UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1994

INSTRUMENTAL-VARIABLE ESTIMATION OF GRAVITY EQUATIONS

Koen Jochmans Vincenzo Verardi

11 November 2019

We present an instrumental-variable approach to estimate gravity equations. Our procedure accommodates the potential endogeneity of policy variables and is fully theory-consistent. It is based on the model in levels and accounts for multilateral resistance terms by means of importer and export fixed effects. The implementation is limited-information in nature, and so is silent on the form of the mechanism that drives the actual policy decisions. The procedure spawns specification tests for the validity of the instruments used as well as a test for exogeneity. We estimate gravity equations from five cross-sections of bilateral-trade data where the policy decision of interest is the engagement in a free trade agreement. We rely on the interaction of the countries in the pair with third-party trading partners to construct a credible instrumental variable based on the substantial transitivity in the formation of trade agreements that is observed in the data. This instrument is strongly correlated with the policy variable. Our point estimate of the average impact of a free trade agreement increases over the sampling period, starting at 61% and clocking o_ at a 117% increase in bilateral trade volume. Not correcting for endogeneity yields stable estimates of around 25%.

INSTRUMENTAL-VARIABLE ESTIMATION OF GRAVITY EQUATIONS

Koen Jochmans*
University of Cambridge

Vincenzo Verardi†
Université de Namur

November 11, 2019

Abstract

We present an instrumental-variable approach to estimate gravity equations. Our procedure accommodates the potential endogeneity of policy variables and is fully theory-consistent. It is based on the model in levels and accounts for multilateral resistance terms by means of importer and export fixed effects. The implementation is limited-information in nature, and so is silent on the form of the mechanism that drives the actual policy decisions. The procedure spawns specification tests for the validity of the instruments used as well as a test for exogeneity. We estimate gravity equations from five cross-sections of bilateral-trade data where the policy decision of interest is the engagement in a free trade agreement. We rely on the interaction of the countries in the pair with third-party trading partners to construct a credible instrumental variable based on the substantial transitivity in the formation of trade agreements that is observed in the data. This instrument is strongly correlated with the policy variable. Our point estimate of the average impact of a free trade agreement increases over the sampling period, starting at 61% and clocking off at a 117% increase in bilateral trade volume. Not correcting for endogeneity yields stable estimates of around 25%.

JEL Classification: C26, F14

Keywords: endogeneity, fixed effects, gravity equation, instrumental variable, multilateral resistance, free trade agreement.

*E-mail: kj345@cam.ac.uk. Address for correspondence: Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. Financial support from the European Research Council through Starting Grant n° 715787 is gratefully acknowledged.

†E-mail: vverardi@unamur.be. Financial support from the FNRS is gratefully acknowledged.

1 Introduction

The gravity equation has a long history. 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 analysis of bilateral trade patterns. The recent literature has made great strides towards credible estimation of the gravity equation; see Head and Mayer (2014) for a survey. The preferred specification features importer and exporter fixed effects to control for multilateral resistance terms (Anderson 1979, Anderson and van Wincoop 2003) and is estimated in levels rather than in log-linearized form (see Redding and Venables 2004 and Santos Silva and Tenreyro 2006, respectively).

An important issue that arises when taking this preferred model to the data is how to handle the (potential) endogeneity of policy variables, such as membership of a currency union or participation in preferential trade agreements. Indeed, the data suggest that countries take such 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). The resulting reverse causality would invalidate the estimates of all the model parameters as well as the consequent policy experiments and welfare calculations. While the literature has long since recognized this problem (see, e.g., Rose 2004), coming up with a satisfactory solution has proven difficult.

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). Such an approach may not be satisfactory for the following reasons. First, the existing 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, estimation from short panel data leads to estimators that are asymptotically biased (Weidner and Zylkin, 2019). Consequently, the hypothesis tests and confidence intervals reported are invalid. Fourth, if current policy variables react to existing trade flows, the policy variables are not (strictly) exogenous. This renders any fixed-effect estimator inconsistent.

An alternative to relying on panel data would be to resort to an instrumental-variable approach. Egger, Larch, Staub and Winkelmann (2011) have taken a full-information view, simultaneously estimating the gravity equation together with an explicit parametrization of the mechanism that underlies trade policy decisions.[1,2] Such an approach requires strong parametric assumptions, some of which are non-refutable and appear to be in conflict with some stylized facts, such as the (conditional) heteroskedasticity of trade flows (Santos Silva and Tenreyro, 2006). Furthermore, although consistent (under the assumption of correct specification), the parameter estimates in both equations are again asymptotically biased, invalidating any inferential statements based on them.

Here we present a limited-information approach to instrumental-variable estimation of the gravity equation. Our strategy is to construct a set of orthogonality conditions that simultaneously difference-out the importer and exporter fixed effects and control for the endogeneity of policy variables. It allows to remain agnostic about how currency unions or preferential trade agreements are formed and is fully compatible with the theoretical foundations underlying the gravity equation. In this sense, our approach is like two-stage least squares, although it is designed for the model in levels, which is nonlinear. We note that the existence of orthogonality conditions in this setting is not immediate, as it is well known that dealing with endogeneity is difficult in nonlinear models, even without the presence of fixed effects (Blundell and Powell, 2003). We use our moment conditions to construct an estimator that can be understood to be an instrumental-variable extension of Jochmans (2017).

[1] There is earlier work on instrumental-variable estimation of the gravity equation; see Rose (2000) and Barro and Tenreyro (2007), for example. However, these analyses are all based on log-linearized models of trade and do not account for multilateral resistance.

[2] Given the robustness of the pseudo-Poisson estimator to the inclusion of importer and exporter effects one might be tempted to resort to its instrumental-variable version as discussed in Mullahy (1997) and Windmeijer and Santos Silva (1997). The statistical properties of this estimator, however, break down when fixed effects are included.

The variable feared to be endogenous in our specification is a dummy for the presence of a free trade agreement. There is precedent in the search for suitable instruments, although with relatively little success. Head and Mayer (2014, p. 162) note that most variables that plausibly cause trade agreements also appear in the trade equation itself. Rose (2004, p. 110) experimented with measures of democracy and polity, and measures of freedom, civil rights and political rights, but reported that these are only quite weak instruments. An apparently unexplored direction in the quest for instruments is to recognize that decisions on bilateral trade are not made in isolation. We find high levels of transitivity in the formation of free trade agreements in our 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). In the data that we look at, the percentage of trade flows subject to a trade agreement is never more than 0.6%. This number increases to over 80% if we focus on country pairs in which the countries have such an agreement with at least one common third party. 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 country-specific variables, i.e., through the importer and exporter effects. This is fully consistent with the structural gravity model (Anderson and van Wincoop 2003, Anderson and Yotov 2010), where the multilateral resistance terms absorb the impact of trade agreements with third-country trading partners. Hence, the validity of our instrument is theoretically grounded. It is also supported by overidentification tests that we report on below.

We apply our estimator to data from Helpman, Melitz and Rubinstein (2008), which cover multiple years. Point estimates of the distance elasticity and border effect take on conventional values. Our estimate of the average impact of a free trade agreement increases over time, being 61% in the initial year and 116% in the last year. In contrast, not correcting for endogeneity yields estimates of around 25% that do not change much over time. Given identification through our instrumental variable we can test whether free trade agreements can be considered endogenous in a statistical sense. We find increasingly strong evidence for this as we move toward the end of the sampling period, with the p-value for the null of exogeneity clocking off at 0.06.

2 Specification and approach

We have cross-sectional data on bilateral trade between n countries. Let $t_{i,j}$ denote the trade flow from exporter i to importer j and let $\boldsymbol{x}_{i,j}$ be a p-vector of covariates that capture trade costs between i and j. These covariates invariably include measures such a geographical distance, common border and language dummies, and an indicator of the existence of preferential trade agreements.

Model for trade flows The gravity equation of Anderson and van Wincoop (2003) states that

$$t_{i,j} = \exp(\alpha_{i,n} + \gamma_{j,n} + \boldsymbol{x}_{i,j}^{\top}\boldsymbol{\beta})\, \varepsilon_{i,j}, \tag{2.1}$$

where $\alpha_{i,n}$ and $\gamma_{j,n}$ are, respectively, exporter and importer effects and $\varepsilon_{i,j}$ is an unobserved error term. The importer and exporter effects are explicitily indexed by the sample size n to highlight that they capture multilateral resistance and depend on the country's interaction with third-party trading partners. The primary object of interest in (2.1) is the vector $\boldsymbol{\beta}$. Throughout, the importer and exporter effects will be treated as fixed (and, hence, are conditioned upon when taking expectations). In the structural gravity model the $\alpha_{i,n}$ and $\gamma_{j,n}$ may be of use in their own right. Estimates of these that satisfy such model's equilibrium constraints can be obtained in the usual manner (Failly, 2015) once an estimator of $\boldsymbol{\beta}$ has been constructed.[3]

[3] With $\hat{\boldsymbol{\beta}}$ denoting a consistent estimator of $\boldsymbol{\beta}$, the importer and exporter effects can be obtained via a pseudo-Poisson regression of the normalized trade flows $t_{i,j}/\exp(\boldsymbol{x}_{i,j}^{\top}\hat{\boldsymbol{\beta}})$ on a set of importer and exporter dummies. When $\hat{\boldsymbol{\beta}}$ converges at the rate n^{-1} the (robust) standard errors from this pseudo-Poisson routine are valid in large samples.

The standard approach to estimation proceeds under the assumption that the errors are (conditionally) independent and that

$$\mathbb{E}(\varepsilon_{i,j}|\boldsymbol{x}_{1,2},\ldots,\boldsymbol{x}_{n,n-1}) = 1. \tag{2.2}$$

This mean-independence assumption validates the use of the pseudo-Poisson estimator (Gouriéroux, Monfort and Trognon, 1984) advocated by Santos Silva and Tenreyro (2006) and the estimator of Jochmans (2017). While the identifying restriction (2.2) is plausible for geographical characteristics that facilitate trade—such as distance or having a common border—it is difficult to maintain for most policy variables, such as the establishment of preferential trade agreements; see Rose (2004) and Baier and Bergstrand (2007, 2009), for example.

Orthogonality conditions An instrumental-variable version of (2.2) would take the form

$$\mathbb{E}(\varepsilon_{i,j}|\boldsymbol{z}_{1,2},\ldots,\boldsymbol{z}_{n,n-1}) = 1, \tag{2.3}$$

where $\boldsymbol{z}_{i,j}$ is a q_1-vector of instrumental variables. However, an estimator based on it has not been proposed. We do so here.

To see how (2.3) can be used to construct an estimator of $\boldsymbol{\beta}$ we introduce the shorthand

$$u_{i,j}(\boldsymbol{\beta}) := \frac{t_{i,j}}{\exp(\boldsymbol{x}_{i,j}^\top\boldsymbol{\beta})},$$

which is known up to $\boldsymbol{\beta}$. Consider two exporter-importer pairs, (i,j) and (i',j'). Equation (2.3), together with the absence of serial correlation in the errors, immediately implies that

$$\mathbb{E}\left(u_{i,j}(\boldsymbol{\beta})\,u_{i',j'}(\boldsymbol{\beta})\right|\boldsymbol{z}_{1,2},\ldots,\boldsymbol{z}_{n,n-1}) = \exp(\alpha_i+\alpha_{i'}+\gamma_j+\gamma_{j'}).$$

Next consider the pairs (i,j') and (i',j), which involve the same countries but concern different trade flows. Then, again,

$$\mathbb{E}\left(u_{i,j'}(\boldsymbol{\beta})\,u_{i',j}(\boldsymbol{\beta})\right|\boldsymbol{z}_{1,2},\ldots,\boldsymbol{z}_{n,n-1}) = \exp(\alpha_i+\alpha_{i'}+\gamma_j+\gamma_{j'}).$$

The right-hand side in each of these equations is identical. Consequently, taking differences,

$$\mathbb{E}\left(u_{i,j}(\boldsymbol{\beta})\,u_{i',j'}(\boldsymbol{\beta}) - u_{i,j'}(\boldsymbol{\beta})\,u_{i',j}(\boldsymbol{\beta})\right|\boldsymbol{z}_{1,2},\ldots,\boldsymbol{z}_{n,n-1}) = 0.$$

This conditional moment condition implies unconditional moment conditions of the form

$$\mathbb{E}\left(\tilde{\boldsymbol{z}}_{i,j,i',j'}\left\{u_{i,j}(\boldsymbol{\beta})\,u_{i',j'}(\boldsymbol{\beta}) - u_{i,j'}(\boldsymbol{\beta})\,u_{i',j}(\boldsymbol{\beta})\right\}\right) = \boldsymbol{0}, \tag{2.4}$$

where $\tilde{\boldsymbol{z}}_{i,j,i',j'}$ is a q_2-vector of transformations of the q_1-vectors $\boldsymbol{z}_{1,n}, \ldots, \boldsymbol{z}_{n,n-1}$. Equation (2.4) is an orthogonality condition in the same spirit as the usual normal equations for two-stage least squares, but it accounts for the nonlinearity of the model as well as for the presence of fixed effects.[4] A generalized method-of-moments (GMM) estimator based on (2.4) can be constructed in a similar fashion as in Jochmans (2017). To maintain focus we relegate details to a later section.

Related work Our limited-information view is to be contrasted with the full-information route of Egger, Larch, Staub and Winkelmann (2011). Their approach is to endogenise the decision to establish a preferential trade agreement by complementing (2.1) with an explicit binary-choice model for it. A tight parametrization of the unobservables allows to estimate this system in two steps. An obvious limitation of this approach is that it requires the whole system of equations to be correctly specified. This is problematic as, for example, their assumptions are at odds with the stylized fact that trade data are (conditionally) heteroskedastic (Santos Silva and Tenreyro, 2006). In any event, even under correct specification, the presence of fixed effects implies that the estimators implemented by Egger, Larch, Staub and Winkelmann (2011) are asymptotically biased, so that any inference procedure based on them is incorrect.[5]

[4]The derivation of our moment condition continues to go through when the errors are correlated at the importer and/or exporter level. The implied estimator remains consistent. It is, however, not clear how to obtain cluster-robust standard errors under this type of dependence. We stress that the use of such clustered standard errors for pseudo-Poisson, although not uncommon in practice, is not theoretically grounded.

[5]The problem stems from the presence of fixed effects in the first-stage probit specification. The probit estimator is used to construct an auxiliary regressor for the second-stage pseudo-Poisson estimator but is asymptotically biased. The bias carries over to the second-stage estimator. The same issue would 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.

The moment conditions in (2.4) do not require a model for how preferential trade agreements come about. They also do not restrict the (conditional) distribution of the error term, allowing for heteroskedasticity of arbitrary form, for example. Furthermore, (2.4) is free of importer and exporter effects; they have effectively been differenced-out. Moreover, our orthogonality conditions separate estimation of the fixed effects from inference on the parameter of interest. This does not only eliminate the need to solve a high-dimensional optimization problem. It also prevents the associated standard errors from suffering from the large bias that is observed in those of the pseudo-Poisson estimator, which result from the estimation noise in the importer and exporter effects (Jochmans 2017, Pfaffermayer 2019).

3 Data, instruments, and results

Bilateral-trade data The data we use are taken from Helpman, Melitz and Rubinstein (2008). They are panel data on trade between 158 countries. We will estimate gravity equations for the years 1985–1989, but will use data going back to 1981. The left plot in Figure I shows the evolution of the average volume of bilateral trade over this period. As discussed in Helpman, Melitz and Rubinstein (2008), the increase in trade volume over time was mostly driven by the growth of trade between countries that traded with each other in both directions at the beginning of the sampling period, and not be the creation of new trade partnerships. The regressors that we will use are standard. They are (i) 'distance', the log distance between the capitals of the respective countries (in kilometers); (ii) 'colony', a dummy indicating whether one of the countries in the dyad ever colonized the other; (iii) 'border', a dummy representing the existence of a common physical border between both countries; (iv) 'language', a dummy capturing if both countries share a common language; (v) 'fta', a dummy measuring whether the countries belong to a common regional trade agreement. Table I contains descriptive statistics for all but the last regressor. Note that these variables do not change over time. The percentage of trade flows that benefit from a preferential trade agreement on the other hand has steadily increased over the sampling

Figure I: Evolution of 'trade', 'fta', and 'common fta'

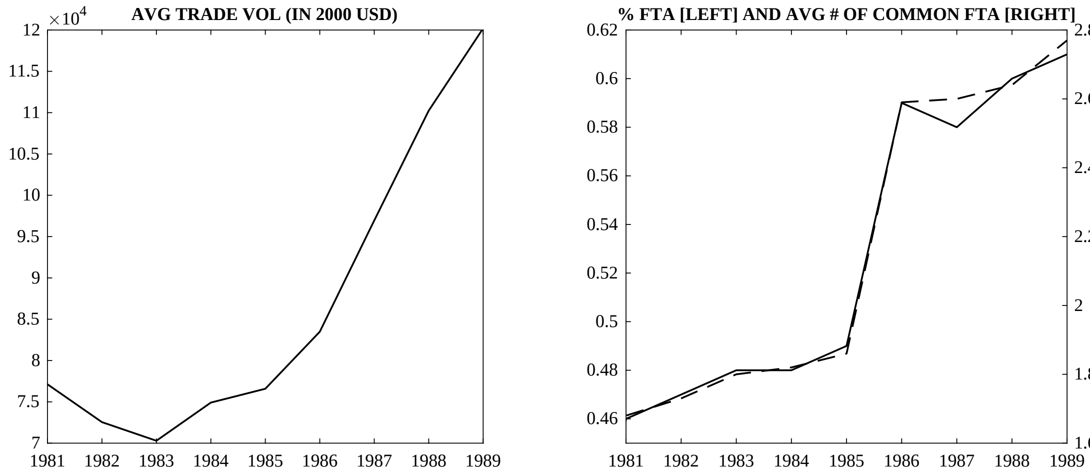

Figure notes. Left plot: Average bilateral trade volume by year. Right plot: percentage of trade flows that benefit from a preferential trade agreement (solid line; measured against the left axis) and average number of common free trade agreement partners (dashed-dotted line; measured against the right axis) by year.

period. The full line in the right plot of Figure I shows this evolution (against the left axis). While only 0.46% of all trade flows were covered by an agreement in 1981, this number reached .61% by 1989. We refer to Helpman, Melitz and Rubinstein (2008, Appendix I) for additional details on the data and on their construction.

Instrument selection Although the literature has recognized the endogeneity of policy variables in the gravity equation, finding good instruments has proven difficult. On the one hand, most variables that plausibly cause trade agreements also appear in the trade

Table I: Descriptive statistics of regressors (constant across all years)

variable	description	mean	std	min	max
distance	log distance between capitals (in km)	4.177	0.781	-0.151	5.661
colony	former colonial tie dummy	0.100	0.098	0	1
border	common border dummy	0.017	0.131	0	1
language	common language dummy	0.286	0.452	0	1

equation itself (Head and Mayer, 2014, p. 162). On the other hand, as noted by Rose (2004, p. 110) variables such as measures of democracy and polity, or measures of freedom, civil rights and political rights are typically only weakly-correlated with the policy variables that are feared to be endogenous.

The variables just mentioned are all bilateral in nature, in that they do not take into account the relationship of the exporter and importer with their other trading partners. However, networks typically feature a high degree of transitivity (see, e.g., Newman 2010, pp. 198–204 for a definition and discussion). Recall from Figure I that the probability that a randomly-drawn exporter-importer pair has established a free trade agreement ranges from 0.46% to 0.6%. On the other hand, the probability that this is the case given that both have a free trade agreement with at least one common third country is around 80% in all years. This suggests that the number of common free trade agreement partners—'common fta', say—is a relevant instrument. The correlation between 'fta' and 'common fta' is always above .80. This large value is to be contrasted with the very small correlations that are found for other potential instrumental variables. In our data, for example, the 'religion' variable used by Helpman, Melitz and Rubinstein (2008) (albeit for different purposes) has a correlation with 'fta' ranging between 0.037 and 0.051.[6] The literature on the formation of trade agreements has acknowledged its strong transitivity (Egger and Larch 2008, Chen and Joshi 2010) but its potential use as an instrumental variable appears to have been overlooked. The right plot in Figure I further shows how, over time, the average number of common free trade agreement partners (dashed line, measured against the right axis) moves closely with the fraction of trade flows (solid line, measured against the left axis) that are subject to a free trade agreement.

To control for importer and exporter effects our moment conditions in (2.4) are based on comparisons of (normalized) bilateral trade flows, $u_{i,j}(\boldsymbol{\beta})$, to trade flows involving a different exporter (i') and a different importer (j'). Consequently, we construct our instrument for

[6]Calculations for measures of political stability in the data of Egger, Larch, Staub and Winkelmann (2011) gave similarly low correlations.

'fta' using a cross-fit procedure, as

$$\sum_{k \neq i', j'} \text{`fta'}_{i,k} \, \text{`fta'}_{k,j},$$

where the reference third countries i' and j' are excluded from the sum.

Identification through instrumental variables is obtained through the assumption that the instrument itself is not driving bilateral trade flows. In our case, this means that the fact that the exporter or importer in a given country-pair has a free trade agreement with a third country affects their trade volume only through their multilateral resistance terms. While any exclusion restriction can be called into question, ours is implied by the structural gravity equation and is, hence, consistent with the theory from which our empirical specification is derived.

We will further exploit the panel structure of our data by augmenting our instrument with further lags. This is useful in gaining efficiency, and so in obtaining more precise coefficient estimates. It also leads to overidentification and thus allows us to construct a Sargan statistic. This statistic can be used to test the validity of our moment conditions. Although we hasten to stress that the interpretation of such a test should be done with care (see, e.g., Newey 1985 or Guggenberger 2012), we find below that this test is generally supportive of our exclusion restriction and, with it, of multilateral resistance as a way to capture third-country effects.

Given the availability of panel data an alternative route to identification that has found applicability elsewhere would be to instrument 'fta' by its own first (and/or further) lag. Here, a case for credible identification along this route appears more difficult to make, and a theoretical model that can be used for guidance seems absent. We calculated incremental Sargan tests for the validity of the first lag (not reported here) and generally found that these tests do not support the validity of lagged levels of 'fta' as an instrument in the current context.

Estimation results Table II contains coefficient estimates for the gravity equation in (2.1) estimated by pseudo-Poisson (PMLE) and our instrumental-variable estimator (IV).

Standard errors are provided in parentheses below the point estimates and p-values for the null that the coefficient in question is zero (against a two-sided alternative) are reported in brackets. We find distance elasticities around $-.65$. The existence of colonial ties and sharing a physical border both are economically and statistically significant factors of trade flows. The point estimates for these effects obtained by IV are somewhat larger than those delivered by PMLE, but the difference is small relative to the standard error. Sharing a common language does not have a significant impact on trade in any of the years. The magnitudes of all these coefficients are in line with those obtained elsewhere in the literature.

The results obtained by PMLE and IV differ most in their estimate of the importance of free trade agreements. While the PMLE coefficient estimates are small and fairly constant over time—with an average of .240 and standard deviation of .029— instrumentation gives coefficient estimates that are both larger and display a (crude) upward trend over time, with an average of .642 and standard deviation of .110. This is summarized visually in Figure II. The left plot gives the coefficient estimates while the right plot contains the corresponding average marginal effect of a free trade agreement. PMLE (x) estimates the latter from 21% in 1985 to 31% in 1989, corresponding to a 10 pp. increase over five years. The instrumental-variable estimates ($*$) are always larger, being 61% in 1985 and 117% in 1989, which is an almost 60 pp. increase over the sampling period. There is, of course, a standard error on our point estimates. The standard errors on IV are larger than on those obtained by PMLE, both because of the instrumentation and the fact that PMLE standard errors tend to be much too small.[7] Dealing with both endogeneity and two-way fixed effects is quite demanding on the data, even more so in highly-unbalanced regressor designs, and reasonable standard errors should reflect this.

[7] For the average marginal effect the larger standard error is also mechanical, as it is partly due to the increase in the estimated coefficient. Indeed, the effect is estimated as $(e^{\hat{\beta}} - 1)$, with $\hat{\beta}$ the coefficient estimate on 'fta'. Its (delta-method) standard error is $e^{\hat{\beta}} \times \text{se}(\hat{\beta})$. This increases with $\hat{\beta}$ even if the standard error on $\hat{\beta}$ remains the same.

Figure II: The impact of 'fta' on trade flows

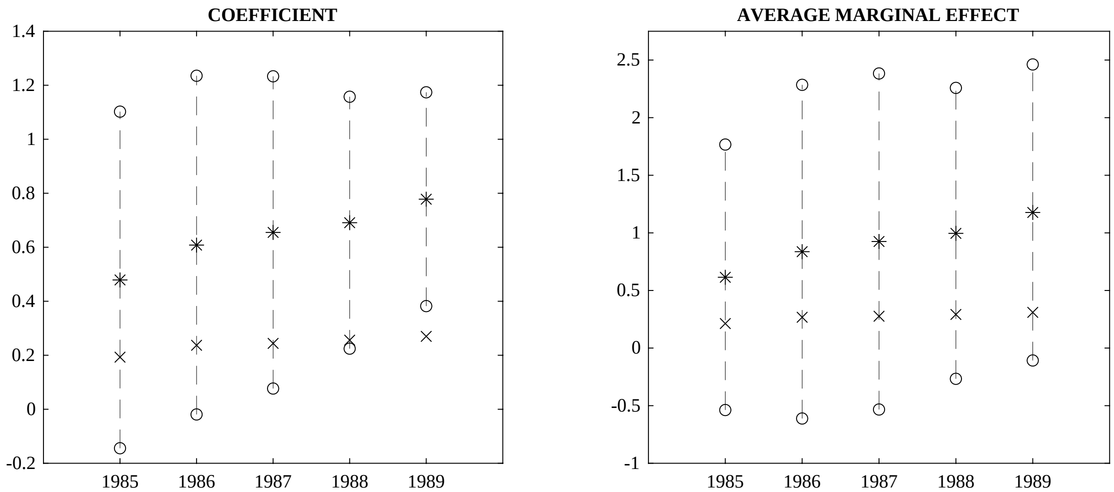

Figure notes: Point estimates ($*$) and 95% confidence intervals ($- - \circ$) obtained by instrumental-variable estimation together with point estimates (x) obtained by pseudo-Poisson.

The lower part of Table II provide the values and associated p-values (in brackets) of two Sargan statistics. Such statistics are a natural by-product of our estimation procedure. The first Sargan statistic, 'exogeneity', can be used to assess whether there is strong statistical evidence that 'fta' is endogenous. The p-value decreases over the sampling period, clocking off at .065, and thus providing less credibility for the null of exogeneity. The second Sargan statistic ('validity') can be used to shed light on the null that our instruments are valid. We do not find strong support against the validity of our instrumental variable in any of the cross sections. Of course, as usual, this interpretation of both Sargan tests is conditional on having achieved identification.

4 Estimator and Monte Carlo

Details on the estimator A sample counterpart to the left-hand side of (2.4) takes the form

$$\boldsymbol{s}_n(\boldsymbol{b}) := \sum_{i=1}^{n} \sum_{j \neq i} \sum_{i' \neq i,j} \sum_{j' \neq i,i',j} \tilde{\boldsymbol{z}}_{i,j,i',j'} \left\{ u_{i,j}(\boldsymbol{b})\, u_{i',j'}(\boldsymbol{b}) - u_{i,j'}(\boldsymbol{b})\, u_{i',j}(\boldsymbol{b}) \right\}, \tag{4.5}$$

Table II: Estimation results

	PMLE					IV				
	1985	1986	1987	1988	1989	1985	1986	1987	1988	1989
distance	-.670	-.640	-.652	-.686	-.664	-.688	-.630	-.633	-.650	-.622
	(.044)	(.043)	(.041)	(.036)	(.035)	(.058)	(.053)	(.050)	(.048)	(.049)
	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
colony	.612	.532	.478	.448	.463	.745	.592	.550	.493	.547
	(.115)	(.106)	(.103)	(.098)	(.098)	(.180)	(.159)	(.152)	(.151)	(.151)
	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.001]	[.000]
border	.745	.771	.747	.625	.621	.977	1.027	1.008	.771	.689
	(.113)	(.112)	(.103)	(.075)	(.076)	(.193)	(.192)	(.184)	(.131)	(.178)
	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]	[.000]
language	-.110	-.061	-.056	-.066	-.055	-.111	-.027	-.029	-.019	.009
	(.092)	(.090)	(.087)	(.088)	(.087)	(.157)	(.152)	(.141)	(.130)	(.129)
	[.231]	[.497]	[.521]	[.457]	[.524]	[.480]	[.859]	[.838]	[.883]	[.945]
fta	.193	.237	.244	.256	.270	.479	.608	.655	.691	.778
	(.088)	(.087)	(.084)	(.077)	(.066)	(.318)	(.320)	(.295)	(.238)	(.202)
	[.000]	[.006]	[.004]	[.001]	[.000]	[.132]	[.058]	[.026]	[.004]	[.000]
exogeneity	—	—	—	—	—	.204	1.077	1.546	2.058	3.405
						[.652]	[.299]	[.214]	[.151]	[.065]
validity	—	—	—	—	—	6.743	7.730	7.742	8.878	9.712
						[.150]	[.172]	[.258]	[.262]	[.286]

Table notes: Robust standard errors in parentheses. p-values in brackets. ‘validity’ is the value of the Sargan test statistic for the null that the instruments are valid. ‘exogeneity’ is the value of the Sargan test statistic for the null that ‘fta’ is exogenous. The instrumental-variable estimator uses all lags of our ‘common fta’ variable ranging back to 1981.

where we assume without loss of generality that the summand is symmetric in i and i' as well as in j and j'. The optimal (two-step) GMM estimator based on this vector of sample moments is equal to

$$\hat{\hat{\boldsymbol{\beta}}} := \arg\min_{\boldsymbol{b}} \boldsymbol{s}_n(\boldsymbol{b})^\top \hat{\boldsymbol{V}}_n^{-1} \boldsymbol{s}_n(\boldsymbol{b}),$$

where $\hat{\boldsymbol{V}}_n$ is an estimator of the asymptotic variance of $\boldsymbol{s}_n(\boldsymbol{\beta})$. To construct this matrix a preliminary estimator of $\boldsymbol{\beta}$ is needed. An intuitive choice is to use the GMM estimator $\hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{b}} \boldsymbol{s}_n(\boldsymbol{b})^\top \boldsymbol{s}_n(\boldsymbol{b})$, i.e., the estimator that assigns the same weight to each moment condition. Without additional prior information this is a natural choice. Standard GMM theory (Hansen, 1982) implies that the large-sample behavior of $\hat{\hat{\boldsymbol{\beta}}}$ does not depend on the first-step estimator used. With this auxiliary estimator at hand we construct the plug-in estimator

$$\hat{\boldsymbol{V}}_n := \sum_{i=1}^{n} \sum_{j \neq i} \boldsymbol{v}_{i,j}(\hat{\boldsymbol{\beta}})\, \boldsymbol{v}_{i,j}(\hat{\boldsymbol{\beta}})^\top,$$

where $\boldsymbol{v}_{i,j}(\boldsymbol{b})$ is defined as

$$\sum_{i' \neq i,j} \sum_{j' \neq i,i',j} \left\{ (\check{\boldsymbol{z}}_{i,j,i',j'} + \check{\boldsymbol{z}}_{i',j',i,j}) - (\check{\boldsymbol{z}}_{i,j',i',j} + \check{\boldsymbol{z}}_{i',j,i,j'}) \right\} \left\{ u_{i,j}(\boldsymbol{b})\, u_{i',j'}(\boldsymbol{b}) - u_{i,j'}(\boldsymbol{b})\, u_{i',j}(\boldsymbol{b}) \right\}.$$

The structure of $\hat{\boldsymbol{V}}_n$ is non-standard because the summands in $\boldsymbol{s}_n(\boldsymbol{b})$ are not independent. It is nonetheless straightforward to construct. We refer to Jochmans (2017) for details and discussion.

Under conventional regularity conditions our GMM estimator is asymptotically normal. An estimator of its covariance matrix is $\hat{\hat{\boldsymbol{\Omega}}}_n := \big(\hat{\hat{\boldsymbol{Q}}}_n^\top \hat{\hat{\boldsymbol{V}}}_n^{-1} \hat{\hat{\boldsymbol{Q}}}_n\big)^{-1} / n(n-1)$, where we denote by $\hat{\hat{\boldsymbol{Q}}}_n$ the Jacobian matrix of the moment conditions at $\hat{\hat{\boldsymbol{\beta}}}$ and $\hat{\hat{\boldsymbol{V}}}_n$ is defined as $\hat{\boldsymbol{V}}_n$ but constructed using $\hat{\hat{\boldsymbol{\beta}}}$ in stead of $\hat{\boldsymbol{\beta}}$. The square-root of the diagonal entries of $\hat{\hat{\boldsymbol{\Omega}}}_n$ provide valid standard errors on $\hat{\hat{\boldsymbol{\beta}}}$.

When we have more moments than parameters to estimate our model is overidentified and the criterion function evaluated at its minimizer satisfies

$$n(n-1)\, \boldsymbol{s}_n(\hat{\hat{\boldsymbol{\beta}}})^\top \hat{\boldsymbol{V}}_n^{-1} \boldsymbol{s}_n(\hat{\hat{\boldsymbol{\beta}}}) \overset{a}{\sim} \chi^2_{q-p}$$

when all moments in (2.4) hold. Consequently, this quantity can be used to test the validity of (some or all of) the moment conditions in the usual way (see Sargan 1958 and Hansen 1982).

Although the multiple sums in the empirical moments in (4.5) may suggest that our estimator is cumbersome to compute, this is not the case. Careful re-arrangement makes the evaluation of the criterion function straightforward in any matrix-based language. An efficient Stata© implementation of our procedure for settings where $\check{\boldsymbol{z}}_{i,j,i',j'} = \check{\boldsymbol{z}}_{i,j}$ is made available through `ssc`[8].

Numerical illustrations We next provide simulation results on the performance of the instrumental-variable estimator. Our design has two covariates $\boldsymbol{x}_{i,j} = (x^1_{i,j}, x^2_{i,j})^\top$ and three instruments $\boldsymbol{z}_{i,j} = (z^1_{i,j}, z^2_{i,j}, z^3_{i,j})^\top$. The design is symmetric in the sense that $\boldsymbol{x}_{i,j} = \boldsymbol{x}_{j,i}$ and $\boldsymbol{z}_{i,j} = \boldsymbol{z}_{j,i}$ for all pairs (i,j). For each pair we draw the first regressor from a lognormal distribution. Hence, $x^1_{i,j}$ is continuous and non-negative, mimicking geographical distance. We next generate the binary covariate via the threshold-crossing rule $x^2_{i,j} = 1\{\boldsymbol{z}^\top_{i,j}\boldsymbol{\gamma} \geq \epsilon_{i,j}\}$, where

$$\begin{pmatrix} \log \varepsilon_{i,j} \\ \log \varepsilon_{j,i} \\ \epsilon_{i,j} \end{pmatrix} \sim \boldsymbol{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \right)$$

for correlation ρ. This setup ensures that the dummy variable is endogenous in (2.1) as long as $\rho \neq 0$. For our three instrumental variables we set $z^1_{i,j} = x^1_{i,j}$ and, as excluded instruments, use $z^2_{i,j} = 1\{1 + x^1_{i,j} < u_{i,j}\}$ and $z^3_{i,j} = 1\{z^2_{i,j} \geq v_{i,j}\}$, where the errors $(u_{i,j}, v_{i,j})$ are bivariate standard normal. Both these instruments are dummies. The first one is generated with a success probability that is decreasing in $x^1_{i,j}$. The second one has success probability .841 when $z^2_{i,j} = 1$ and .500 when $z^2_{i,j} = 0$. These instruments broadly behave like a common border dummy and a common language dummy, respectively. In our simulations we set $\boldsymbol{\gamma} = (-1, 1, 1)^\top$ and $\boldsymbol{\beta} = (-1, 1)^\top$, ensuring that $x^1_{i,j}$ has a negative

[8] The package `ivgravity` can be installed from within Stata© by typing `ssc install ivgravity` in the command window.

Table III: Monte Carlo results for $n = 50$

ρ		mean	median	std	se	rej freq
-0.50	$\hat{\hat{\beta}}_1$	-1.0011	-1.0008	0.0317	0.0331	0.0444
	$\hat{\hat{\beta}}_2$	0.9829	0.9913	0.1796	0.1870	0.0354
	validity	—	—	—	—	0.0552
	exogeneity	—	—	—	—	0.9966
-0.25	$\hat{\hat{\beta}}_1$	-1.0006	-1.0008	0.0300	0.0305	0.0478
	$\hat{\hat{\beta}}_2$	0.9869	0.9928	0.1783	0.1789	0.5910
	validity	—	—	—	—	0.0489
	exogeneity	—	—	—	—	0.5836
0	$\hat{\hat{\beta}}_1$	-1.0016	-1.0027	0.0273	0.0282	0.0446
	$\hat{\hat{\beta}}_2.$	0.9791	0.9781	0.1755	0.1774	0.0474
	validity	—	—	—	—	0.0534
	exogeneity	—	—	—	—	0.0444
0.25	$\hat{\hat{\beta}}_1$	-1.0021	-1.0032	0.0262	0.0263	0.0588
	$\hat{\hat{\beta}}_2$	0.9741	0.9667	0.1872	0.1815	0.0592
	validity	—	—	—	—	0.0472
	exogeneity	—	—	—	—	0.4622
0.50	$\hat{\hat{\beta}}_1$	-1.0027	-1.0037	0.0246	0.0246	0.0560
	$\hat{\hat{\beta}}_2$	0.9686	0.9566	0.1934	0.1920	0.0584
	validity	—	—	—	—	0.0506
	exogeneity	—	—	—	—	0.9836

Table notes: All results obtained over 10,000 Monte Carlo replications. True values: $\boldsymbol{\beta} = (-1, 1)^\top$. Nominal size of all tests is .05.

Table IV: Monte Carlo results for $n = 100$

ρ		mean	median	std	se	rej freq
-0.50	$\hat{\hat{\beta}}_1$	-1.0002	-1.0001	0.0160	0.0161	0.0482
	$\hat{\hat{\beta}}_2$	0.9958	0.9985	0.0906	0.0906	0.0440
	validity	—	—	—	—	0.0490
	exogeneity	—	—	—	—	1.0000
-0.25	$\hat{\hat{\beta}}_1$	-0.9998	-0.9997	0.0148	0.0149	0.0472
	$\hat{\hat{\beta}}_2$	0.9973	0.9978	0.0886	0.0879	0.0482
	validity	—	—	—	—	0.0542
	exogeneity	—	—	—	—	0.9912
0	$\hat{\hat{\beta}}_1$	-1.0007	-1.0011	0.0135	0.0138	0.0500
	$\hat{\hat{\beta}}_2$	0.9925	0.9926	0.0867	0.0877	0.0472
	validity	—	—	—	—	0.0474
	exogeneity	—	—	—	—	0.0490
0.25	$\hat{\hat{\beta}}_1$	-1.0009	-1.0013	0.0128	0.0129	0.0502
	$\hat{\hat{\beta}}_2$	0.9914	0.9891	0.0906	0.0904	0.0518
	validity	—	—	—	—	0.0486
	exogeneity	—	—	—	—	0.9832
0.50	$\hat{\hat{\beta}}_1$	-1.0008	-1.0012	0.0121	0.0122	0.0552
	$\hat{\hat{\beta}}_2$	0.9913	0.9873	0.0988	0.0962	0.0586
	validity	—	—	—	—	0.0510
	exogeneity	—	—	—	—	1.0000

Table notes: All results obtained over 10,000 Monte Carlo replications. True values: $\boldsymbol{\beta} = (-1, 1)^\top$. Nominal size of all tests is .05.

impact on both $t_{i,j}$ and $x^2_{i,j}$. In what follows We instrument by setting $\tilde{\boldsymbol{z}}_{i,j,i',j'} = \boldsymbol{z}_{i,j}$ for all (i', j') in (2.4).

Tables III and IV contain simulation results for our estimator for samples of size $n = 50$ and $n = 100$, respectively, as obtained over 10,000 Monte Carlo replications. The designs vary in the severity of the endogeneity, as governed by $\rho \in \{-.50, -.25, 0, .25, .50\}$. The tables contain the mean, median, and standard deviation of the point estimates, together with the average estimated standard errors and the rejection frequency of the two-sided t-test for the null that the coefficient in question is equal to its true value. We also provide the rejection frequency of Sargan's overidentification test (that tests whether the instruments are valid) and the exogeneity test. All tests were implemented with critical values corresponding to a 5% significance level.

Our estimator is close to unbiased for all designs and both sample sizes considered. The estimated standard errors of our estimator also perform well, being close (on average) to the actual standard deviations over the Monte Carlo replication. They tend to be slightly too large here when $\rho < 0$ and slightly too small when $\rho > 0$. Nonetheless, the bias is small in magnitude, and the t-tests based on them provide reliable inference throughout. The same can be concluded for both Sargan tests. Although we do not report it here (as it was not designed for this situation) the pseudo-Poisson estimator suffers from substantial bias for all non-zero values of ρ. The t-tests based on this estimator almost always reject, making them an unreliable tool for inference in situations where endogeneity is feared to be present.

5 Conclusion

We have introduced an instrumental-variable approach to estimate the gravity equation of Anderson and van Wincoop (2003). Our procedure is meant to accommodate the potential endogeneity of policy variables and is fully theory-consistent, in the sense of Head and Mayer (2014). They are based on the model in levels and account for multilateral resistance terms by means of importer and exporter fixed effects. The implementation is limited-information in nature, and so is silent on the determinants that drive the actual policy decisions. A Stata© implementation of our estimator is available.

We estimate gravity equations for multiple cross-sections of bilateral-trade data where the policy decision of interest is the engagement in a free trade agreement. We rely on the interaction of the countries in the pair with third-party trading partners to construct a credible instrumental variable based on the substantial transitivity in the formation of trade agreements that is observed in the data. This instrument is strongly correlation with the policy variable and consistent with models of structural gravity. Our estimate of the average marginal effect of a free trade agreement varies over time, with values ranging from 61% to 117%, implying more than a doubling of trade volume. This puts trade agreements broadly on par, in terms of magnitude of their impact, with sharing a common border or having a colonial history. In contrast, not correcting for endogeneity yields estimates of around 25%.

References

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.

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.

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

Blundell, R. W. and J. L. Powell (2003). Endogeneity in nonparametric and semiparametric regression models. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances In Economics and Econometrics*, Volume II, Chapter 8, pp. 312–357. Cambridge University Press.

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

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.

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.

Failly, T. (2015). Structural gravity and fixed effects. *Journal of International Economics 97*, 76–85.

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

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 (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica 52*, 701–720.

Guggenberger, P. (2012). A note on the (in)consistency of the test of overidentifying restrictions and the concepts of true and pseudo-true parameters. *Economics Letters 117*, 901–904.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

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.

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

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.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics 29*, 229–256.

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

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

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. 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.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*, 393–415.

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

Weidner, M. and T. Zylkin (2019). Bias and consistency in three-way gravity models. Mimeo.

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.