INSTRUMENTAL-VARIABLE ESTIMATION OF EXPONENTIAL REGRESSION MODELS WITH TWO-WAY FIXED EFFECTS WITH AN APPLICATION TO GRAVITY EQUATIONS

Koen Jochmans[*]
Toulouse School of Economics, Université Toulouse 1 Capitole

Vincenzo Verardi[†]
Université de Namur

This version: November 19, 2021

Abstract

This paper introduces instrumental-variable estimators for exponential-regression models that feature two-way fixed effects. These techniques allow us to develop a theory-consistent approach to the estimation of cross-sectional gravity equations that can accommodate the endogeneity of policy variables. We apply this approach to a data set in which the policy decision of interest is the engagement in a free trade agreement. We explore ways to exploit the transitivity observed in the formation of trade agreements to construct instrumental variables with considerable predictive ability. Within a bilateral model, the use of these instruments has strong theoretical foundations. We obtain point estimates of the partial effect of a preferential-trade agreement on trade volume that range between 20% and 30% and find no statistical evidence of endogeneity.

JEL Classification: C23, C26, F14

Keywords: bias correction, count data, differencing estimator, endogeneity, fixed effects, gravity equation, instrumental variable, transitivity.

---

[*]E-mail: `koen.jochmans@tse-fr.eu`.

[†]E-mail: `vverardi@unamur.be`.

Jochmans gratefully acknowledges support from the European Research Council through grant n$^{\text{o}}$ 715787 (MiMo), and from the French Government and the ANR under the Investissements d' Avenir program, grant ANR-17-EURE-0010. Verardi gratefully acknowledges financial support from the FNRS. Comments from three referees on an earlier version have been much appreciated. We are grateful to Peter Egger, Mario Larch, Kevin Staub, and Rainer Winkelmann for discussing their replication material with us. This paper supersedes 'Instrumental-variable estimation of gravity equations' (Cambridge Working Paper in Economics: 1994). The differencing estimator developed in this paper can be implemented in Stata. The module `ivgravity` can be installed from within Stata by typing `ssc install ivgravity` in the command line.

Introduction

Exponential-regression models are a workhorse tool in applied work, with applications in the economics of health (Cameron, Trivedi, Milne and Piggott 1988; Deb and Trivedi 1997; Powell and Seabury 2018), innovation and investment (Hausman, Hall and Griliches 1984; Head and Reis 2008), international trade (Santos Silva and Tenreyro, 2006) and the estimation of wage equations and production functions (Blackburn 2007; Sun, Henderson and Kumbhakar 2011).

The pseudo-poisson maximum-likelihood estimator (Gouriéroux, Monfort and Trognon 1984a,b) has emerged as the default approach to the estimation of exponential-regression models, following influential work by Santos Silva and Tenreyro (2006). This technique is well known to enjoy certain robustness properties with respect to the inclusion of fixed effects (Wooldridge 1999; Fernández-Val and Weidner 2016; Weidner and Zylkin 2021) which are often included in applications to panel and network data to control for unobserved confounding factors. The inclusion of fixed effects need not, however, resolve all endogeneity concerns (Arzaghi and Henderson 2008 provide a discussion on this in their application, for example).

Although instrumental-variable estimators for models with an exponential link function have been proposed (Mullahy 1997; Windmeijer and Santos Silva 1997) and, indeed, have found wide applicability (Tenreyro 2007, Werner 2015, and Cagé, Hervé and Mazoyer 2020 are just a few examples), they do not behave well in the presence of fixed effects (Jochmans, 2021). In this paper we consider exponential models for double-indexed $n \times m$ data where fixed effects are included in each of the two dimensions of the data. In that case the fixed-effect instrumental-variable estimators are consistent if both $n, m \to \infty$ but only have a proper limit distribution when n and m converge at the same rate, i.e., when n/m converges to a finite non-zero constant. Under such rectangular-array asymptotics, the limit distribution is incorrectly centered, though, implying that test statistics (or confidence sets) constructed from it do not have correct size (or coverage).

We first derive the leading bias in the generalized method-of-moment (GMM) estimator of Mullahy (1997) and consider an analytical correction to the estimator that re-centers the limit distribution around zero, thereby salvaging inference procedures based on it. Through extensive simulations we find that the correction removes a considerable amount of bias from the point estimator. However, the presence of fixed effects also introduces bias in the estimator of the (asymptotic) variance; similar observations have been made for the pseudo-poisson estimator (Jochmans 2017; Pfaffermayr 2019), but the problem appears to be somewhat more severe here. This bias is important even in quite large samples and, together with the remaining bias in the point estimator, can lead to inferential procedures with unsatisfactory performance.

We, therefore, next set out to construct a GMM estimator based on moment conditions from which the fixed effects have been 'differenced-out'. This turns out to be feasible in our context by following arguments along the lines in Jochmans (2017). Separating inference on the common parameters from the estimation of the fixed effects is useful as it leads to estimators that are (consistent and) asymptotically unbiased as $n, m \to \infty$, independent of their relative magnitude. Further, as the moment conditions underlying the estimator are free of fixed effects, the associated estimator of the asymptotic variance matrix, in turn, also does not suffer from incidental-parameter bias. In our simulations we find that this procedure provides accurate estimates and reliable inference across all designs, uniformly outperforming the (bias-corrected) fixed-effect estimator.

The gravity equation for international trade flows is an interesting application of our techniques. The gravity equation is an exponential model linking bilateral trade flows to various measures of trade costs. Here, the inclusion of importer and exporter fixed effects has micro-economic foundations (Eaton and Kortum 2002; Anderson and van Wincoop 2003). A concern that arises when taking the gravity equation to the data is whether we can plausibly treat policy variables, such as membership of a currency union or the decision to participate in a preferential-trade agreement, as exogenous (see, e.g., Rose 2000, 2004 and Baier and Bergstrand 2004, 2007).

While the trade literature has long since recognized this problem, coming up with a satisfactory solution has proven difficult. Early work using instrumental variables (e.g., Rose 2000 and Barro and Tenreyro 2007) did not account for general-equilibrium effects and ignored issues of nonlinearity that are now well-understood to be of great importance (Santos Silva and Tenreyro, 2006). As an alternative, Egger, Larch, Staub and Winkelmann (2011) set up a (nonlinear) simultaneous-equation model that incorporates the constraints imposed by general equilibrium. This strategy requires the imposition of strong parametric assumptions and is not robust to misspecification.

An additional concern is that the search for suitable instruments turns out to be quite complicated. First, Head and Mayer (2014, p. 162) note that most variables that plausibly cause trade agreements also appear in the trade equation itself. Second, Rose (2004, p. 110), who experimented with measures of democracy and polity, and measures of freedom, civil rights and political rights, found that this type of variable is only weakly correlated with policy decisions.

As a reaction to these concerns, the profession has favored an approach that exploits time-series variation in the form of panel data (Baier and Bergstrand 2007; Glick and Rose 2016), including multi-way fixed effects to handle any endogeneity concerns (Weidner and Zylkin 2021 introduce approaches to bias correction in this framework). Such an approach may, however, not be satisfactory for the following three reasons. First, the micro-foundations for the gravity model apply to cross-sectional data and are questionable bases for panel data (Head and Mayer, 2014, p. 189). Second, relying solely on time-series variation rules out the possibility to estimate distance elasticities, border effects, and the impact of other determinants of trade that are fixed across time. Third, if current policy decisions react to existing trade flows, the policy variables are not (strictly) exogenous. This renders the pseudo-poisson estimator inconsistent. The feedback issue seems a reasonable concern. Indeed, the data suggest that countries take policy decisions (at least partially) in response to the size of existing trade flows (Santos Silva and Tenreyro 2010, p. 59; Head and Mayer 2014, p. 162).

We apply our differencing estimator to the trade data of Egger, Larch, Staub and Winkelmann (2011). Here, the policy variable feared to be endogenous is the decision to enter into a preferential-trade agreement. Egger, Larch, Staub and Winkelmann (2011) search for exogenous variation in such decisions by using indicators of a shared colonial history and for whether or not the countries in question used to be part of the same country. Using the same set of instruments we find similar results as they do (after a correction to their implementation has been applied; see below). However, our results show that this set of instruments is quite weak, implying wide confidence intervals around the point estimates.

In our quest for alternative instruments we recognize that decisions on bilateral trade policy are not made in isolation. We find high levels of transitivity in the formation of free trade agreements in the data. Moreover, trade within a country pair is much more likely to be subject to a free trade agreement if the respective countries have such an agreement with one or more common third parties. Similar findings are reported in Egger and Larch (2008) and Chen and Joshi (2010). This shows that the number of common free-trade partners is a relevant instrument. The argument underlying the validity of this variable as an instrument is that free trade agreements concluded with third-party countries affect bilateral trade flows only through the importer and exporter effects. Such a mechanism is fully consistent with the theory underlying the cross-sectional gravity model (Anderson and van Wincoop 2003, Anderson and Yotov 2010). Hence, the validity of our instrument is theoretically grounded. We also show how validity and relevance are implied by the specification of Egger, Larch, Staub and Winkelmann (2011) (as well as generalizations thereof in various directions). As we discuss in more detail below, validity can be more difficult to justify if trade agreements are the outcome of a multilateral bargaining process.

Our new instrument allows to obtain more precise point estimates of the impact of free-trade agreements. Across our different specification the partial effect (not taking into account estimation uncertainty) ranges from 20% and 30%. Furthermore, in the data used, we do not find statistical evidence that preferential-trade agreements are created endogenously.

1 Model specification and estimators

We have double-indexed data on a scalar outcome, $Y_{i,j} \geq 0$, a vector of regressors, $\boldsymbol{X}_{i,j}$, and a vector of instruments, $\boldsymbol{Z}_{i,j}$. These data may come in the form of traditional panel data or as data on the pairwise interaction between agents, i.e., as a graph. A workhorse specification when dealing with non-negative outcomes variables is the exponential model

$$Y_{i,j} = \exp\left(A_i + B_j + \boldsymbol{X}'_{i,j}\boldsymbol{\vartheta}\right) V_{i,j}, \tag{1.1}$$

where A_i and B_j are unobserved effects and $V_{i,j}$ is a latent disturbance term. We are interested in estimating the parameter vector $\boldsymbol{\vartheta}$ under the mean-independence assumption

$$\mathbb{E}(V_{i,j}|\boldsymbol{Z}) = \mathbb{E}(V_{i,j}) = 1, \tag{1.2}$$

where we let $\boldsymbol{Z}$ be the collection of $\boldsymbol{Z}_{i',j'}$ for all pairs (i', j'). In doing so we will treat the unobserved effects as fixed. Hence, here and later, all expectations implicitely condition on them. To be clear, mean independence states that $\mathbb{E}(V_{i,j}|\boldsymbol{Z}) = \mathbb{E}(V_{i,j})$. Due to the presence of fixed effects, presuming this mean to be equal to unity amounts to an innocuous normalization.

1.1 Fixed-effect estimator

The first estimator of $\boldsymbol{\vartheta}$ that we consider is a fixed-effect version of the (generalized) method-of-moments estimator of Mullahy (1997). This estimator is based on the set of orthogonality conditions

$$\sum_{(i,j)} \mathbb{E}(\boldsymbol{Z}_{i,j}(V_{i,j} - 1)) = \boldsymbol{0}, \tag{1.3}$$

where the sum ranges over all pairs (i, j) in the data, and

$$\sum_{j \in [i]_1} \mathbb{E}(V_{i,j} - 1) = 0, \qquad \sum_{i \in [j]_2} \mathbb{E}(V_{i,j} - 1) = 0, \tag{1.4}$$

where $[i]_1$ denotes the set of values j for which we observe the pair (i, j) in the data and, similarly, $[i]_2$ denotes the set of values j for which we observe the pair (j, i) in the data.

This is a convenient way of dealing with missing observations assuming, of course, that the missingness is at random. We will let $m_i := |\,[i]_1|$ and $n_i := |\,[i]_2|$, where $|\cdot|$ denotes the cardinality of a set, and will denote the dimensions of the data as n and m, respectively. For example, in a $n \times m$ panel, m_i is the number of time series observations on individual i while n_j is the number of observations in cross section j. In a directed graph, m_i is the number of vertices starting at node i and n_i is the number of vertices arriving at that node. We will assume throughout that n_j/n and m_i/m are bounded away from zero for all indices i and j. Of course, $\sum_i m_i = \sum_j n_j =: r$ is the total number of observations available in the data.

Estimator Equation (1.4) implies a (just-identified) system of estimating equations for the fixed effects for a given value of the parameter vector $\boldsymbol{\vartheta}$. Solving these equations amounts to choosing estimates for the fixed effects such that the implied residuals, $\hat{V}_{i,j}(\boldsymbol{\vartheta})$ (say), satisfy

$$\sum_{j \in [i]_1} (\hat{V}_{i,j}(\boldsymbol{\vartheta}) - 1) = 0, \qquad \sum_{i \in [j]_2} (\hat{V}_{i,j}(\boldsymbol{\vartheta}) - 1) = 0,$$

that is, that their sample means across each of the two indices are equal to unity. Plugging the residuals so obtained back into the empirical counterpart to Equation (1.3) then yields a profiled estimating equation for $\boldsymbol{\vartheta}$. The implied GMM estimator of $\boldsymbol{\vartheta}$ may then be written as

$$\arg\min_{\boldsymbol{\vartheta}} \left(\sum_{(i,j)} \boldsymbol{Z}_{i,j} (\hat{V}_{i,j}(\boldsymbol{\vartheta}) - 1) \right)' \boldsymbol{A} \left(\sum_{(i,j)} \boldsymbol{Z}_{i,j} (\hat{V}_{i,j}(\boldsymbol{\vartheta}) - 1) \right),$$

for a weight matrix $\boldsymbol{A}$. The usual arguments suggest using the inverse of (an estimator of) the variance of the moment conditions as weight matrix. With (conditionally) uncorrelated errors this variance equals

$$\boldsymbol{\Omega} := \sum_{(i,j)} \mathbb{E}(\boldsymbol{Q}_{i,j} \boldsymbol{Q}_{i,j}' (V_{i,j} - 1)^2),$$

where $\boldsymbol{Q}_{i,j} := \boldsymbol{Z}_{i,j} - m_i^{-1} \sum_{j' \in [i]_1} \mathbb{E}(\boldsymbol{Z}_{i,j'}) - n_j^{-1} \sum_{i' \in [j]_2} \mathbb{E}(\boldsymbol{Z}_{i',j}) + r^{-1} \sum_{(i',j')} \mathbb{E}(\boldsymbol{Z}_{i',j'})$ is the deviation of $\boldsymbol{Z}_{i,j}$ from its population linear projection on the space spanned by the fixed effects. In practice, this leads to a two-step estimator, using a preliminary estimator $\hat{\boldsymbol{\vartheta}}$, computed using a known weight matrix, to estimate $\boldsymbol{\Omega}$ in a first-step by

$$\hat{\boldsymbol{\Omega}} := \sum_{(i,j)} \hat{\boldsymbol{Q}}_{i,j} \hat{\boldsymbol{Q}}_{i,j}' (\hat{V}_{i,j}(\hat{\boldsymbol{\vartheta}}) - 1)^2,$$

where $\hat{\boldsymbol{Q}}_{i,j} := \boldsymbol{Z}_{i,j} - m_i^{-1} \sum_{j' \in [i]_1} \boldsymbol{Z}_{i,j'} - n_j^{-1} \sum_{i' \in [j]_2} \boldsymbol{Z}_{i',j} + r^{-1} \sum_{(i',j')} \boldsymbol{Z}_{i',j'}$, and subsequently solving the minimization problem again, now with $\boldsymbol{A}$ set to $\hat{\boldsymbol{\Omega}}^{-1}$, to arrive at the two-step estimator, $\hat{\hat{\boldsymbol{\vartheta}}}$.

Bias correction The need to estimate the fixed effects implies that the approach just described is subject to the incidental-parameter problem of Neyman and Scott (1948). Moreover, it will deliver an inconsistent estimator of $\boldsymbol{\vartheta}$, in general, unless both n and m grow large. Calculations underlying this conclusion are provided elsewhere (Jochmans, 2021). The estimator (when properly normalized) will have a well-behaved limit distribution under such an asymptotic scheme provided that n/m converges to a positive and finite constant, that is, that both dimensions grow at the same rate, although it will be incorrectly centered. Consequently, to ensure that inference based on this distributional approximation is size correct in large samples bias correction is necessary.

The problem arises from the bias that estimation of the fixed effects induces in the profiled estimating equation for $\boldsymbol{\vartheta}$. Using similar arguments as Fernández-Val and Weidner (2016) we find that, under the assumption of (conditionally) independent errors, the bias is

$$\sum_{(i,j)} \mathbb{E}(\boldsymbol{Z}_{i,j}(\hat{V}_{i,j}(\boldsymbol{\vartheta}) - 1)) = \boldsymbol{\beta} + o(n) + o(m),$$

for $\boldsymbol{\beta} := \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2$ with

$$\boldsymbol{\beta}_1 := -\sum_{i} \left(\sum_{j \in [i]_1} \frac{\mathbb{E}(\boldsymbol{Q}_{i,j} V_{i,j}(V_{i,j} - 1))}{m_i} - \frac{1}{2} \sum_{j \in [i]_1} \frac{\mathbb{E}(\boldsymbol{Q}_{i,j} V_{i,j})}{m_i} \sum_{j \in [i]_1} \frac{\mathbb{E}((V_{i,j} - 1)^2)}{m_i} \right),$$

$$\boldsymbol{\beta}_2 := -\sum_{j} \left(\sum_{i \in [j]_2} \frac{\mathbb{E}(\boldsymbol{Q}_{i,j} V_{i,j}(V_{i,j} - 1))}{n_j} - \frac{1}{2} \sum_{i \in [j]_2} \frac{\mathbb{E}(\boldsymbol{Q}_{i,j} V_{i,j})}{n_j} \sum_{i \in [j]_2} \frac{\mathbb{E}((V_{i,j} - 1)^2)}{n_j} \right),$$

which are of order n and m, respectively. The Jacobian of the profiled moment conditions is

$$\boldsymbol{\Upsilon} := \sum_{(i,j)} \mathbb{E}(V_{i,j}(\boldsymbol{Z}_{i,j}\boldsymbol{X}'_{i,j} - \boldsymbol{Q}_{i,j}\boldsymbol{P}'_{i,j})),$$

where $\boldsymbol{P}_{i,j}$ is the residual of a population projection of $\boldsymbol{X}_{i,j}V_{i,j}$ on the space spanned by the fixed effects. Then, as $n, m \to \infty$ at the same rate we get the distributional approximation

$$\hat{\hat{\boldsymbol{\vartheta}}} \overset{a}{\sim} \boldsymbol{N}\left(\boldsymbol{\vartheta} - (\boldsymbol{\Upsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\beta}, (\boldsymbol{\Upsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Upsilon})^{-1}\right),$$

which is not centered at $\boldsymbol{\vartheta}$.

A bias-corrected estimator based on the above findings is easily constructed. It takes the form

$$\hat{\hat{\boldsymbol{\vartheta}}} + (\hat{\hat{\boldsymbol{\Upsilon}}}'\hat{\hat{\boldsymbol{\Omega}}}^{-1}\hat{\hat{\boldsymbol{\Upsilon}}})^{-1}\hat{\hat{\boldsymbol{\Upsilon}}}'\hat{\hat{\boldsymbol{\Omega}}}^{-1}\hat{\boldsymbol{\beta}},$$

where $\hat{\boldsymbol{\beta}} := \hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\beta}}_2$, with

$$\hat{\boldsymbol{\beta}}_1 := -\sum_i \left(\sum_{j\in[i]_1} \frac{\hat{\boldsymbol{Q}}_{i,j}\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})(\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})-1)}{m_i} - \frac{1}{2}\sum_{j\in[i]_1} \frac{\hat{\boldsymbol{Q}}_{i,j}\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})}{m_i} \sum_{j\in[i]_1} \frac{(\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})-1)^2}{m_i} \right),$$

$$\hat{\boldsymbol{\beta}}_2 := -\sum_j \left(\sum_{i\in[j]_2} \frac{\hat{\boldsymbol{Q}}_{i,j}\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})(\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})-1)}{n_j} - \frac{1}{2}\sum_{i\in[j]_2} \frac{\hat{\boldsymbol{Q}}_{i,j}\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})}{n_j} \sum_{i\in[j]_2} \frac{(\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})-1)^2}{n_j} \right),$$

$\hat{\hat{\boldsymbol{\Omega}}}$ is constructed in the same way as $\hat{\boldsymbol{\Omega}}$ given above but now using $\hat{\hat{\boldsymbol{\vartheta}}}$ in the place of $\hat{\boldsymbol{\vartheta}}$, and

$$\hat{\hat{\boldsymbol{\Upsilon}}} := \sum_{(i,j)} \hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})\,(\boldsymbol{Z}_{i,j}\boldsymbol{X}'_{i,j} - \hat{\boldsymbol{Q}}_{i,j}\hat{\boldsymbol{P}}'_{i,j}).$$

Here, $\hat{\boldsymbol{P}}_{i,j}$ is the sample version of $\boldsymbol{P}_{i,j}$. It is defined in the same way as $\hat{\boldsymbol{Q}}_{i,j}$, only with $\boldsymbol{Z}_{i,j}$ replaced by $\boldsymbol{X}_{i,j}\hat{V}_{i,j}(\hat{\hat{\boldsymbol{\vartheta}}})$ throughout. For this corrected estimator we can use the distributional approximation

$$\hat{\hat{\boldsymbol{\vartheta}}} + (\hat{\hat{\boldsymbol{\Upsilon}}}'\hat{\hat{\boldsymbol{\Omega}}}^{-1}\hat{\hat{\boldsymbol{\Upsilon}}})^{-1}\hat{\hat{\boldsymbol{\Upsilon}}}'\hat{\hat{\boldsymbol{\Omega}}}^{-1}\hat{\boldsymbol{\beta}} \overset{a}{\sim} \boldsymbol{N}\left(\boldsymbol{\vartheta}, (\boldsymbol{\Upsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Upsilon})^{-1}\right),$$

under rectangular-array asymptotics. Automatic corrections based on a jackknife procedure can equally be concocted. However, such an approach would require substantially stronger homogeneity restrictions on the distribution of the data (Fernández-Val and Weidner, 2016). It is also not obvious how to deal with missing data in such a procedure (Dhaene and Jochmans, 2015).

Remark It is important to be clear on the fact that the fixed-effect estimator does not reduce to the pseudo-poisson estimator when regressors are instrumented by themselves. Rather, an instrumental-variable version of that approach would be based on the moment conditions

$$\sum_{(i,j)} \mathbb{E}(\boldsymbol{Z}_{i,j} U_{i,j}) = \mathbf{0},$$

together with $\sum_{j \in [i]_1} \mathbb{E}(U_{i,j}) = 0$ and $\sum_{i \in j_2} \mathbb{E}(U_{i,j}) = 0$ for all i and j present in the data, where

$$U_{i,j} := Y_{i,j} - \exp(A_i + B_j + \boldsymbol{X}'_{i,j}\boldsymbol{\vartheta}) = \exp(A_i + B_j + \boldsymbol{X}'_{i,j}\boldsymbol{\vartheta})\,(V_{i,j} - 1) \tag{1.5}$$

is an additive error. An estimator based on the moment equation in the above display (in a context without fixed effects) was discussed in Windmeijer and Santos Silva (1997). However, this moment equation is not implied by our mean-independence condition in (1.2).

The assumption $\mathbb{E}(U_{i,j}|\boldsymbol{Z}) = 0$ would imply the moment conditions above. However, as inspection of (1.5) makes apparent, this condition cannot hold if $\mathbb{E}(V_{i,j}|\boldsymbol{Z}) = 1$ but $\mathbb{E}(V_{i,j}|\boldsymbol{X}) \neq 1$, in general. Hence, the instrumental-variable poisson-type estimator will be inconsistent under our assumptions. Our mean-independence assumption is natural in the context of a system of simultaneous equations and has equally been preferred in related contexts elsewhere (Terza 1998, Egger, Larch, Staub and Winkelmann 2011, Wooldridge 2014, and Jochmans 2015). We refer to Mullahy (1997, Section III.B) and Windmeijer and Santos Silva (1997, Section 2.1) for additional discussion on the difference between the two mean-independence conditions.

Note that, even if $\mathbb{E}(U_{i,j}|\boldsymbol{Z}) = 0$ holds, the introduction of fixed effects into the instrumental-variable estimator of Windmeijer and Santos Silva (1997) again creates an incidental-parameter problem (Jochmans, 2021). Moreover, the well-known robustness of the pseudo-poisson estimator to the inclusion of fixed effects does not carry over to its instrumental-variable counterpart.

1.2 Differencing estimator

The second estimator of $\boldsymbol{\vartheta}$ that we consider is an instrumental-variable generalization of the differencing estimator put forth in Jochmans (2017). This approach is based on the construction of moment conditions that are free of fixed effects. To see how this can be done, let $W_{i,j} := Y_{i,j}/\exp(\boldsymbol{X}_{i,j}'\boldsymbol{\vartheta}) = \exp(A_i + B_j)\, V_{i,j}$. Because $\mathbb{E}(V_{i,j}|\boldsymbol{Z}) = 1$ we have that

$$\mathbb{E}(W_{i,j}|\boldsymbol{Z}) = \exp(A_i + B_j).$$

Next, take two distinct data pairs (i, j) and (i', j') and observe that

$$\mathbb{E}\left(W_{i,j}\, W_{i',j'} |\, \boldsymbol{Z}\right) = \exp(A_i + A_{i'} + B_j + B_{j'})$$

follows when $V_{i,j}$ and $V_{i',j'}$ are (conditionally) uncorrelated. Similarly, for the data pairs (i, j') and (i', j),

$$\mathbb{E}\left(W_{i,j'}\, W_{i',j} |\, \boldsymbol{Z}\right) = \exp(A_i + A_{i'} + B_j + B_{j'})$$

follows in the same way. Noting that the right-hand side of both equations is identical we can take differences and iterate expectations to obtain the unconditional moment conditions

$$\sum_{(i,j)} \sum_{i' \in [j]_2} \sum_{j' \in [i]_1 \cap [i']_1} \mathbb{E}\left(\boldsymbol{Z}_{i,j}(W_{i,j}\, W_{i',j'} - W_{i,j'}\, W_{i',j})\right) = \mathbf{0}, \tag{1.6}$$

which can be used in a GMM procedure to deliver an estimator that will be consistent and asymptotically unbiased as the sample size grows large, irrespective of the relative growth rate of n and m. In fact, one of the dimensions of the data could be held fixed. We will not consider this situation further here, however, as, in that case, a simpler differencing strategy, building on Chamberlain (1992), may be applied to one dimension of the data only; see Jochmans (2021).

Estimator With our moment conditions in hand the implied GMM estimator of $\boldsymbol{\vartheta}$ takes the form

$$\arg\min_{\boldsymbol{\vartheta}} \left(\sum_{(i,j)} \boldsymbol{Z}_{i,j} \overline{W}_{i,j}(\boldsymbol{\vartheta}) \right)' \boldsymbol{A} \left(\sum_{(i,j)} \boldsymbol{Z}_{i,j} \overline{W}_{i,j}(\boldsymbol{\vartheta}) \right),$$

where $\boldsymbol{A}$ is again a chosen weight matrix and we have introduced the notational shorthand

$$\overline{W}_{i,j}(\boldsymbol{\vartheta}) := \sum_{i' \in [j]_2} \sum_{j' \in [i]_1 \cap [i']_1} \left(\frac{Y_{i,j}\, Y_{i',j'}}{\exp((\boldsymbol{X}_{i,j} + \boldsymbol{X}_{i',j'})'\boldsymbol{\vartheta})} - \frac{Y_{i,j'}\, Y_{i',j}}{\exp((\boldsymbol{X}_{i,j'} + \boldsymbol{X}_{i',j})'\boldsymbol{\vartheta})} \right).$$

Let $\check{\boldsymbol{\vartheta}}$ denote the one-step GMM estimator obtained using a known weight matrix. Then, again, the optimal two-step estimator under (conditionally) uncorrelated errors, $\check{\check{\boldsymbol{\vartheta}}}$, may then be calculated using the inverse of

$$\check{\boldsymbol{\Sigma}} := \sum_{(i,j)} \check{\boldsymbol{S}}_{i,j} \check{\boldsymbol{S}}_{i,j}',$$

as (an estimator of the) optimal weight matrix. Here,

$$\check{\boldsymbol{S}}_{i,j} := 4 \sum_{i' \in [j]_2} \sum_{j' \in [i]_1 \cap [i']_1} ((\boldsymbol{Z}_{i,j} + \boldsymbol{Z}_{i',j'}) - (\boldsymbol{Z}_{i,j'} + \boldsymbol{Z}_{i',j}))(\check{W}_{i,j}\, \check{W}_{i',j'} - \check{W}_{i,j'}\, \check{W}_{i',j}),$$

where $\check{W}_{i,j} := Y_{i,j} / \exp(\boldsymbol{X}_{i,j}' \check{\boldsymbol{\vartheta}})$. This construction arises because our empirical moment conditions are not simple sample averages. Moreover, as is apparent from their definition, $\overline{W}_{i,j}(\boldsymbol{\vartheta})$ and $\overline{W}_{i',j'}(\boldsymbol{\vartheta})$ are not independent. The construction of $\check{\boldsymbol{\Sigma}}$ accounts for this using a projection argument (e.g., van der Vaart 2000); see Jochmans (2017, 2018) for related applications of this device.

To state the large-sample distribution of $\check{\check{\boldsymbol{\vartheta}}}$, let

$$\boldsymbol{\Xi} := \sum_{(i,j)} \sum_{i' \in [j]_2} \sum_{j' \in [i]_1 \cap [i']_1} \mathbb{E} \left(\boldsymbol{Z}_{i,j} (W_{i,j}\, W_{i',j'}\, (\boldsymbol{X}_{i,j} + \boldsymbol{X}_{i',j'})' - W_{i,j'}\, W_{i',j}\, (\boldsymbol{X}_{i,j'} + \boldsymbol{X}_{i',j})') \right)$$

be the Jacobian of the moment conditions underlying the estimator and write $\boldsymbol{\Sigma}$ for the population version of $\check{\boldsymbol{\Sigma}}$. Then

$$\check{\check{\boldsymbol{\vartheta}}} \overset{a}{\sim} \boldsymbol{N}(\boldsymbol{\vartheta}, (\boldsymbol{\Xi}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\Xi})^{-1})$$

as $n, m \to \infty$. Recall that, in contrast with the approximation for the (bias-corrected) fixed-effect estimator, here, no requirement on the relative growth rate of n and m is needed.

Our approach differences-out the fixed effects but estimates of the latter can, of course, still be obtained if desired. One way in which this can be done is by applying the two-way pseudo-poisson estimator to the transformed outcome $\check{W}_{i,j}$. Under asymptotics where both $n, m \to \infty$, the estimation noise that is introduced by replacing $\boldsymbol{\vartheta}$ by $\hat{\boldsymbol{\vartheta}}$ is negligible in large samples, implying that the usual (robust) standard errors are valid to perform inference on the incidental parameters.

1.3 Numerical assessment

We report on the performance of the estimators introduced above in a series of simulation experiments. Because models with an exponential link function are used in a wide range of different scenarios we provide results for different types of data generating processes. They cover count data, continuous outcomes, as well as mixed continuous/discrete outcomes where there is a mass point at zero.

To maintain coherency across these different data generating processes we introduce endogeneity into the model through a classical omitted-variable argument. We first generate $Z_{i,j} \sim N(1/2, 1)$ and $O_{i,j} \sim N(-1/2, 1)$ independently and next construct the single regressor

$$X_{i,j} = Z_{i,j} + O_{i,j}.$$

Next, we generate the outcome $Y_{i,j}$ in such a way that

$$\mathbb{E}(Y_{i,j}|X_{i,j}, O_{i,j}) = \exp(X_{i,j}\vartheta + O_{i,j}) =: \mu_{i,j}.$$

Marginally on the variable $O_{i,j}$, we have $\mathbb{E}(Y_{i,j}|X_{i,j}) = \exp(X_{i,j}\vartheta)\ \mathbb{E}(\exp(O_{i,j})|X_{i,j})$, which is not equal to $\exp(X_{i,j}\vartheta)$. We will report results for $\vartheta = 1$. To generate the outcome for the different scenarios we adopt the specifications in Jochmans (2017), as described in more detail below.

Data generation Count data are simulated from the Poisson model and from three negative-binomial models. In the Poisson case the conditional mean and variance both equal the arrival rate, $\mu_{i,j}$. Note that this implies that the variance is a function of the omitted regressor. The negative-binomial model is a mixture model over Poisson models, where the arrival rate has a Gamma distribution with positive shape and scale parameters

θ and $p_{i,j} = (1+\mu_{i,j}/\theta)^{-1}$, respectively. In this case the variance exceeds the mean by $\theta\mu_{i,j}^2$. By setting $\theta \in \{1, 5, 10\}$ we cover several degrees of overdispersion.

Non-negative continuous outcomes are generated through an exponential-regression model with log-normal disturbances. More precisely, we generate outcomes as $\mu_{i,j}\,\varepsilon_{i,j}$ with

$$\varepsilon_{i,j} \sim \log N(-1/2 \log(1+\sigma_{i,j}^2),\, \log(1+\sigma_{i,j}^2))$$

for $\sigma_{i,j}^2 \in \{1, \mu_{i,j}^{-1}, 1+\mu_{i,j}^{-1}, \mu_{i,j}^{-2}\}$. These four cases yield different patterns of (conditional) heteroskedasticity in the outcome. The first specification has homoskedastic errors. The second specification has Poisson-type errors, with the conditional mean and variance being equal. The third specification gives an overinflated variance as in a negative-binomial model with $\theta = 1$. The fourth specification, corresponds to homoskedastic outcomes. Note, though, that, even in this case, the conditional distribution will dependent on the regressors through higher-order moments

To incorporate a mass point at zero we generate outcome from a χ^2 distribution with $d_{i,j}$ degrees of freedom, where $d_{i,j}$ is drawn from a negative-binomial distribution with shape parameter θ and scale parameter $p_{i,j} = (1+\mu_{i,j}/\theta)^{-1}$, where $\theta \in \{5, 15\}$. Here, zero has (conditional) probability mass $(1-p_{i,j})^\theta$.

Simulation results The above configurations yield ten data generating processes. For each we simulated data sets of size $(n,m) \in \{(100,25), (100,50), (100,100), (250,250)\}$, and estimated ϑ by the (fixed-effect) pseudo-poisson estimator (PMLE), the fixed-effect instrumental-variable estimator (FE-IV) and its bias-corrected version (BC), as well as the estimator based on differencing out the fixed effects (DIFF). Here, PMLE is, of course, inconsistent. It is included only to gauge the extent of the endogeneity problem. Tables 1 and 2 provide the median, the interquartile range (IQR) and the empirical coverage of 95% confidence intervals for each of the four estimators as obtained over $5,000$ Monte Carlo replications.

PMLE is upward biased by roughly 50% in all designs. Its confidence intervals contain the true parameter value in almost none of the replications. Hence, endogeneity is a relevant issue here.

FE-IV, too, is upward biased in all designs. Contrary to PMLE, though, this is incidental-parameter bias, not endogeneity bias; it is of the order $n^{-1} + m^{-1}$. The bias is less pronounced in some of the log-normal data generating processes. It is more severe in the cases where the outcome is discrete or mixed continuous/discrete, accounting for 15% to 30% of the point estimate when $(n, m) = (100, 25)$ and 6% to 12% of the point estimate when $(n, m) = (100, 100)$, for example. Although it diminishes as we move to the designs with larger (n, m), the improvement is insufficient to yield reliable inference. In these cases the confidence intervals show substantial undercoverage (or, equivalently, the t-test heavily overrejects under the null.) These observations are a manifestation of our results above and show that bias correction is needed.

BC removes about half of the bias from FE-IV. This results in a large improvement in coverage rates. Even with large (n, m) some undercoverage remains, though. When $(n, m) = (100, 100)$ it ranges from 5 to 10 percentage points in all but the three last designs. This can be explained by the (downward) bias in the standard errors of the fixed-effect estimator. Similar observations have been made for the pseudo-poisson estimator (under exogeneity) by Jochmans (2017) and Pfaffermayr (2019), and by Weidner and Zylkin (2021) in a panel data setting. In the last three designs the lingering presence of relatively more bias further hurts the coverage rates, bringing them down to 60% to 70%, far below their theoretical rate of 95%.

As DIFF is based on moment conditions that are free of incidental parameters it only suffers from standard nonlinearity bias, which is $(nm)^{-1}$ in our setting. The estimator performs well for all configurations of (n, m). Its bias is uniformly (across designs) smaller than that of all the other estimators while its variability is comparable to that of the other estimators. The absence of any substantial bias, together with the fact that the standard errors do not suffer from incidental-parameter bias, explains why the DIFF estimator yields confidence intervals with close to correct coverage for all data generating processes and for all configurations of (n, m).

In conclusion our simulation results strongly point in favor of our differencing estimator.

Table 1: Simulation results

	MEDIAN				IQR				COVERAGE (95%)			
	PMLE	FE-IV	BC	DIFF	PMLE	FE-IV	BC	DIFF	PMLE	FE-IV	BC	DIFF
					$(n, m) = (100, 25)$							
Poisson	1.495	1.153	1.061	1.003	0.060	0.103	0.092	0.073	0.001	0.224	0.719	0.945
Negbin ($\theta = 1$)	1.479	1.184	1.088	1.006	0.102	0.129	0.114	0.086	0.000	0.220	0.633	0.949
Negbin ($\theta = 5$)	1.489	1.159	1.066	1.004	0.069	0.105	0.094	0.074	0.001	0.213	0.705	0.946
Negbin ($\theta = 10$)	1.492	1.157	1.064	1.004	0.067	0.105	0.093	0.073	0.001	0.219	0.711	0.951
Normal (1)	1.479	1.002	1.002	1.001	0.094	0.046	0.051	0.057	0.000	0.910	0.884	0.956
Normal (2)	1.495	1.055	1.027	1.004	0.061	0.053	0.056	0.058	0.002	0.599	0.803	0.946
Normal (3)	1.482	1.048	1.026	1.003	0.095	0.062	0.067	0.074	0.001	0.706	0.801	0.942
Normal (4)	1.494	1.123	1.074	1.023	0.062	0.069	0.072	0.088	0.001	0.204	0.543	0.887
Mixture ($\theta = 5$)	1.493	1.307	1.190	1.015	0.077	0.160	0.148	0.111	0.001	0.047	0.284	0.929
Mixture ($\theta = 15$)	1.495	1.300	1.182	1.012	0.072	0.153	0.142	0.102	0.001	0.046	0.302	0.935
					$(n, m) = (100, 50)$							
Poisson	1.493	1.087	1.025	1.003	0.047	0.059	0.055	0.051	0.000	0.343	0.859	0.946
Negbin ($\theta = 1$)	1.480	1.096	1.030	1.003	0.078	0.075	0.069	0.061	0.002	0.392	0.823	0.940
Negbin ($\theta = 5$)	1.489	1.088	1.025	1.003	0.056	0.062	0.056	0.052	0.001	0.351	0.853	0.946
Negbin ($\theta = 10$)	1.493	1.085	1.022	1.002	0.050	0.062	0.056	0.052	0.001	0.359	0.864	0.951
Normal (1)	1.480	1.001	1.001	1.001	0.069	0.034	0.037	0.039	0.002	0.922	0.897	0.947
Normal (2)	1.493	1.036	1.014	1.003	0.046	0.038	0.041	0.042	0.001	0.675	0.860	0.940
Normal (3)	1.481	1.036	1.017	1.003	0.073	0.046	0.049	0.052	0.002	0.724	0.832	0.951
Normal (4)	1.493	1.089	1.048	1.017	0.048	0.050	0.053	0.066	0.001	0.231	0.617	0.887
Mixture ($\theta = 5$)	1.493	1.179	1.083	1.007	0.059	0.093	0.088	0.077	0.002	0.081	0.566	0.928
Mixture ($\theta = 15$)	1.494	1.177	1.080	1.007	0.053	0.091	0.085	0.076	0.001	0.086	0.574	0.933

PMLE: pseudo-poisson estimator. FE-IV: fixed-effect instrumental-variable estimator. BC: bias-corrected fixed-effect instrumental-variable estimator. DIFF: differencing estimator.

Table 2: Simulation results (cont'd)

	MEDIAN				IQR				COVERAGE (95%)			
	PMLE	FE-IV	BC	DIFF	PMLE	FE-IV	BC	DIFF	PMLE	FE-IV	BC	DIFF
$(n, m) = (100, 100)$												
Poisson	1.494	1.054	1.011	1.001	0.036	0.040	0.038	0.036	0.000	0.428	0.901	0.953
Negbin ($\theta = 1$)	1.481	1.060	1.014	1.001	0.062	0.049	0.047	0.044	0.001	0.458	0.879	0.948
Negbin ($\theta = 5$)	1.490	1.055	1.012	1.002	0.042	0.042	0.040	0.037	0.000	0.435	0.904	0.954
Negbin ($\theta = 10$)	1.493	1.054	1.010	1.001	0.040	0.041	0.038	0.036	0.000	0.444	0.895	0.950
Normal (1)	1.482	0.999	0.999	0.999	0.057	0.025	0.028	0.029	0.000	0.939	0.912	0.955
Normal (2)	1.494	1.026	1.009	1.002	0.036	0.028	0.030	0.031	0.000	0.680	0.879	0.941
Normal (3)	1.482	1.025	1.010	1.002	0.056	0.033	0.036	0.038	0.000	0.760	0.863	0.945
Normal (4)	1.495	1.069	1.035	1.014	0.036	0.037	0.040	0.050	0.000	0.208	0.626	0.875
Mixture ($\theta = 5$)	1.491	1.122	1.046	1.006	0.045	0.059	0.057	0.055	0.000	0.088	0.688	0.935
Mixture ($\theta = 15$)	1.494	1.121	1.045	1.005	0.041	0.057	0.055	0.053	0.000	0.083	0.689	0.939
$(n, m) = (250, 250)$												
Poisson	1.496	1.021	1.002	1.000	0.019	0.014	0.015	0.014	0.000	0.475	0.934	0.944
Negbin ($\theta = 1$)	1.487	1.022	1.003	1.000	0.033	0.018	0.018	0.017	0.000	0.526	0.920	0.944
Negbin ($\theta = 5$)	1.492	1.021	1.002	1.000	0.022	0.015	0.015	0.015	0.000	0.481	0.932	0.943
Negbin ($\theta = 10$)	1.495	1.022	1.003	1.001	0.020	0.015	0.015	0.015	0.000	0.448	0.930	0.952
Normal (1)	1.488	1.000	1.000	1.000	0.027	0.010	0.011	0.011	0.000	0.943	0.926	0.954
Normal (2)	1.497	1.013	1.003	1.001	0.018	0.012	0.012	0.013	0.000	0.634	0.902	0.944
Normal (3)	1.488	1.012	1.003	1.001	0.030	0.014	0.015	0.016	0.000	0.762	0.907	0.949
Normal (4)	1.495	1.039	1.018	1.007	0.018	0.016	0.018	0.024	0.000	0.079	0.571	0.891
Mixture ($\theta = 5$)	1.494	1.051	1.012	1.000	0.022	0.023	0.022	0.023	0.000	0.074	0.816	0.947
Mixture ($\theta = 15$)	1.496	1.051	1.012	1.001	0.022	0.023	0.023	0.023	0.000	0.082	0.823	0.946

PMLE: pseudo-poisson estimator. FE-IV: fixed-effect instrumental-variable estimator. BC: bias-corrected fixed-effect instrumental-variable estimator. DIFF: differencing estimator.

2 Gravity equation for trade flows

The gravity equation has a long history in international trade. While its origins can be traced back to Tinbergen (1962), the work of Eaton and Kortum (2002) and Anderson and van Wincoop (2003) has provided the gravity model with micro foundations, establishing its place as the workhorse method for the econometric analysis of bilateral trade patterns. The recent literature has made great strides towards credible estimation of the gravity equation. Including importer and exporter fixed effects to control for third-country effects through multilateral resistance terms (see Anderson 1979, Anderson and van Wincoop 2003, and Redding and Venables 2004) has become standard. Further, the equation is now estimated in levels rather than in log-linearized form to deal with such empirically-relevant issues as the existence of zero trade flows between countries and conditional heteroskedasticity (see Santos Silva and Tenreyro 2006). Consequently, the gravity equation is of the form in (1.1), with $Y_{i,j} \geq 0$ a measure of trade intensity (typically, the volume of trade flows) from exporter i to importer j and $\boldsymbol{X}_{i,j}$ a set of covariates that capture trade costs between i and j such as geographical distance, for example.

A difficult issue is how to handle the (potential) endogeneity of policy variables such as the participation in preferential trade agreements (see, e.g., Baier and Bergstrand 2004, 2007, 2009, Frankel and Rose 1998, 2002, and Rose 2004). Although the literature has long since recognized this problem, tackling it within a theory-consistent setting has proven difficult. Early work on instrumental-variable estimation of the gravity equation was based on log-linearized models of trade and did not account for general-equilibrium effects (e.g., Rose 2000 and Barro and Tenreyro 2007).

2.1 A model for trade flows and free-trade agreements

Egger, Larch, Staub and Winkelmann (2011) set up a simultaneous-equation system for bilateral trade flows and the decision to participate to a preferential-trade agreement. The main equation follows Anderson and van Wincoop (2003) and is of the form in (1.1). The second equation is a threshold-crossing specification as in

$$FTA_{i,j} = \begin{cases} 1 & \text{if } H_{i,j} \geq \tilde{V}_{i,j} \\ 0 & \text{if not} \end{cases}, \tag{2.7}$$

where $H_{i,j}$ is a function (assumed to be known up to a set of parameters) of the vector of instrumental variables $\boldsymbol{Z}_{i,j}$. Simultaneity is introduced by allowing $V_{i,j}$ and $\tilde{V}_{i,j}$ to be dependent for a given country pair (but independent across country pairs). The model of Egger, Larch, Staub and Winkelmann (2011) imposes the joint-normality assumption that

$$\begin{pmatrix} \log V_{i,j} \\ \tilde{V}_{i,j} \end{pmatrix} \sim \boldsymbol{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right), \tag{2.8}$$

independent of $\boldsymbol{Z}_{i,j}$. Here, endogeneity is present unless the disturbances are independent. By normality independence is equivalent to the single-parameter restriction that $\rho = 0$. This setup implies that decisions on preferential-trade agreements follow the probit model $\mathbb{P}(FTA_{i,j} = 1 | \boldsymbol{Z}_{i,j}) = \Phi(H_{i,j})$ and that

$$\mathbb{E}(V_{i,j} | \boldsymbol{X}, \boldsymbol{Z}) = FTA_{i,j} \frac{\Phi(\rho\sigma + H_{i,j})}{\Phi(H_{i,j})} + (1 - FTA_{i,j}) \frac{1 - \Phi(\rho\sigma + H_{i,j})}{1 - \Phi(H_{i,j})}.$$

Once the aggregator $H_{i,j}$ is chosen, the model can be estimated in two stages. First, the probit model is fitted to the data to construct an estimator of $H_{i,j}$. Next, a pseudo-poisson regression is performed, using the probit fit to correct the conditional mean, to recover the parameters $\boldsymbol{\vartheta}$ and $\rho\sigma$. Wooldridge (2014) discusses this approach (as well as related procedures in different settings) in more detail.

Implementation Egger, Larch, Staub and Winkelmann (2011) work with the familiar linear-index structure

$$H_{i,j} = C_i + D_j + \boldsymbol{Z}_{i,j}' \boldsymbol{\nabla},$$

where $\boldsymbol{\nabla}$ is an unknown parameter vector. Here, C_i and D_j are importer and exporter fixed effects and $\boldsymbol{Z}_{i,j}$ contains variables that capture political affinities and proxies for iceberg transportation costs, along with all trade-cost variables (except for $FTA_{i,j}$) that make up the vector $\boldsymbol{X}_{i,j}$ in (1.1). This model specification is inspired by influential work by Baier and Bergstrand (2004) where it was shown that this type of specification does well in predicting (in sample) existing trade agreements.

We remark that the presence of fixed effects in the first-stage specification introduces asymptotic bias in the two-step estimation procedure. The (estimated) fixed effects and index coefficients from the probit model appear in the correction term that is applied to the exponential-regression model and introduce non-negligible bias in the pseudo-poisson estimator (as well as in the probit estimator, of course). The problem would equally occur if the second stage would be performed by least squares on a log-linearized gravity equation. See Fernández-Val and Vella (2011) and Dhaene and Jochmans (2015) for further discussion and illustrations. In principle, these biases could be corrected for via suitable extensions of the theory and methods proposed in Fernández-Val and Vella (2011). Such results are not available at present and are not pursued here.

Empirical comparison The specification in (2.7)–(2.8) is encompassed by our model in (1.1)–(1.2). Indeed, our (limited-information) specification is agnostic about the precise manner in which preferential-trade agreements come about. It does not impose parametric restrictions or homoskedasticity, nor does it demand independence between instruments and errors. We compared the two approaches by re-analysing the data set of Egger, Larch, Staub and Winkelmann (2011). These data constitute a cross-section of (directed) trade flows between 126 countries in the year 2005.

Table 3 provides definitions and some descriptive statistics of the variables included in the data. The trade-cost variables are standard. Egger, Larch, Staub and Winkelmann (2011) used three instrumental variables for the formation of trade agreements. These are binary indicators of whether or not one of the countries in the pair was (at some point in time) colonized by the other, whether or not they were colonized by the same (third) country, and whether or not they have (at some point in time) been part of the same country.

The first two columns in Table 4 contain point estimates (and associated standard

Table 3: Descriptive statistics

VARIABLE	MEAN	STD	MIN	MAX	DESCRIPTION
TRADE	305.93	3,257.27	0.00	213,763.06	Nominal exports (in million US dollar)
FTA	0.22	0.42	0.00	1.00	Free trade agreement in place (by 2005)
DIST	8.20	0.83	3.25	9.42	Log of distance
BORD	0.02	0.14	0.00	1.00	Common border
LANG	0.14	0.35	0.00	1.00	Common language
COLONY	0.02	0.12	0.00	1.00	(Former) colonial relationship
COMCOL	0.08	0.27	0.00	1.00	(Former) common colonizer
CURCOL	0.01	0.09	0.00	1.00	Colonial relationship after 1945
SMCTRY	0.01	0.09	0.00	1.00	One country used to be part of the other
CONT	0.23	0.42	0.00	1.00	On the same continent
DURAB	29.40	29.22	0.00	100.00	Durability index of political regime
POLCOMP	8.90	19.94	0.00	98.00	Political competition index
AUTOC	7.99	18.95	0.00	1.00	Autocracy index
COUNTRIES	126				
OBSERVATIONS	15,750				

errors) for the parameters of the gravity equation that are valid under the assumption that trade agreements are formed exogenously. These are obtained using pseudo-poisson (PMLE) and our differencing estimator (DIFF 1). For the latter this means that we instrumented all covariates by themselves. The next two columns in Table 4 correct for endogeneity by using the instrumental variables of Egger, Larch, Staub and Winkelmann (2011). These are the three binary variables COLONY, COMCOL, and SMCTRY. FIML refers to the two-step estimator of Egger, Larch, Staub and Winkelmann (2011). DIFF 2 is our differencing estimator. The fifth and sixth column, and the approach underlying them, will be discussed in the next subsection.

The results for PMLE and FIML in Table 4 differ from those reported in Table 2 of Egger, Larch, Staub and Winkelmann (2011). Inspection of their replication material, available through the publisher's website, reveals that two exporter fixed effects and two importer fixed effects are dropped in the computation of all estimators. In our context (and provided that a constant term is included) it suffices to drop one of each to deal with the fact that fixed effects can only be identified up to location; the results are invariant to which two are chosen.

The point estimates obtained through these four procedures are fairly similar, when taking into account estimation uncertainty. Of primary interest here is the coefficient on the policy variable, or its marginal effect. The latter is the relative change in the trade flow stemming from an exogenous change of FTA status (from zero to one), i.e., (with ϑ the coefficient on FTA)

$$e^{\vartheta} - 1.$$

(Calculation of the full effect—i.e., taking into account general equilibrium conditions—could equally be done but the details of the calculation depend on the model at hand and, in particular, the structure of the multilateral resistance terms that are captured by the fixed effects.) The differencing estimators give the smallest marginal effect of free-trade agreements, with an estimated 26% increase for DIFF 1 and an estimated 17% increase for DIFF 2. The latter is, however, very imprecisely estimated. This suggests that the explanatory power of the instruments might be fairly limited. This motivates our search for

Table 4: Estimation results

	EXOGENOUS FTA		ENDOGENOUS FTA		ENDOGENOUS FTA	
	PMLE	DIFF 1	FIML	DIFF 2	DIFF 3	DIFF 4
FTA	0.386	0.229	0.443	0.153	0.259	0.195
	(0.071)	(0.111)	(0.168)	(0.760)	(0.162)	(0.122)
DIST	-0.618	-0.743	-0.608	-0.748	-0.736	-0.716
	(0.035)	(0.073)	(0.042)	(0.189)	(0.073)	(0.067)
BORD	0.649	0.809	0.651	0.787	0.803	0.8411
	(0.060)	(0.179)	(0.061)	(0.223)	(0.177)	(0.176)
LANG	0.213	0.243	0.218	0.254	0.243	0.3024
	(0.063)	(0.087)	(0.060)	(0.088)	(0.087)	(0.075)
CONT	0.285	0.185	0.272	0.221	0.178	0.2262
	(0.065)	(0.100)	(0.074)	(0.199)	(0.111)	(0.097)
DURAB	-0.003	-0.001	-0.003	-0.001	-0.0001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
POLCOMP	0.086	0.078	0.089	0.077	0.078	0.076
	(0.020)	(0.021)	(0.021)	(0.023)	(0.021)	(0.022)
AUTOC	-0.070	-0.078	-0.073	-0.078	-0.079	-0.084
	(0.029)	(0.021)	(0.030)	(0.028)	(0.020)	(0.021)
CURCOL	0.396	0.352	0.384	0.314	0.355	0.383
	(0.189)	(0.228)	(0.192)	(0.218)	(0.162)	(0.217)
Sargan test (overidentification)			—	2.308	—	10.592
p-value			—	[0.315]	—	[0.105]
Sargan test (exogeneity)			0.119	0.217	0.048	0.599
p-value			[0.731]	[0.642]	[0.827]	[0.439]

PMLE: pseudo-poisson estimator. DIFF 1: our differencing estimator instrumenting FTA by itself. FIML: the estimator of Egger, Larch, Staub and Winkelmann (2011) where endogeneity of FTA is corrected for by a parametric control function using COLONY, COMCOL, and SMCTRY as instruments. DIFF 2: our differencing estimator with FTA instrumented for by COLONY, COMCOL, and SMCTRY. DIFF 3: our differencing estimator instrumenting FTA by (cross-fitted) C_FTA. DIFF 4: our differencing estimator instrumenting FTA by frequencies of (cross-fitted) C_FTA. Importer and exporter fixed effects are included in all procedures.

an alternative set of instruments in the next section. The instruments do pass a standard overidentification test. The corresponding point estimate obtained by FIML is a 55% increase. This number, however, ignores the asymptotic bias in the two-step estimator reported on above. The smaller standard error on the FIML estimate is an artifact of the tightly parametrized specification that underlies it.

Both FIML and DIFF 2 allow to test whether the data suggest that trade agreements are, in fact, formed endogenously. Interestingly, both approaches do not find evidence for this here. Indeed, with values of .73 and .64, respectively, the p-values of these tests both imply that the null of exogeneity would not be rejected at any conventional significance level.

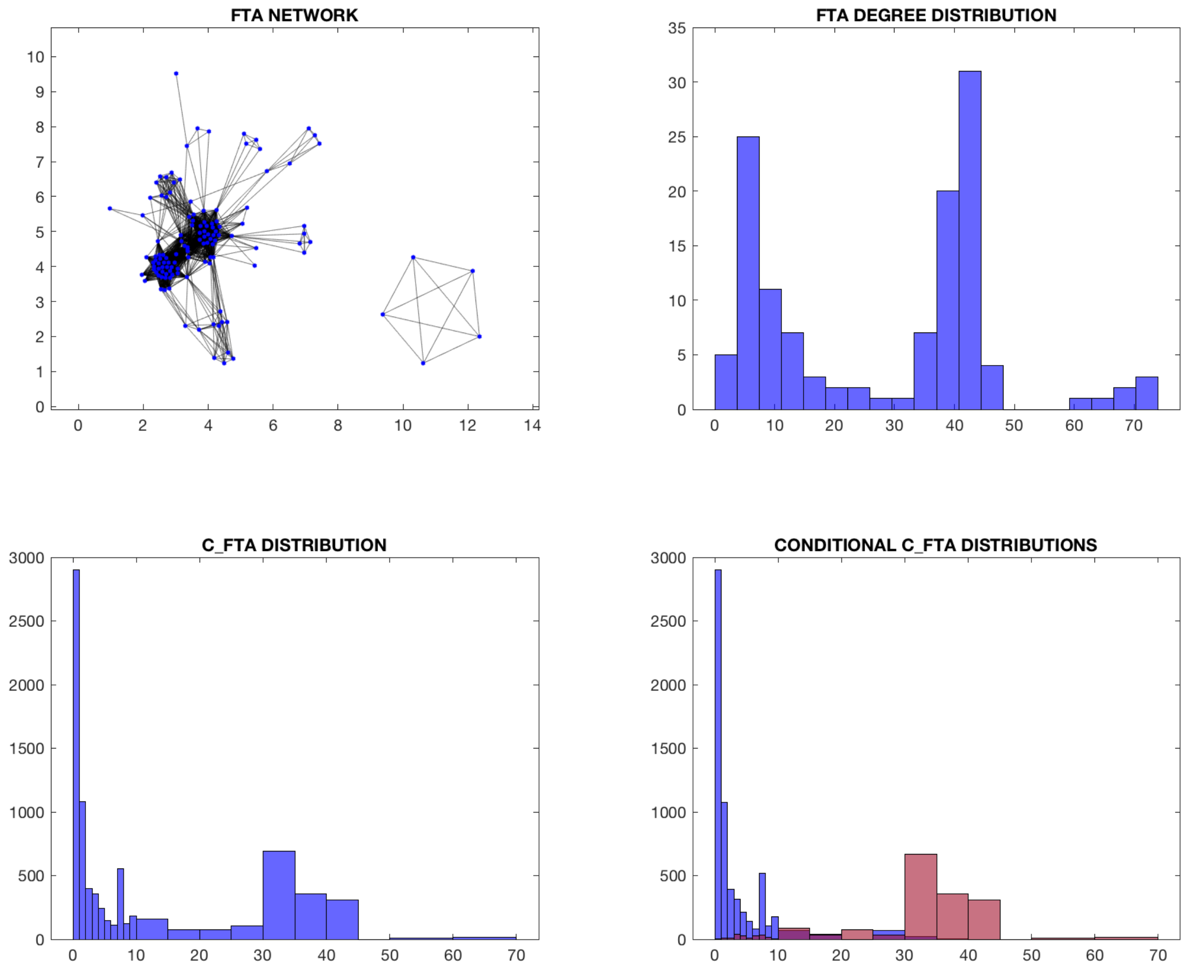
2.2 Leveraging dependence in FTA decisions

The upper left plot of Figure 1 presents the network of preferential-trade agreements in our data. Here, the nodes of the graph (blue dots) are the countries. An edge (black line) is present between two countries if there is a free-trade agreement in place between them. This network is undirected.

The network exhibits substantial heterogeneity in the degree, i.e., in the number of countries with whom an agreement has been made. The upper right plot in Figure 1 contains the histogram of the degree. It is tri-modal.

The figure also reveals the existence of clusters of countries. A useful measure here is the clustering coefficient (e.g., Newman 2010) which, in our context, is the probability that trade between two countries is governed by a trade agreement if both countries have such an agreement in place with a common third-party country. In our data this probability is .784. On the other hand, the unconditional probability of a country pair's trade being subject to a preferential-trade agreement is only .223. This strong transitivity is in line with observations made elsewhere; see Egger, Egger and Greenaway (2008), Egger and Larch (2008), Chen and Joshi (2010), Baldwin and Jaimovich (2012), and Baier, Bergstrand and Mariutto (2014).

Figure 1: Transitivity in FTA formation



Transitivity index The strong clustering shows that there is predictive content about the trade-deal decision between a given pair of countries in their respective decisions to strike such deals with other (third-party) countries. This motivates the construction of instrumental variables for $FTA_{i,j}$ based on

$$C\_FTA_{i,j} := \sum_k FTA_{i,k}\, FTA_{k,j},$$

which is simply the number of countries with whom both Country i and Country j trade on preferential terms. The lower left plot in Figure 1 summarizes this distribution via its histogram. Note that the bins are of unequal length to be able to better differentiate between the different values in the left tail. The lower right plot in the figure, in turn, gives the distribution of the transitivity index (on the same set of bins) conditional on whether (red) or not (blue) a free-trade agreement is in place. There is overlap in the support of these distributions but the mass is distributed very differently. This is in line with the high clustering coefficient reported on above.

The contagion index of Baldwin and Jaimovich (2012, Eq. (11)) for country pair (i, j) is

$$\sum_k \left(\frac{TRADE_{i,j}}{\sum_{j'} TRADE_{i,j'}} \right) \left(\frac{TRADE_{k,j}}{\sum_{i'} TRADE_{i',j}} \right) FTA_{k,j}.$$

It is related to but different from the transitivity index. Its construction is in the spirit of the shift-share approach (see, e.g., Borusyak, Hull and Jaravel 2021). The contagion index has recently been used by Aichele, Felbermayr and Heiland (2016) as an instrument for free-trade agreements (in a log-linearized model). Whether it is suitable for this is questionable, however, as the weights in its contruction are functions of trade flows which themselves, again, depend on FTA decisions.

The transitivity index exploits the network structure of international trade. As such it has a connection to recent contributions such as Lawless (2009), Chaney (2014), and Morales, Sheu and Zahler (2019). These recent models argue that firms tend to export to countries similar to their prior destinations. While our approach is purely cross-sectional, this 'extended gravity' channel is dynamic in nature, and calls for trade within country pair (i, j) in a given time period to depend on their trade costs with third-party trading partners (and, thus, in general also their trade-agreement arrangements) in the previous period.

Validity The argument underlying the validity of the transitivity index as an instrument is that preferential-trade agreements concluded with third-party countries affect bilateral trade flows only through the importer and exporter effects. These effects absorb both all country-specific variables and multilateral resistance. Such a mechanism is fully consistent with the theory underlying the cross-sectional gravity model (Anderson and van Wincoop 2003, Anderson and Yotov 2010), where multilateral resistance terms capture any and all third-party effects. Hence, the validity of our instrument is theoretically grounded. Indeed, the original purpose of multilateral resistance in Anderson and van Wincoop (2003) was precisely to let bilateral trade flows adjust to a change in trade-facilitating conditions with third countries, such as the signing of a free trade agreement.

A potential threat to identification comes from interdependency or the presence of strategic behavior in the formation of trade agreements. Such a situation is not captured by a first-stage equation that is bilateral in nature, such as the one of Egger, Larch, Staub and Winkelmann (2011), for example. Indeed, if policy decisions are the outcome of a simultaneous game, then $FTA_{i,j}$ explicitely depends on $FTA_{i',j'}$ for all other country pairs (i', j'). An implication of this is that the transitivity index will violate our key exclusion restriction.

A related difficulty is that some trade agreements are not bilateral in nature but, rather, the outcome of a joint decision of multiple countries (for example, the European Union). Again, in this case a first-stage equation that is bilateral in nature would be misspecified. The precise way in which blocs negotiate trade agreements can reasonably be expected to be heterogenous across blocs. Furthermore, different member countries may carry relatively more or less weight in the final decision. This would imply, in turn, that the degree to which the decision to establish preferential trade terms is endogenous will be heterogenous across members of the bloc.

Cross-fitting Observe that, when endogeneity is indeed present, the transitivity index of country pairs (i, j') and (i', j), $C\_FTA_{i',j}$ and and $C\_FTA_{i',j}$, are correlated with $V_{i,j}$ in (1.1), as they involve $FTA_{i,j}$ in their construction. This violates the condition in (1.2) underlying our results for the (bias-corrected) fixed-effect estimator of Mullahy (1997). On the other hand, because the differencing estimator is based on interactions between two exporters (i, i') and two importers (j, j') only, it is easy to see that (functions of) the transitivity index can be used as instrument in that context, provided that we construct $C\_FTA_{i,j}$ as

$$\sum_{k \neq i', j'} FTA_{i,k}\, FTA_{k,j}$$

when forming our moment condition (1.6). Note that this correction depends on (i', j'). The cross-fit correction ensures that the instrument does not involve free-trade decisions of the countries in (i, j) with the countries in (i', j'), which are used to difference-out the fixed effects.

An example It is instructive to place the transitivity index within the framework of Baier and Bergstrand (2004) and Egger, Larch, Staub and Winkelmann (2011). A stripped-down (and symmetrized, as is the case in our data) version of the first-stage equation in (2.7) has

$$FTA_{i,j} = FTA_{j,i} = \begin{cases} 1 & \text{if } C_i + C_j \geq \tilde{V}_{i,j} \\ 0 & \text{if not} \end{cases}, \tag{2.9}$$

where C_i is again a country-specific effect. The latter suffice to generate transitivity; no additional explanatory variables are needed in the first-stage equation. Indeed, maintaining (for simplicity) the assumption in (2.8) that the $\tilde{V}_{i,j}$ are independent standard normal across pairs (i, j) we have

$$\begin{aligned} \mathbb{E}(FTA_{i,k}\, FTA_{i,j}) &= \mathbb{E}(\Phi(C_i + C_k)\, \Phi(C_i + C_j)) \\ &\neq \mathbb{E}(\Phi(C_i + C_k))\, \mathbb{E}(\Phi(C_i + C_j)) = \mathbb{E}(FTA_{i,k})\, \mathbb{E}(FTA_{i,j}), \end{aligned}$$

implying relevance of the transitivity index. Furthermore, as $C\_FTA_{i,j}$ is a function of $C_{i'}$ (for all i') and of $\tilde{V}_{i,j'}$ ($j' \neq j$) and $\tilde{V}_{i',j}$ ($i' \neq i$) we can write

$$\mathbb{E}(V_{i,j}|C\_FTA_{i,j}) = \mathbb{E}\left(\mathbb{E}\left(V_{i,j}\left|\begin{array}{l}\tilde{V}_{i,1},\ldots,\tilde{V}_{i,j-1},\tilde{V}_{i,j+1},\ldots,\tilde{V}_{i,n}\\ \tilde{V}_{1,j},\ldots,\tilde{V}_{i-1,j},\tilde{V}_{i+1,j},\ldots,\tilde{V}_{n,j}\end{array}\right.,C_1,\ldots,C_n\right)\right|C\_FTA_{i,j}\right)$$

revealing that the independence of the $\tilde{V}_{i,j}$ across country pairs (i,j) together with the mean-independence of the $V_{i,j}$ of the random effects is more than enough to ensure that $\mathbb{E}(V_{i,j}|C\_FTA_{i,j}) = \mathbb{E}(V_{i,j}) = 1$, thereby implying the validity of the transitivity index as an instrumental variable.

These conclusions do not hinge on the precise form of the first-stage model used in this illustration. The key feature that creates relevance is that there is dependence in a given country's trade policy with respect to different trading partners. This would appear to be an ingredient of any reasonable specification. Validity requires that trade policy decisions of other country pairs are (conditionally) independent. This is the case in conventional bilateral specifications, such as the model of Egger, Larch, Staub and Winkelmann (2011). As discussed above, this need not be the case in situations where trade policy decisions are taken in a strategic manner, however. Note also that the transitivity index remains a proper instrumental variable in cases where the bilateral decision process to engage in a preferential trade agreement is heterogeneous (and this in an unspecified way) across country pairs. A conventional threshold-crossing specification with an index that is additively-separable in (scalar) country effects is in no way necessary.

We report the results from a small set of simulations inspired by this illustration. Our simulated data concern 125 countries. We first draw country effects independently from a (continuous) uniform distribution on [0,1] and next generate a single (symmetric) binary regressor via (2.9) with standard-normal errors. Outcomes are then generated through the ten different designs as in the simulation section in the previous section. In Table 5 DIFF 1 refers to our differencing estimator instrumenting the covariate by itself while DIFF 2 uses the induced transitivity index. The latter is, of course, less precise. The thing to take away here, though, is that, even in this very basic specification, our procedure

Table 5: Simulation results

	MEDIAN		IQR		COVERAGE (95%)	
	DIFF1	DIFF 2	DIFF1	DIFF 2	DIFF1	DIFF 2
Poisson	0.999	1.003	0.056	0.149	0.950	0.948
Negbin ($\theta = \ \ 1$)	0.998	1.003	0.082	0.225	0.958	0.956
Negbin ($\theta = \ \ 5$)	1.000	0.996	0.063	0.161	0.950	0.958
Negbin ($\theta = 10$)	1.001	1.001	0.058	0.158	0.952	0.960
Normal (1)	1.002	0.998	0.060	0.173	0.952	0.960
Normal (2)	0.999	1.001	0.056	0.148	0.952	0.960
Normal (3)	1.001	1.008	0.078	0.233	0.949	0.947
Normal (4)	1.001	1.001	0.057	0.134	0.944	0.952
Mixture ($\theta = \ \ 5$)	1.002	0.991	0.104	0.273	0.967	0.955
Mixture ($\theta = 15$)	1.005	0.997	0.102	0.263	0.952	0.950

DIFF 1: our differencing estimator instrumenting the covariate by itself. DIFF 2: our differencing estimator using the induced transitivity index as instrument.

yields informative inference. Furthermore, this is achieved without the need to identify conventional 'outside' instrumental variables. Indeed, in our example, such variables are absent from the model. This is useful in the trade context because finding such instruments there has proven difficult.

Empirical results We can now return to the results in Table 4. The last two columns there report instrumental-variable estimates based on the transitivity index. The first of these, DIFF 4, corresponds to the case where FTA is instrumented for by the index itself. With a coefficient estimate of .259 we estimate the marginal effect of signing a trade agreement at 29.6%. This gain is twice as large as that obtained using traditional instruments and is estimated much more precisely. DIFF 4 achieves overidentification by using empirical frequencies of the transitivity index as instruments. Moreover, here, we use as instruments indicator variables that switch on for country pairs that have exactly zero, exactly one, exactly two, between 3 and 5, between 6 and 10, between 11 and 20, or more than twenty preferential trading partners in common. This is an approximation to the optimal instrument in this context. Estimation precision improves compared to DIFF 3, and the marginal effect is now estimated at 22%. Further, the Sargan overidentification test has a p-value of .105, giving some confidence that the transitivity index is a valid instrument in our data.

Use of the transitivity index as an instrumental variable yields much more precise point estimates of the importance of trade liberalization than do the more traditional instruments; compare DIFF 3 and DIFF 4 with DIFF 2. On the other hand, the point estimates themselves are quite similar and, indeed, not significantly different from the point estimate of DIFF 1, obtained under the presumption that trade-agreement formation is, in fact, exogenous. Furthermore, a test of exogeneity, based either on DIFF 3 or DIFF 4, again does not allow a rejection of the null of exogeneit at any of the conventional significance levels.

2.3 Discussion

Overall findings Using several different choices for instrumental variables within a theory-consistent specification of the gravity equation we obtained estimates of the partial effect of a free-trade agreement ranging between 20% and 30%. These numbers are in the same range as the point estimates obtained from fixed-effect (pseudo-poisson) regressions on (different) panel data sets; Larch, Wanner, Yotov and Zylkin (2019) report a partial effect of 18%, for example. Using dissagregated panel data, Weidner and Zylkin (2021) find heterogeneous effects by industry ranging broadly from 0% up to 20%. Across all industries their partial-effect estimate is 9%.

We do not find strong statistical evidence of endogeneity in free-trade agreements. In all specifications considered we cannot reject the null of exogeneity at any of the usual significance levels. The point estimates obtained under the assumption of exogeneity, too, are similar to the ones obtained through instrumental variables (when taking into account standard errors).

Accounting for clustering The analysis above was performed under the assumption that the errors are uncorrelated across country pairs. This excludes within-importer and within-exporter dependence. This is a maintained assumption in the (theoretical) literature on fixed-effect estimators. The statistical properties of such estimators in the presence of strong clustering patterns of this type are unknown, at present. We note, though, that our differencing estimator remains consistent if errors are dependent within importer and withing exporter. Hence, the point estimates reported on above remain interpretable. The standard errors, however, would no longer correctly reflect the estimation uncertainty. Of course, clustering would normally be expected to increase standard errors so, if anything, the confidence intervals around the impact of free-trade agreements on trade would be expected to widen further.

Dealing with zero trade flows Our framework does not explicitely model zero trade flows. Structural models of gravity like that of Anderson and van Wincoop (2003) do not naturally generate zeros. Such zeros are, however, observed in typical trade data sets; in the data we use they make up for 37% of the observations. Our econometric specification is not inconsistent with zeros and, indeed, the estimation procedures introduced here can handle them (recall the simulation results for designs where the outcome had a mass point at zero, for example, or see the discussion in Santos Silva and Tenreyro 2011). It could, however, be fruitful to consider an explicit model for such zero trade flows. Theoretical models of trade that generate zeros are introduced by Helpman, Melitz and Rubinstein (2008) and Eaton, Kortum and Sotelo (2012), among others. A (statistical) two-part model, such as in Egger, Larch, Staub and Winkelmann (2011), for example, would be one way to proceed. It would allow to differentiate between the extensive and intensive margin of trade. How to extend the estimation approaches constructed here to such a model is a question that is left for future work.

Conclusion

In this paper we have introduced two estimators for two-way exponential-regression models by instrumental variables. The first is a bias-corrected fixed-effect estimator. The second estimator is a 'differencing' estimator that is based on moment conditions that are free of fixed effects. Theoretical arguments, supported by an extensive set of simulation results, favor the second technique.

We applied the differencing estimator to a cross-sectional gravity equation for trade flows that features importer and exporter fixed effects. In this setting, the policy variable feared to be endogenous is the decision to establish a preferential-trade agreement. Conventional choices of instrumental variables gave very imprecisely estimated partial effects. As an alternative we have constructed estimators where the decision on a free-trade agreement between a pair of countries is instrumented for by (functions of) the same decision of the countries in the pair with the other countries in the data. The relevance and validity condition for such variables to be proper instruments can be supported by theoretical specifications of the gravity equation. We have further illustrated how they are implied by common econometric specifications of how bilateral trade agreements are formed, and have highlighted causes for potential violation of the exclusion restriction. The latter come from considerations that are not well captured by bilateral specifications.

While the point estimates obtained using these new instruments are considerably more precise, we do not find strong evidence in our data that free-trade agreements are formed endogenously. Moreover, our point estimates (when taking into account estimation noise) are similar to those obtained using methods that presume exogeneity. The literature has (using various different methods) obtained partial-effect estimates that span a large range (Baier, Yotov and Zylkin, 2019). It would, therefore, be interesting to see whether our conclusions generalize to other data sets, covering different (or more) countries and other time periods.

References

Aichele, R., G. Felbermayr, and I. Heiland (2016). Going deep: The trade and welfare effects of TTIP revisited. Mimeo.

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review 69*, 106–116.

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

Anderson, J. E. and Y. V. Yotov (2010). The changing incidence of geography. *American Economic Review 100*, 2157–2186.

Arzaghi, M. and J. V. Henderson (2008). Networking off Madison avenue. *Review of Economic Studies 75*, 1011–1038.

Baier, S. L. and J. H. Bergstrand (2004). Economic determinants of free trade agreements. *Journal of International Economics 64*, 29–63.

Baier, S. L. and J. H. Bergstrand (2007). Do free trade agreements actually increase members international trade? *Journal of International Economics 71*, 72–95.

Baier, S. L. and J. H. Bergstrand (2009). Estimating the effects of free trade agreements on international trade flows using matching estimators. *Journal of International Economics 77*, 63–76.

Baier, S. L., J. H. Bergstrand, and R. Mariutto (2014). Economic determinants of free trade agreements revisited: Distinguishing sources of interdependence. *Review of International Economics 22*, 31–58.

Baier, S. L., Y. V. Yotov, and T. Zylkin (2019). On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. *Journal of International Economics 116*, 206–226.

Baldwin, R. and D. Jaimovich (2012). Are free trade agreements contagious? *Journal of International Economics 88*, 1–16.

Barro, R. and S. Tenreyro (2007). Economic effects of currency unions. *Economic Enquiry 45*, 2–23.

Blackburn, M. L. (2007). Estimating wage differentials without logarithms. *Labour Economics 14*, 73–98.

Borusyak, K., P. Hull, and X. Jaravel (2021). Quasi-experimental shift-share research designs. Forthcoming in *Review of Economic Studies*.

Cagé, J., N. Hervé, and B. Mazoyer (2020). Social media and newsroom production decisions. Mimeo.

Cameron, A. C., P. K. Trivedi, F. Milne, and J. Piggott (1988). A microeconometric model of the demand for health care and health insurance in Australia. *Review of Economic Studies 55*, 85–106.

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

Chaney, T. (2014). The network structure of international trade. *American Economic Review 104*, 3600–3634.

Chen, M. X. and S. Joshi (2010). Third-country effects on the formation of free trade agreements. *Journal of International Economics 82*, 238–248.

Deb, P. and P. K. Trivedi (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics 12*, 313–336.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*, 991–1030.

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

Eaton, J., S. Kortum, and S. Sotelo (2012). International trade: Linking micro to macro. Mimeo.

Egger, H., P. Egger, and D. Greenaway (2008). The trade structure effects of endogenous regional trade agreements. *Journal of International Economics 74*, 278–298.

Egger, P. and M. Larch (2008). Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics 76*, 384–399.

Egger, P., M. Larch, K. E. Staub, and R. Winkelmann (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy 3*, 113–143.

Fernández-Val, I. and F. Vella (2011). Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics 163*, 144–162.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel data models with large N, T. *Journal of Econometrics 192*, 291–312.

Frankel, J. A. and A. K. Rose (1998). The endogeneity of the optimum currency area criteria. *Economic Journal 108*, 1009–1025.

Frankel, J. A. and A. K. Rose (2002). An estimate of the effect of common currencies on trade and income. *Quarterly Journal of Economics 117*, 437–466.

Glick, R. and A. K. Rose (2016). Currency unions and trade: A post-EMU reassessment. *European Economic Review 87*, 78–91.

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

Gouriéroux, C., A. Monfort, and A. Trognon (1984b). Pseudo maximum likelihood methods: Theory. *Econometrica 52*, 681–700.

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

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

Head, K. and J. Reis (2008). FDI as an outcome of the market for corporate control: Theory and evidence. *Journal of International Economics 74*, 2–20.

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

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

Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics 99*, 478–485.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics 36*, 705–713.

Jochmans, K. (2021). Bias in instrumental-variable estimators of fixed-effect models for count data. Mimeo.

Larch, M., J. Wanner, Y. V. Yotov, and T. Zylkin (2019). Currency unions and trade: A PPML reassessment with high-dimensional fixed effects. *Oxford Bulletin of Economics and Statistics 81*, 487–510.

Lawless, M. (2009). Firm export dynamics and the geography of trade. *Journal of International Economics 77*, 1149–1172.

Morales, E., G. Sheu, and A. Zahler (2019). Extended gravity. *Review of Economic Studies 86*, 2668–2712.

Mullahy, J. (1997). Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics 79*, 586–593.

Newman, M. E. J. (2010). *Networks: An Introduction.* Oxford University Press.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Pfaffermayr, M. (2019). Gravity models, PPML estimation and the bias of the robust standard errors. *Applied Economics Letters 26*, 1467–1471.

Powell, D. and S. Seabury (2018). Medical care spending and labor market outcomes: Evidence from workers compensation reforms. *American Economic Review 108*, 2995–3027.

Redding, S. and T. Venables (2004). Economic geography and international inequality. *Journal of International Economics 62*, 53–82.

Rose, A. K. (2000). One money, one market: The effect of common currencies on trade. *Economic Policy 15*, 7–45.

Rose, A. K. (2004). Do we really know that the WTO increases trade? *American Economic Review 94*, 98–114.

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

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

Santos Silva, J. M. C. and S. Tenreyro (2010). Currency unions in prospect and retrospect. *Annual Review of Economics 2*, 51–74.

Sun, K., D. J. Henderson, and S. C. Kumbhakar (2011). Biases in approximating log production. *Journal of Applied Econometrics 26*, 708–714.

Tenreyro, S. (2007). On the trade impact of nominal exchange rate volatility. *Journal of Development Economics 82*, 485–508.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics 84*, 129–154.

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

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

Weidner, M. and T. Zylkin (2021). Bias and consistency in three-way gravity models. *Journal of International Economics 132*, 103513.

Werner, T. (2015). Gaining access by doing good: The effect of sociopolitical reputation on firm participation in public policy making. *Management Science 61*, 1989–2011.

Windmeijer, F. A. G. and J. M. C. Santos Silva (1997). Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics 12*, 281–294.

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

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics 182*, 226–234.