	0.01	0.07	0.06	0.05	†
SD	0.13	0.35	0.23	0.99	†
Med	0.00	0.05	0.02	−0.01	0.00
5th	−0.01	−0.56	−0.20	−1.49	−0.01
95th	0.03	0.76	0.46	1.73	0.17

Table 11 continued

	Gam	P	NB	Gau	IG
Variance function 5 (Inverse Gamma)					
CR	996	999	999	982	996
Mean	−0.03	−0.01	−0.00	0.07	−0.05
SD	0.23	0.18	0.14	0.27	0.26
Med	0.00	0.02	0.01	0.06	0.00
5th	−0.10	−0.10	−0.06	−0.22	−0.35
95th	0.03	0.09	0.07	0.41	0.06
Variance function 6 (Other)					
CR	1,000	1,000	1,000	1,000	691
Mean	0.02	0.06	0.05	0.05	0.10
SD	0.15	0.25	0.21	0.92	1.94
Med	0.00	0.05	0.02	0.03	0.00
5th	−0.01	−0.41	−0.20	−1.50	−0.01
95th	0.03	0.53	0.33	1.60	0.07

Displayed summary statistics are for t-statistics from iterative structural GLM estimators, based on robust asymptotic standard errors. See notes of Table 3 for more information

Table 12 Negative Binomial Quasi-generalized pseudo-maximum likelihood estimation (Baseline specification, 1,000 replications)

	$C = 10$		$C = 50$	
	IS	FE	IS	FE
Variance function 1 (Gamma)				
CR	565	966	438	784
Mean	1.000	0.989	1.000	0.997
SD	0.008	0.097	0.001	0.013
Med	1.000	0.986	1.000	0.997
5th	0.994	0.843	0.998	0.975
95th	1.003	1.164	1.002	1.018
Variance function 2 (Poisson)				
CR	202	0	8	0
Mean	1.001	–	0.999	–
SD	0.006	–	0.002	–
Med	1.001	–	0.999	–
5th	0.998	–	0.997	–
95th	1.009	–	1.002	–
Variance function 3 (Negative Binomial)				
CR	630	931	697	923
Mean	1.000	0.995	1.000	0.998
SD	0.008	0.106	0.001	0.018

Table 12 continued

	C = 10		C = 50	
	IS	FE	IS	FE
Med	1.000	0.987	1.000	0.998
5th	0.995	0.833	0.998	0.970
95th	1.003	1.186	1.002	1.028
Variance function 4 (Gaussian)				
CR	354	0	28	0
Mean	1.001	–	1.000	–
SD	0.006	–	0.002	–
Med	1.001	–	1.000	–
5th	0.997	–	0.997	–
95th	1.007	–	1.002	–
Variance function 5 (Inverse Gamma)				
CR	256	352	235	364
Mean	1.000	0.984	1.000	0.974
SD	0.004	0.063	0.001	0.029
Med	1.000	0.978	1.000	0.970
5th	0.993	0.931	1.000	0.957
95th	1.003	1.071	1.000	0.987
Variance function 6 (Other)				
CR	348	0	34	0
Mean	1.002	–	1.000	–
SD	0.006	–	0.002	–
Med	1.001	–	0.999	–
5th	0.998	–	0.997	–
95th	1.011	–	1.003	–

See Table 3

References

Adams N, Crowder M, Hand DJ, Stephens D (ed) (2004) Methods and models in statistics. In Honour of Professor John Nelder, FRS. Imperial College Press, London, UK

Allen T, Arkolakis C, Takahashi Y (2014) Universal gravity unpublished manuscript. Yale University, New Haven

Anderson JE, van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. Am Econ Rev 93(1):170–192

Anderson JE, Yotov YV (2010) The changing incidence of geography. Am Econ Rev 100(5):2157–2186

Anderson JE, Yotov YV (2012) Gold standard gravity. NBER Working Papers No 17835

Arkolakis C, Costinot A, Rodríguez-Clare A (2012) New trade models, same old gains? Am Econ Rev 102(1):94–130

Aviat A, Coeurdacier N (2007) The geography of trade in goods and asset holdings. J Int Econ 71:22–51

Baier SL, Bergstrand JH (2009) Bonus vetus ols: a simple method for approximating international trade-cost effects using the gravity equation. J Int Econ 77(1):77–85

Baltagi BH (2013) Econometric analysis of panel data, 5th edn. Wiley, Hoboken

Baltagi BH, Egger PH, Pfaffermayr M (2014) Panel data gravity models of international trade. In: Baltagi BH (ed) The Oxford handbook of panel data. Oxford University Press, Oxford, forthcoming

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. J Int Econ 89(1):110–121

Bosquet C, Boulhol H (2009) Gravity, log of gravity and the "distance puzzle" Unpublished manuscript. Université Paris Panthéon-Sorbonne, Paris

Bosquet C, Boulhol H (2014) Applying the glm variance assumption to overcome the scale-dependence of the negative binomial qgpml estimator. Econometric Rev 33(7):772–784

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, Cambridge

Charbonneau KB (2012) Multiple fixed effects in nonlinear panel data models—theory and evidence unpublished manuscript. Princeton University, Princeton

Dekle R, Eaton J, Kortum SS (2007) Unbalanced trade. Am Econ Rev 97(2):351–355

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

Egger PH, Nigai S (2014) Structural gravity with dummies only. CEPR Discussion Paper

Egger PH, Larch M, Staub KE, Winkelmann R (2011) The trade effects of endogenous preferential trade agreements. Am Econ J: Econ Policy 3(1):113–143

Egger PH, Larch M, Staub KE (2012) The log of gravity with correlated sectors: an application to structural estimation of bilateral goods and services trade. CEPR Discussion Paper No. 9051, Revised version

Fally T (2014) Structural gravity with fixed effects, unpublished manuscript. University of Colorado, Colorado

Fernández-Val I, Weidner M (2013) Individual and time effects in nonlinear panel models with large n, t. CeMMAP working paper CWP60/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies

GeoDist Database (2013) Centre d'Études prospectives et d'informations internationales, Accessed in Oct 2013 on http://www.cepii.fr

Gómez Herrera E (2013) Comparing alternative methods to estimate gravity models of bilateral trade. Empir Econ 44:1087–1111

Hahn J, Newey W (2004) Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72(4):1295–1319

Head K, Mayer T (2014) Gravity equations: workhorse, toolkit, and cookbook. In: Gopinath, Helpman, Rogoff (eds) Handbook of international economics, vol 4

Martin W, Pham CS (2008) Estimating the gravity model when zero trade flows are frequent. Working Paper Economics Series 2008_03, Deakin University

Martínez Zarzoso I (2013) The log of gravity revisited. Appl Econ 45(3):311–327

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London

Mélitz J, Toubal F (2014) Native language, spoken language, translation and trade. J Int Econ 93(2):351–363

Nelder JA, Wedderburn RW (1972) Generalized linear models. J R Stat Soc A 135(3):370–384

Novy D (2013) International trade without ces: estimating translog gravity. J Int Econ 89(2):271–282

Pierce DA, Schafer DW (1986) Residuals in generalized linear models. J Am Stat Assoc 81:977–986

Santos Silva JMC, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88(4):641–658

Santos Silva JMC, Tenreyro S (2011) Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. Econ Lett 112:220–222

Santos Silva JMC, Tenreyro S, Windmeijer F (2015) Is it different for zeros? Discriminating between models for non-negative data with many zeros. J Econom Methods (forthcoming)

Waugh ME (2010) International trade and income differences. Am Econ Rev 100(5):2093–2124