NBER WORKING PAPER SERIES

GOLD STANDARD GRAVITY

James E. Anderson
Yoto V. Yotov

Working Paper 17835
http://www.nber.org/papers/w17835

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2012

A much earlier version of a portion of this work was presented at the NBER ITI meetings, Spring 2010 and the Venice Trade Costs conference, June 2010. We thank participants for helpful comments, especially Dave Donaldson, Keith Head and Michael Waugh. We also thank Arthur Lewbel for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Gold Standard Gravity
James E. Anderson and Yoto V. Yotov
NBER Working Paper No. 17835
February 2012
JEL No. F1

ABSTRACT

This paper provides striking confirmation of the restrictions of the structural gravity model of trade. Structural forces predicted by theory explain 95% of the variation of the fixed effects used to control for them in the recent gravity literature, fixed effects that in principle could reflect other forces. This validation opens avenues to inferring unobserved sectoral activity and multilateral resistance variables by equating fixed effects with structural gravity counterparts. Our findings also provide important validation of a host of general equilibrium comparative static exercises based on the structural gravity model.

James E. Anderson
Department of Economics
Boston College
Chestnut Hill, MA 02467
and NBER
james.anderson.1@bc.edu

Yoto V. Yotov
Drexel University
LeBow College of Business
Department of Economics and International Business
Matheson Hall, Suite 503-C
Philadelphia, PA 19104
yotov@drexel.edu

The extensive gravity model literature moved from pulp fiction to high brow shelves with the development of the structural gravity model by Anderson and van Wincoop (2003) and its success in explaining the border puzzle posed by McCallum (1995). This paper provides the first empirical test of structural gravity. The results are a gold standard benchmark.[1] Structural gravity forces account for 95% of variation in product/importer/time and product/exporter/time fixed effects estimated from empirical gravity equations for 18 manufacturing sectors and 76 countries from 1990-2002. Similar results are found in a robustness check on different and perhaps special data for 28 goods and services sectors in Canada's provinces from 1997-2007: 96% of variation is explained. These results provide an empirical justification for comparative static applications of structural gravity.[2] Perhaps more important, they justify inference of unobservable multilateral resistances and unobservable or distrusted sales and expenditure variables from estimated fixed effects and structural gravity restrictions.

Gravity estimation following Feenstra (2004) usually features importer and exporter country fixed effects as controls in trade flow equations. Econometric problems of exogeneity and omitted variables are demolished when fixed effects replace the theoretically indicated size and multilateral resistance variables. Another potential advantage is that this specification is agnostic as to whether the fixed effects are explained solely or at all by the structural gravity forces. Bilateral trade cost proxies such as distance have consistently estimated coefficients; but if the structural model is valid, much structural information is lost — the bank building is blown up to get at the safe inside.

The methodological novelty of this paper is to compare the patterns of fixed effects in the rubble to a separate reconstruction of structural patterns predicted by theory. Estimated

[1] "The gold standard" is a pervasive metaphor in medical and health research for the most certain medical knowledge or best test, the meaning we intend. The metaphor suggests the highly probable fixity of exchange rates in the gold standard era. The connotation of crisis when the test is failed is not intended.

[2] Comparative static applications start with Anderson and van Wincoop's own comparative static experiment with the effect of removing the Canada-US border barrier, and have proliferated in the subsequent literature, especially with variations on the Eaton-Kortum (2002) model that nests structural gravity within a Ricardian production model.

bilateral trade costs are combined with independent data on total sales and expenditures to construct a facsimile of the theoretical structural gravity model. If structural gravity theory is right, these constructs should equal the estimated fixed effects. The results reported below show that pure structural gravity forces explain almost all of the variation in estimated fixed effects generated from gravity regressions.

The very high goodness of fit was a big surprise to us. The structural gravity forces include multilateral resistance terms solved from highly nonlinear equation systems derived from structural gravity theory. The equations use shipments data, expenditure data and bilateral trade cost estimates that are all measured with error. In principle the directional (importer and exporter) product/country/time fixed effects could be generated by an agnostic model containing many other variables not in the structural gravity model. Even if the structural model were valid, these considerations formed our prior belief that the test was risky in Popper's (1963) sense. Popper's riskiness criterion implies that the result is an impressive validation of the model.

Our reconstruction method is applied to gravity in this paper, but is presumably more widely applicable when fixed effects are used to estimate in the context of a structural model. It may often be possible to analyze estimated fixed effects to reveal information about the structure from which they fell.

Section 1 sets out the structural gravity model and its empirical implementation. Section 2 presents the goodness of fit comparisons. Section 3 subjects the results to robustness checks. Section 4 concludes with drawing out implications for future research. Appendix A describes the data used for the main results. Appendix B describes the Canadian data used in a robustness check.

1 The Structural Gravity Model

Anderson (1979) derives the structural gravity model under the assumptions of identical preferences represented by a globally common Constant Elasticity of Substitution (CES) sub-utility function and product differentiation by place of origin (Armington). Bergstrand (1989) shows that this model nests inside a monopolistic competition structure that determines the size of total shipments in each originating country. Anderson and van Wincoop (2004) argue that under technical assumptions that generate trade separability, structural gravity can nest inside a wide variety of general equilibrium models that determine the size of sales and expenditures in each country, the role of gravity being to determine the distribution pattern of given total sales and expenditures.[3]

At the sectoral level (Anderson and van Wincoop, 2004) the resulting model is:

$$X_{ij}^k = \frac{E_j^k Y_i^k}{Y^k} \left(\frac{T_{ij}^k}{P_j^k \Pi_i^k} \right)^{1-\sigma_k} \tag{1}$$

$$(\Pi_i^k)^{1-\sigma_k} = \sum_j \left(\frac{T_{ij}^k}{P_j^k} \right)^{1-\sigma_k} \frac{E_j^k}{Y^k} \tag{2}$$

$$(P_j^k)^{1-\sigma_k} = \sum_i \left(\frac{T_{ij}^k}{\Pi_i^k} \right)^{1-\sigma_k} \frac{Y_i^k}{Y^k}, \tag{3}$$

where X_{ij}^k denotes the value of shipments at destination prices from origin i to destination j in goods class k. E_j^k is the expenditure at destination j on goods in k from all origins. Y_i^k denotes the sales of goods k at destination prices from i to all destinations, while Y^k is the total output, at delivered prices, of goods k. $T_{ij}^k \geq 1$ denotes the variable trade cost factor on shipment of commodities from i to j in class k, and σ_k is the elasticity of substitution

[3] Anderson (2011) shows that exactly the same system as (1)-(3) below can be derived from two other foundations that feature selection of heterogeneous agents on an extensive margin. One is a model of heterogeneous buyers and the other is a combined production and distribution model of heterogeneous sellers (Eaton and Kortum, 2002) in a Ricardian production framework. The key assumption leading to the 'as if' CES structure is that the heterogeneous agents are distributed according to the Type II (Frechet) extreme value distribution. In these interpretations, the dispersion parameter of the distribution is equivalent to the elasticity of substitution minus 1 in the CES interpretation.

across goods in k. Finally, Π_i^k and P_j^k are multilateral resistance (MR) terms (Anderson and van Wincoop, 2003) that are theoretically derived average outward and inward resistance to shipments toward all destinations and from all origins, respectively.

Multilateral resistance is not observable, but it can be estimated in a two step process based on estimating a stochastic version of (1) by using exporter and importer fixed effects to control for $Y_i^k(\Pi_i^k)^{\sigma_k-1}$ and $E_j^k(P_j^k)^{\sigma_k-1}$. $(T_{ij}^k)^{1-\sigma_k}$ is estimated using proxies for bilateral trade costs. The second step is to solve (2)-(3) for the multilateral resistances, given the Y_i^ks and E_j^k along with the estimated $(T_{ij}^k)^{1-\sigma_k}$s. The system only solves for the multilateral resistances up to a normalization, which may be chosen for convenience (Anderson and Yotov, 2010a), as its value is irrelevant to the methods below.[4]

The novelty of structural gravity theory is the multilateral resistance variables.[5] Equations (2)-(3) derive from world market clearance equations (one for the national variety of each country in each product line) and national budget constraints (one for each country in each product class). See Anderson and Yotov (2010a,b) for detailed analysis of calculated multilateral resistances from estimated gravity equations. These applications reveal large variation in multilateral resistance across product lines, countries and direction of trade. Because multilateral resistance can be interpreted as sellers' (for outward) and buyers' (for inward) incidence of all trade costs and because these trade costs are large themselves, it is important to know how believable are estimated multilateral resistances such as those of Anderson and Yotov (2010a,b).

Turning to the estimation of the gravity equations, following the standard practice in the gravity literature, bilateral trade costs are approximated here by a set of observable proxy

[4]The normalization is needed because if $\{\Pi^0, P^0\}$ is a solution to (2)-(3) then so is $\{\lambda\Pi^0, P^0/\lambda\}$ for any $\lambda > 0$. In our methods below, as indeed in (1)-(3) after dividing the elements of (2)-(3) through by the left hand side variable, the Πs and Ps appear as a product, hence λ cancels.

[5]Anderson (1979) noted their presence but did not propose a solution for estimating or calculating them, still less performing comparative statics with them.

variables:

$$T_{ij}^{k\,1-\sigma_k} = e^{\sum_{m=1}^{4}\beta_m^k \ln DIST_{ij}^m + \beta_5^k BRDR_{ij} + \beta_6^k LANG_{ij} + \beta_7^k CLNY_{ij} + \sum_{i=8}^{83}\beta_i^k SMCTRY_{ij}}. \tag{4}$$

Here, $\ln DIST_{ij}^m$ is the logarithm of bilateral distance between trading partners i and j. We follow Eaton and Kortum (2002) to decompose distance effects into four intervals, $m \in [1, 4]$. The distance intervals, in kilometers, are: [0, 3000); [3000, 7000); [7000, 10000); [10000, maximum]. $BRDR_{ij}$ captures the presence of contiguous borders. $LANG_{ij}$ and $CLNY_{ij}$ account for common language and colonial ties, respectively. Finally, $SMCTRY_{ij}$ is a set of country-specific dummy variables equal to 1 when $i = j$ and zero elsewhere.[6] These variables capture the effect of crossing the international border by shifting up internal trade, all else equal. Use of the internal trade dummies has the advantage of exogeneity, in contrast to direct measures of forces that discriminate between internal and international trade.

The next step toward estimation is to use (4) to substitute for the power transform of bilateral trade costs in (1). The Poisson pseudo-maximum-likelihood (PPML) estimator of Santos-Silva and Tenreyro (2007) is used to address the issues of heteroskedasticity and zeroes in bilateral trade flows.[7] The PPML technique is used to estimate the following econometric specification of the gravity model for each class of goods in our sample:

$$\begin{aligned} X_{ij,t}^k &= \exp[\beta_0 + \sum_{m=1}^{4}\beta_m^k \ln DIST_{ij}^m + \beta_5^k BRDR_{ij} + \beta_6^k LANG_{ij} + \beta_7^k CLNY_{ij} + \\ &\quad + \sum_{i=8}^{84}\beta_i^k SMCTRY_{ij} + \eta_{i,t}^k + \theta_{j,t}^k] + \epsilon_{ij,t}^k, \forall i, j, t; \end{aligned} \tag{5}$$

[6]It should be noted that we can only identify country-specific coefficients β_i^k in a panel setting, which has been used to obtain our main results. Lacking observations for enough degrees of freedom, we have to impose common cross country SMCTRY coefficients in our yearly estimates, which are used in the robustness analysis.

[7]The choice of estimator turns out to be irrelevant in practice in our data. Experiments with ordinary least squares (dropping zero bilateral trade observations) and with the Helpman, Melitz and Rubinstein (2008) selection estimator give the same results for relative trade costs. Realizing that gravity equations can only estimate relative trade costs, results are the same after normalization of the trade costs. The irrelevant differences in estimated levels of trade costs arise from differing implicit normalizations of the estimators.

where $\eta_{i,t}^{k}$ denotes the set of time-varying source-country fixed effects that control for the (log of) outward multilateral resistances along with total sales $Y_{i,t}^{k}$, and $\theta_{j,t}^{k}$ denotes the fixed effects that control for the (log of) inward multilateral resistances along with total expenditures $E_{j,t}^{k}$.[8]

The panel data estimation approach allows us to identify separate country-specific estimates of the international border variables, $SMCTRY_{ij}$, an important dimension of heterogeneity as we verify below. A well-known disadvantage of pooling gravity data over consecutive years is "that dependent and independent variables cannot fully adjust in a single year's time" (p.8 Cheng and Wall, 2005). To address this critique, we use four-year lags and employ data for 1990, 1994, 1998, and 2002.

To estimate (5) and to construct the trade cost indexes needed for our test of structural gravity, we use data on trade flows, output, expenditures, bilateral distances, contiguous borders, colonial ties, and common language from Anderson and Yotov (2010b). Their sample covers the period 1990-2002 for 76 countries and 18 manufacturing commodities aggregated on the basis of the United Nations' 3-digit International Standard Industrial Classification (ISIC) Revision 2. Appendix A lists the countries and commodities in the sample and provides details on the data sources and all variables. In addition, as a robustness check, we experiment with an alternative data set covering Canada's provincial trade for 28 sectors, including services, for the period 1997-2007. See Appendix B for the results and Appendix C for description of the data.

Without going into details, we note that the gravity estimates of (5) are reasonable, intuitive and comparable to the aggregate gravity estimates from the existing literature. See Anderson and Yotov (2010b) for a detailed discussion of the properties of the gravity estimates and the multilateral resistances.

[8] Anderson and van Wincoop (2003) use full information methods to estimate the multilateral resistances. Feenstra (2004) advocates the directional, country-specific fixed effects approach. To estimate the effects of the Canadian Agreement on Internal Trade (AIT), Anderson and Yotov (2010) use panel data with time-varying, directional (source and destination), country-specific fixed effects. Olivero and Yotov (forthcoming) formalize their econometric treatment of the MR terms in a dynamic gravity setting.

2 The Fit of Structural Gravity

Structural gravity theory implies that the expectation of the estimated (denoted with a hat) fixed effects $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$ from equation (5) should be equal to the structural gravity term:

$$\eta^k_{i,t} + \theta^k_{j,t} = \ln\left[Y^k_i(\Pi^k_i)^{\sigma_k-1}E^k_{j,t}(P^k_{j,t})^{\sigma_k-1}/Y^k_t\right], \forall i, j, k, t.$$

Here the absence of a hat denotes the theoretical expected value. The main point of this paper is the remarkably close alignment of theory with prediction, empirical gravity is 95% explained by estimated structural gravity forces.

2.1 Simple Fit Measures

The structural gravity term is built up from the estimation of (5) and the restrictions of system (2)-(3). The estimated regression coefficients permit construction of the ($1-\sigma_k$ power transforms of) estimated bilateral trade costs for each trading pair, year and commodity in our sample. These estimated trade costs are used along with data on the E^k_js and Y^k_is in system (2)-(3) to solve for (the $1-\sigma_k$ power transforms of) the multilateral resistance terms. The latter combine with the sales and expenditure data to yield the composite structural gravity term

$$[Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k-1}E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}].$$

Data on the elasticities of substitution σ_ks are not needed for any calculations here because both the bilateral trade costs and the multilateral resistance terms enter the gravity system and our tests powered to $1-\sigma_k$.

The excellent performance of the structural gravity model is revealed by examining the residuals of the estimated fixed effects from their theoretically predicted values:

$$r^k_{ij,t} = \widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t} - \ln\left[Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k-1}E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}\right], \forall i, j, k, t. \quad (6)$$

The residuals $r^k_{ij,t}$ are interpreted as the residuals from a regression of the estimated fixed effects $\widehat{\eta}^k_{i,t}+\widehat{\theta}^k_{j,t}$ on the constructed estimate of the structural gravity term $\ln{[Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k-1}E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}]}$, constraining the slope coefficient to equal 1 and the intercept to equal 0.

By construction the fixed effects estimated from (5), $\widehat{\eta}^k_{i,t}$ and $\widehat{\theta}^k_{j,t}$, are estimated as deviations from the US. Thus, by construction, the composite variable $\ln{[Y^k_i(\widehat{\Pi}^k_i)^{\sigma_k-1}E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}/Y^k_t]}$ is also a deviation from the US value. In principle, differencing from the US cancels out the global scaling variable $1/Y^k_t$ that is otherwise a component of the structural gravity term.

The analysis of variance provides familiar measures of the goodness of fit of structural gravity. Let $V(x)$ denote the variance of the random variable x. The proportion of the variation of fixed effects explained by the structural gravity term is given by $1-V(r)/V(\eta+\theta)$, the simple R^2 of a constructed regression with slope coefficient set equal to 1 and constant term set equal to 0. Table 1 reports constructed simple R^2s for each sector in our sample.[9] As can be seen from panel A of the table, the simple R^2's are very large and mostly above 0.9. The outlier is Coal and Petroleum at 0.62, followed by Apparel at 0.71, and Beverages-Tobacco and Raw Metals at 0.84. The explanation is that these sectors have the worst fitting first-stage gravity regressions (available by request) as well. The relatively poor behavior is attributable to unmeasured asymmetric policy barriers and their movements over time that are prominent for these three sectors. The constructed simple R^2 across all industries is 0.88, and without the problematic sectors, the goodness-of-fit statistic increases to 0.93.

The portion of the variance of the estimated fixed effects $\widehat{\eta}^k_{i,t}+\widehat{\theta}^k_{j,t}$ being explained by the structural gravity term $\ln{[Y^k_i(\widehat{\Pi}^k_i)^{\sigma_k-1}E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}/Y^k_t]}$ is astonishing. Our prior expectations were far more pessimistic, considering the amount of sector-country-time variation in the fixed effects and in the constructed structural gravity terms.[10]

The standard decomposition of variance (applying the theoretical slope coefficient of 1 to

[9] We do not report sector-year goodness-of-fit measures for brevity. However, those statistics (mostly above 0.9) are available by request and are in support of the sectoral findings analyzed here. Furthermore, as shown below, we argue that all time effects are effectively absorbed by the structural gravity term.

[10] The vast majority of the fixed effects are very precisely estimated. At the same time they vary quite a bit across countries and across commodities.

the size and multilateral resistance components) demonstrates that not all the explanatory power of the unit slope regression is due to the size effects. Panel B of Table 1 shows that the multilateral resistance terms $\ln(\widehat{\Pi}_{i,t}^k \widehat{P}_{j,t}^k)^{1-\sigma_k}$ account for between 17% (Machinery) and 39% (Minerals) of the variance of the $\widehat{\eta}_{i,t}^k + \widehat{\theta}_{j,t}^k$s, while the size effect terms $\ln(Y_{i,t}^k E_{j,t}^k)$ account for between 33% (Coal and Petroleum) and 74% (Machinery). The importance of size effects is no surprise based on the large atheoretic gravity literature, but it is notable that here they contribute to explaining the fixed effects in precisely the theoretically predicted form of $Y_i^k E_j^k$, in contrast to the atheoretic practice of using origin and destination GDP with estimated exponents that differ from 1. More importantly, the large portion of variance explained by the multilateral resistance term is striking because prior considerations suggest that it would be risky to use them to predict trade flows. Multilateral resistance is due strictly to structural gravity theory, and is calculated from the solution to the nonlinear system (2)-(3) that places great reliance on all the structural restrictions of the model.

The constructed R^2s from Table 1 fall short of 1 in part due to measurement error in the $E_{j,t}^k$s and $Y_{i,t}^k$s that are drawn from the UNIDO shipments data.[11] To account for the influence of measurement error, we regress the $r_{ij,t}$'s for each sector on time fixed effects ψ_t to control for differing time mean measurement error in the E's and Y's.[12] Additionally, we control for country-specific measurement error with exporter fixed effects ϕ_i and importer fixed effects φ_j along with a sector specific mean error α. The resulting regression equation, estimated with OLS for each sector k, is:

$$\widehat{\eta}_{i,t}^k + \widehat{\theta}_{j,t}^k - \ln\left[E_{j,t}^k (\widehat{P}_{j,t}^k)^{\sigma_k - 1} Y_{i,t}^k (\widehat{\Pi}_{i,t}^k)^{\sigma_k - 1}\right] = \alpha^k + \psi_t^k + \phi_i^k + \varphi_j^k + e_{ij,t}^k, \forall i, j, t; \qquad (7)$$

[11]Theory assumes shipments evaluated at full user prices whereas actual shipments data excludes trade costs paid by users. The exclusion cancels on average in the shares but introduces unknown error to (2)-(3) and thus to estimates of multilateral resistances.

[12]The estimated $\widehat{\psi}_{j,t}^k + \widehat{\phi}_{i,t}^k$ values can be rescaled as factors that shift trade flows of the gravity equation in levels, obtained by exponentiating. These range from around 0.26 to 1.73, plausibly associated with sector fixed effects on the following reasoning. Suppose, plausibly, that the mean measurement error for each goods class k is in proportion to the observable component of global shipments. The range of the estimated country-time fixed effects values is comparable to the range of values implied by the total shipments data $Y^k/[\sum_k Y^k/N]$ where N is the number of sectors. This is $(0.07, 2.51)$ and is stable over time.

where $e^k_{ij,t}$ is a random error term.

Estimation results from equation (7) are omitted for brevity but available by request. Here, we just use the R^2s from the regressions based on (7) to construct the composite goodness-of-fit indexes reported in the first row of panel C in Table 1. For example, we obtain an $R^2 = 0.668$ from (7) for Food and we add the additional explained variation of $\widehat{\eta}^k_{i,t}+\widehat{\theta}^k_{j,t}$ (66.8% of 0.045=1-0.955) to that of the simple $R^2 = 0.955$ from panel A to construct the composite R^2 of 0.985 reported in panel C.

The composite indexes reveal that, in combination, the time, exporter and importer fixed effects contribute to a moderate increase in the explanatory power of structural gravity. Naturally, the contribution of the fixed effects is most pronounced for the problematic sectors. The simple R^2s for Apparel (0.71) and Coal and Petroleum (0.62) increase to 0.91 and 0.80, respectively. On average, across all sectors, adding the additional explained variation by the fixed effects yields a composite R^2 of 0.952: structural gravity explains 95% of the variation in the estimated sector-country-time fixed effects $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$.[13] To demonstrate the performance of the structural gravity model visually, in Figure 1 we plot the kernel density estimated distribution of the r's after correcting for measurement errors. As expected, the residuals are densely clustered around zero, meaning that structural gravity predicts the fixed effects very successfully.

In our next experiment, we show that the time variation of the structural gravity term captures essentially all of the time variation of the fixed effects, an important confirmation due to the substantial time variation in the data. This observation is already implied by the large constructed R^2s from panel A of Table 1, but strengthened by estimating a variant regression based on (7) with time effects only. Composite R^2s based on this specification are reported in row ψ_t of panel C in Table 1. Comparisons between these numbers and the simple R^2s from panel A reveal that the time fixed effects contribute very little, if at all, to the unexplained variation of $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$. The largest difference between the simple R^2s from

[13]The fit improves to 97.6% when the four worst performing sectors (Coal and Petroleum, Apparel, Beverage/Tobacco and Metals) are dropped from the sample.

panel A and the composite R^2s obtained with time fixed effects only is 0.017 for Apparel. This suggests that essentially all time variation in $\widehat{\eta}_{i,t}^k + \widehat{\theta}_{j,t}^k$ is absorbed by the structural gravity term.

Concern about possible large exporter and importer fixed effects on outliers (despite their moderate overall impact on the composite R^2) led us to examine the effect of successive introduction of the fixed effects from the estimation of (7) over the entire range of estimated values. For expositional simplicity, we estimate (7) on a sample pooled across all sectors and all years. Figure 2 illustrates the result of several experiments. The x-axis shows the $r_{ij,t}^k$'s and the y-axis shows the fitted residual values, the $\widehat{e}_{ij,t}^k$'s of (7). A perfect fit of the model with no measurement error in the Es and Ys would result in a fitted line that coincides with a horizontal line at zero.

We start by plotting the fitted line and the corresponding 95 percent confidence band from a specification with sector fixed effects only. The resulting fitted line, labeled "Fitted Values Sector FEs", is not horizontal at zero, but the horizontal line at zero is completely contained within the 95 percent confidence band. The data points (not shown) are densely clustered about 0 on Figure 2 (as Figure 1 shows), where the fixed effects have almost no impact, but for a small proportion of outliers the fixed effects matter significantly.

The individual roles played by the exporter, importer and time fixed effects in shifting the outliers are isolated in two other experiments displayed in Figure 2. First, we introduce the directional (exporter and importer) country fixed effects (ϕ_i and φ_j), in addition to the sector fixed effects, on suspicions that the former might control for activity elasticities not equal to one or other country-specific effects not explained by structural gravity. The resulting fitted line, labeled "Fitted Values Country FEs", is a bit closer to the horizontal line at zero, but well within the 95% confidence band of the original fitted line, obtained with sector fixed effects only. Second, we also introduce a time dimension by adding year fixed effects. The resulting fitted line "Fitted Values All FEs" almost coincides with the line based on the former specification with sector, importer and exporter fixed effects only, confirming the insignificant contribution of time. Based on these experiments, we conclude that the structural gravity term $\ln[E^k_{j,t}{P^k_{j,t}}^{\sigma_k-1}Y^k_{i,t}{\Pi^k_{i,t}}^{\sigma_k-1}]$ absorbs essentially all time and country variation in the directional country-time fixed effects.

2.2 Regression Error Test

It is natural to look for a hypothesis test of structural gravity. Unfortunately a valid test statistic requires the doubtful condition that the observations of

$$\widehat{\eta}^k_{i,t}+\widehat{\theta}^k_{j,t}-\widehat{\alpha}^k-\widehat{\psi}^k_t-\widehat{\phi}^k_i-\widehat{\varphi}^k_j-\ln[Y^k_i(\widehat{\Pi}^k_i)^{\sigma_k-1}E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}/Y^k_t],\forall i,j,k,t$$

are independent draws. Since the (unknown) measurement error process driving the Es and Ys is knitted into the calculated $\widehat{\Pi}$s and $\widehat{P}$s, there is no way to build a model of the dependence or develop a useful sufficient condition for independence.

In contrast, it is feasible to generate a conventional t-test statistic focused on regression error, taking the E,Y data as given. We construct a t-test of the null hypothesis for the residuals of (7), the raw residuals of (6) adjusted for mean measurement error. Drawing bootstrapped standard errors from the estimated data-generating process of the first stage gravity regressions, we obtain 100 bootstrapped first-stage gravity slope and fixed effect estimates. Then for each bootstrapped iteration, we construct a set of multilateral resistances, which we combine with output and expenditures data to construct 100 raw residuals $r^k_{ij,t}(s)$ for $s=1,...,100$. Next, we estimate regression (7) for each bootstrapped iteration s and calculate the estimated residuals $\widehat{e}^k_{ij,t}(s)$. For each iteration s the mean residual across partners in (6) is $\bar{r}^k_t(s)$. Finally, we use the mean of $\bar{r}^k_t(s)$ over iterations s, $\bar{r}^k_t$ and the standard error of the generated distribution of mean residuals $\bar{r}^k_t(s)$ to construct the t-statistic. P-values from one-sample t-tests for each sector in our sample are reported in Panel E of Table 1. Once again, our findings provide strong confirmation in support of the structural gravity model. With the exception of apparel, leather and printing, we cannot reject (at the 5%

level) the null hypothesis that each of the sectoral level residuals have zero mean. For each of these rejections the omission (due to lack of information) of important non-tariff barriers to trade from our gravity estimation is the source of the problem, as opposed to a failure of the structural gravity approach.

2.3 Trade Flows Fit Comparison

The relative performance of structural gravity is alternatively measured by the change in fit of the model to the trade flow data when the fixed effects are replaced by the structural gravity terms. This indirect measure addresses the potential problem that even though the fixed effect and structural measures are close, they may differ substantially in explanatory power in the trade flow model from which parameters are inferred. The comparative trade flow fit check also alleviates the suspicion that because both the fixed effects $\hat{\eta}^k_{i,t} + \hat{\theta}^k_{j,t}$ and the structural gravity terms $\ln[E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k-1}Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k-1}]$ are fitted values (or combinations of fitted values), in some hidden way they are fitted to each other.

The appropriate measure of relative performance is Aikake's (1974) Information Criterion, since we compare non-nested models estimated with PPML. The agnostic fixed effects model has the minimum AIC value in all sectors, necessarily so because its $2 \cdot 75 \cdot 4 = 600$ fixed effects are optimized to fit the more than 20,000 observations of the trade flow data. The 600 values of the structural gravity term cannot do better. The relative probability that the structural model nevertheless minimizes information loss is a measure of the closeness of fit of the structural to the agnostic model, understanding that neither model is the unknown perfect model that completely explains the data. The results are reported in Panel D of Table 1.

In the first row of the panel, we report relative probabilities when the full set of fixed effects are replaced with the structural gravity term $\ln Y^k_i E^k_j(\widehat{P}^k_j)^{\sigma_k-1}(\widehat{\Pi}^k_i)^{\sigma_k-1}$. These probabilities range from 0.871 for Food to 0.307 for Apparel, with 16 of 18 sectors yielding relative probability values greater than the critical value 0.37 (differences in AIC values less than

2) that in practice is often taken as "considerable" support for the alternative (structural) model. The goodness of fit (measured by relative probability) would rise substantially by augmenting the pure structural gravity term with the controls for mean measurement error estimated from equation (7), just as it does in moving from the simple R^2 of Panel A to the composite R^2 of Panel C in Table 1. Overall, the results confirm that the close fit of the estimated fixed effects to their structural gravity counterparts in Panel A is due to similar performance of the agnostic and structural models in explaining the trade flows from which all parameters are inferred.

To gauge the importance of the purely structural multilateral resistance terms, in the second row of panel D, we report relative probabilities obtained by replacing the full set of fixed effects only with the product of (the log of) output and expenditures $\ln Y_i^k E_j^k$, thus omitting the MR effects. Without exception across sectors, omitting the multilateral resistance terms results in worse model performance. Some sectors are more affected than others. Intuitively, the MR effects are stronger for sectors with higher transportation costs (see Wood, Furniture, Paper, Petroleum and Coal). Moreover, once we fail to account for the general equilibrium trade cost effects that are channeled via the multilateral resistances, the resulting specifications for half of the commodities in our sample fail to pass the conventional threshold of 0.37 (differences in AIC values less than 2). This suggests that econometric trade flow models that use size variables as covariates but fail to control for the multilateral resistance terms are misspecified.

3 Smell Tests

Does the gold standard claim advanced here withstand common sense smell tests? This section shows that it does with several robustness checks. First, alternative estimation with much more eclectic use of gravity components does not add any appreciable explanatory power. Second, we obtain similar results with alternative samples, different specifications, and new data sets from Canadian provincial trade in goods and services.

Despite unavoidably biased estimation, it is natural to examine how the structural gravity term and its components perform as regressors in explaining the directional country-time fixed effects $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$ when the coefficients are free to vary to improve the fit. The regression most closely linked to the constructed composite R^2 using the results of regression (7) is

$$\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t} = \gamma^k_0 + \gamma^k_1 \ln [E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k - 1} Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k - 1}] + \epsilon^k_{ij,t}, \forall i, j, \tag{8}$$

for each k and t, and overall. Here $\epsilon^k_{ij,t}$ is a random error term. Estimation is biased because $\epsilon^k_{ij,t}$ is obviously correlated with measurement error in $\ln [E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k - 1} Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k - 1}]$.[14] Hypothesis tests for such examples as estimated $\widehat{\gamma}^k_1 \neq 1$ are thus invalid, but reported anyway for interested readers.

Estimating regression (8) reveals that allowing γ^k_1 to vary from its theoretical value of 1 only very marginally raises the goodness of fit, R^2. Variants of (8) that progressively relax the coefficient restrictions on the *components* of the structural gravity term $\ln [E^k_{j,t}(\widehat{P}^k_{j,t})^{\sigma_k - 1} Y^k_{i,t}(\widehat{\Pi}^k_{i,t})^{\sigma_k - 1}]$, constrained to equal γ^k_1 in (8), further shift the coefficient estimates away from 1 while again very marginally improving goodness-of-fit. In all variants, both the size and multilateral resistance components explain important proportions of the variance of $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$. We conclude from these experiments that there is no persuasive or even modestly credible evidence against restricting the coefficient on the composite structural gravity term to its theoretically required value of 1.

Details of the results by sector are reported in Tables 2-3.[15] The first panel of the table presents the results of estimating (8) for each sector. Most importantly, the R^2s of these estimated regressions are not much larger than the corresponding simple R^2s from panel A

[14] Measurement error in the structural gravity term is due to the activity variables $\{Y^k_i, E^k_j\}$ both directly and in calculation of multilateral resistances, along with the usual estimation error from the gravity coefficients, both of which are correlated with estimation error in the $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$'s.

[15] To take advantage of the additional information contained in the standard errors of the country-specific, directional fixed effects, we estimate (8) using weighted ordinary least squares with weights equal to the inverse squared standard errors of the sum of the fixed effect estimates. The intuition is that more precise estimates should be given higher weights in the estimations.

of Table 1. The largest improvements are for the problematic sectors of Apparel and Coal and Petroleum.

The extra freedom of regression (8) to vary the slope γ_1^k does very slightly improve the fit of the regression, but at the cost of introducing biased estimation of coefficients. Without exception the estimates of the coefficient of γ_1^k are very precisely estimated and relatively close to 1.[16] The estimates of Paper and Electric Products are closest to 1 with values of 0.941 (std.err. 0.001) and 0.940 (std.err. 0.002), respectively, even though a standard F test (invalid but descriptively useful) renders them statistically different from 1. Apparel and Coal and Petroleum are the worst performing sectors, in accordance with our main findings from the previous section. Aggregate estimates of (8) from Table 4 pooled over sectors and time (in column 1) and by year pooled over sectors (in the next four columns) similarly demonstrate close conformation of the data to the restrictions of structural gravity, despite biased estimators.

The performance of separate components of structural gravity is revealed by eclectic regressions reported in the next sections and panels of Tables 2-3. In the lower section of panel A of each table, the composite term is split into its size ($\ln Y_i^k E_j^k$) and multilateral resistance ($\ln(\widehat{P}_j^k)^{\sigma_k-1}(\widehat{\Pi}_i^k)^{\sigma_k-1}$) components estimated from:

$$\widehat{\eta}_{i,t}^k + \widehat{\theta}_{j,t}^k = \tilde{\phi}_0^k + \phi_1^k \ln[(\widehat{P}_{j,t}^k)^{\sigma_k-1}(\widehat{\Pi}_{i,t}^k)^{\sigma_k-1}] + \phi_2^k \ln[E_{j,t}^k Y_{i,t}^k] + \epsilon_{ij,t}^k \qquad (9)$$

The standard F-test (invalid due to biased estimation but used as a descriptive device nonetheless) rejects the null hypotheses that ϕ_1^k and ϕ_2^k are equal to 1, but the main message from the estimation of (9) is that both size and multilateral resistance terms are highly significant economically and statistically (again, with the caveat that biased standard errors are used).

A still more eclectic approach allows for decompositions of the directional country fixed

[16]The standard errors are of course biased.

effects:

$$\hat{\eta}_{i,t}^k = \beta_0^k + +\beta_1^k \ln[(\widehat{\Pi}_{i,t}^k)^{\sigma_k - 1} Y_{i,t}^k] + u_{i,t}^k \tag{10}$$

$$\hat{\eta}_{i,t}^k = \tilde{\beta}_0^k + +\beta_2^k \ln(\widehat{\Pi}_{i,t}^k)^{\sigma_k - 1} + \beta_3^k \ln Y_{i,t}^k + \tilde{u}_{i,t}^k \tag{11}$$

for the source country, and

$$\hat{\theta}_{j,t}^k = \alpha_0^k + \alpha_1^k \ln[(\widehat{P}_{j,t}^k)^{\sigma_k - 1} E_{j,t}^k] + v_{j,t}^k \tag{12}$$

$$\hat{\theta}_{j,t}^k = \tilde{\alpha}_0^k + \alpha_2^k \ln(\widehat{P}_{j,t}^k)^{\sigma_k - 1} + \alpha_3^k \ln E_{j,t}^k + \tilde{v}_{j,t}^k. \tag{13}$$

for the destination country.

Estimates from (10)-(11) giving the success of structural components in predicting outward fixed effects $\eta_{i,t}^k$, by sector pooled across years, are reported in panel B of Tables 2-3. The outward fixed effects are explained with somewhat higher R^2 than are the sum of inward and outward effects from estimating (8) or (9). In addition, we note that overall, the β_1^k estimates from (10) are closer to 1 as well. See Furniture, for example (Table 2, panel B), with an estimate of $\hat{\beta}_1 = .99$ (std.err. 0,014). The inward fixed effects $\theta_{j,t}^k$ are explained with somewhat lower R^2 in panel C, and when structural term is split into its size and multilateral resistance components in the bottom section of panel C, the estimated effects of inward multilateral resistance are often significantly lower and relatively less precisely estimated than the coefficient estimates on expenditures. This arises because the variation in inward multilateral resistance is much less than the variation in outward multilateral resistance. The main message of these experiments is that both size and multilateral resistance terms are economically and statistically highly significant, while the extra freedom of the eclectic regressions to alter the slope coefficients from their theoretical values only very marginally improves goodness of fit.

In (8) all effects of time are attributed to the composite structural gravity term. Robustness checks show that time has essentially no additional explanatory power. Several informative experiments alter the fixed effect components of specification (8) that uses the composite structural gravity term. For brevity, we limit our experiments to the sample pooled across all sectors and years. The results are reported in the top panel of Table 5. In the first column of the table, we report the base estimates, which are obtained with sector fixed effects only. In column 'Time', we introduce year fixed effects to equation (8), in addition to the sector fixed effects. The new results are virtually identical to the ones from column (1). Similar findings are obtained with sector-year fixed effects. See column 'ProdTime' of Table 5.

In column 'Country', we introduce exporter and importer fixed effects (ϕ_i, φ_j), in addition to the sector-time interactions from column 'ProdTime'. Two findings stand out. First, the estimate of the coefficient on the composite term falls significantly. The high positive correlation between the country fixed effects and the size variables in the composite structural terms permits the former to take away some explanatory power from the structural terms. Second, combined with the evidence of insignificant time effects, the high goodness-of-fit statistic $R^2 = 0.96$ implies that there is not much room for improvement by introducing time-varying country-specific characteristics. Furthermore, the small effect of eliminating time-varying country fixed effects is reassuring because time-varying, country-specific border barriers are not modeled in the underlying gravity regression (5) due to lack of data, and might have an influence on the estimated fixed effects drawn from (5).

Next, in the bottom panel of Table 5, we experiment by altering our sample and by employing different data sets. Column 'Internal' reports estimates based on the subsample for internal trade ($i = j$), where the dependent variable is constructed as the sum $\widehat{\eta}_{i,t}^{k} + \widehat{\theta}_{i,t}^{k}$. These results are stronger as compared to the main findings from column (1). The estimate of $\hat{\gamma}_1 = 0.952$ (std.err. 0.004) is very close to 1, but still statistically different from 1, and the R^2 increases as well, suggesting that, on average, the structural gravity component explains almost 97% of the variability in the sum of the directional fixed effects for each country. Additional experiments, available by request, reveal that allowing for country-specific border effects improves the fit of this specification significantly.

Finally, in order to ensure that our results are not due to specific features of the data, we employ an alternative data set covering trade between all Canadian provinces and territories, the US and the rest of the world (ROW) for 28 sectors (19 goods and 9 services) during the period 1997-2007.[17] A notable advantage of this data set is that it enables us to test gravity for services as well as goods. The specific geography and trade of Canada require a new definition of bilateral trade costs, which we describe in Appendix C, but other than that, we employ the same econometric procedures to obtain results. The new findings validate structural gravity resoundingly.

The simple R^2 for the Canadian provincial goods and services trade fixed effects is 0.86, very close to the 0.88 for world manufacturing trade that is reported in Section 2. Split into services and goods the simple R^2s are respectively 0.95 and 0.82. The composite R^2's formed by adding the explained variance from estimating (7) are 0.95 for goods and 0.99 for services, with an overall composite R^2 equal to 0.96. We also confirm that the time fixed effects do not add much explanatory power but those effects vary across goods and services. The composite R^2's constructed with time fixed effects only for goods increases to 0.86, while the R^2 for services remains 0.95. Inspection of the gravity data and estimates by year reveals that the explanation for the effects of time effects for goods is driven by noisier data and poor first-stage gravity performance in 1997.

Compared to the 0.95 composite R^2 for global manufacturing, the application of province/country fixed effects to control for measurement error in (7) plays a larger role in improving goodness of fit for Canadian provincial goods trade. This is due to the prominence of Rest of World (ROW) trade in the gravity model estimated for Canada: specifically, the importer/product fixed effect for ROW $\varphi^k_{ROW,t}$ on the right hand side of (7) explains a sizable portion of the variation in the dependent variable of (7).

The various robustness checks of this section based on (8) and applied to Canadian

[17] These data set is constructed by Anderson et.al. (2012) in an effort to investigate the effects of exchange rates on Canadian trade. See Appendix C for further details.

trade confirm that relaxing the theoretical coefficient restrictions only marginally improves goodness of fit. See the last three columns in the bottom panel of Table 5. In column 'CAAll', which reports results across all sectors, the composite structural term explains by itself close to 95% of the variability of the sum of the gravity fixed effects, and its coefficient estimate of 1.187 (std.err. 0.005) is close to one. Decomposition into goods and services in columns 'CAGoods' and 'CAServices' respectively further confirms our findings. The goodness-of-fit for services is a remarkable .97, while the corresponding statistic for goods is .87.

4 Conclusion

This paper provides a validation of structural gravity theory based on the close fit of estimated fixed effects to their theoretical counterparts. Popper's riskiness criterion implies that the close fit is impressive, a gold standard benchmark. Various robustness checks do not shake this confidence. We conclude with drawing out the implications for future work.

If structural gravity is to be relied upon, it strengthens the credibility of the wide range of comparative static exercises that have become popular in the last decade following the examples of Eaton and Kortum (2002) and Anderson and van Wincoop (2003) with one sector versions. The reliability of disaggregated structural gravity reported on here extends the potential range of such exercises.

Reliance on structural gravity also enables powerful tools for dealing with missing or non-credible data. Empirical research on disaggregated trade (and investment and migration) flows is typically hampered by such data problems. Structural gravity and its estimated bilateral resistances and fixed effects can, with sufficiently rich but incomplete data, generate projected bilateral flows, total shipments and expenditures and inward and outward multilateral resistances and unobservable bilateral trade costs.

References

Aikake, Hirotugu 1974. "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, 19 (6), 716-23.

Anderson, James E. 2011. "The Gravity Model", *Annual Review of Economics*, 3, 133-60.

Anderson, James E. 1979. "A theoretical foundation for the gravity equation", American Economic Review 69, 106-116.

Anderson, James E. and Eric van Wincoop. 2004."Trade Costs", *Journal of Economic Literature*, 42, 691-751.

Anderson, James E. and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review*, 93, 170-92.

Anderson, James E. and Yoto V. Yotov. 2010a. "The Changing Incidence of Geography," *American Economic Review*, vol. 100(5), pages 2157-86.

Anderson, James E. and Yoto V. Yotov. 2010b. "Specialization: Pro- and Anti-Globalizing, 1990-2002," NBER Working Paper 14423.

Bergstrand, Jeffrey H., 1989, "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade," *Review of Economics and Statistics*, Vol. 71, No. 1, February 1989, 143-153.

Cheng, I.-Hui and Howard J. Wall. 2002. "Controlling for heterogeneity in gravity models of trade", Federal Reserve Bank of St. Louis Working Paper vol. 1999-010C.

Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography and Trade", *Econometrica*, 70: 1741-1779.

Feenstra, Robert. 2004. *Advanced International Trade: Theory and Evidence*, Princeton, NJ: Princeton University Press.

Head, Keith and Thierry Mayer. 2000. "Non-Europe : The Magnitude and Causes of Market Fragmentation in Europe", *Weltwirtschaftliches Archiv* 136, 285-314.

Helpman, Elhanan, Marc J. Melitz and Yona Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes", Harvard University, *Quarterly Journal of Eco-*

nomics, 123: 441-487.

Mayer, T. and R. Paillacar and S. Zignago. 2008. "TradeProd. The CEPII Trade, Production and Bilateral Protection Database: Explanatory Notes", CEPII.

Mayer, T. and S. Zignago. 2006. "Notes on CEPIIs distances measures", CEPII.

McCallum, John (1995) "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 1995, 85(3), pp. 615-623.

Olivero, Maria and Yoto V. Yotov. Forthcoming. "Dynamic Gravity: Theory and Empirical Implications," *Canadian Journal of Economics*.

Popper, Karl (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge, London.

Redding, Stephen and Anthony J. Venables (2004), "Economic Geography and International Inequality", *Journal of International Economics* 62, 53-82.

Rose, Andrew K. (2004), "Do WTO members have more liberal trade policy? ", *Journal of International Economics*, 63, 209-235.

Santos Silva, Jorge and Sylvana Tenreyro. 2006. "The Log of Gravity", *Review of Economics and Statistics*, Vol. 88, No. 4: 641-658.

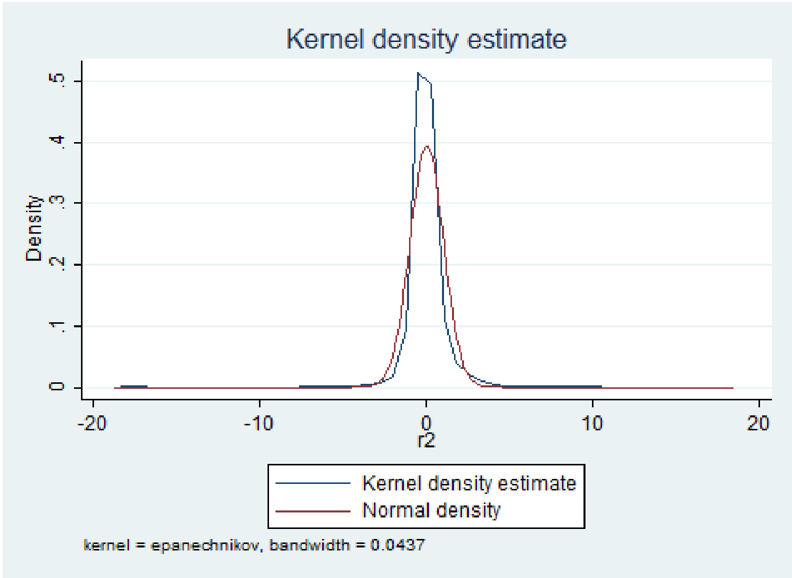

Figure 1: Estimated Distribution of r

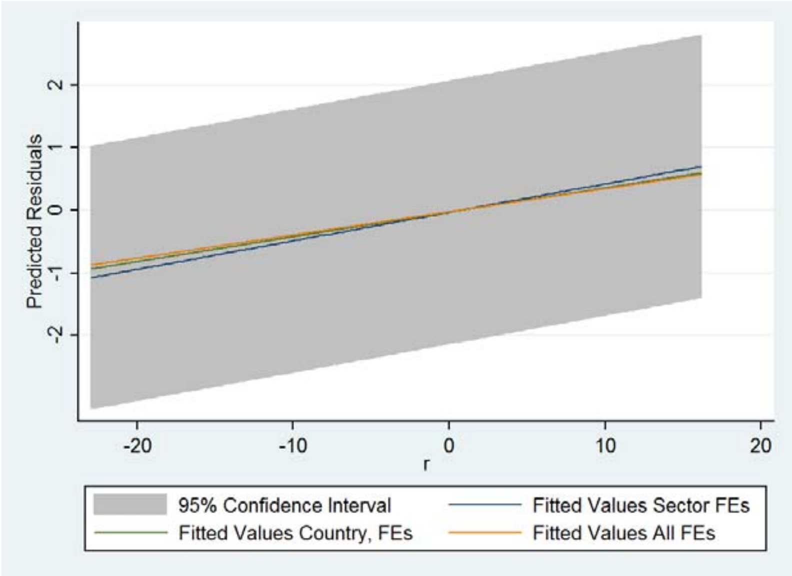

Figure 2: Residual Fixed Effects

Table 1: Main Tests of Structural Gravity, by Product

Dep. Var. $r^k_{ij,t}$	Food	BevTob	Textiles	Apparel	Leather	Wood	Furniture	Paper	Priniting
A. Simple R^2	.955	.841	.901	.708	.809	.884	.896	.958	.951
B. Variance Decomposition									
$\ln[(\widehat{P}^k_{j,t})^{\sigma_k-1}(\widehat{\Pi}^k_{i,t})^{\sigma_k-1}]$	.377	.385	.261	.229	.207	.346	.314	.282	.371
$\ln[E^k_{j,t}Y^k_{i,t}]$	.578	.456	.64	.478	.602	.538	.582	.676	.58
C. Composite R^2									
$\psi_t, \phi_i, \varphi_j$	.985	.904	.964	.909	.918	.956	.963	.982	.982
ψ_t	.955	.843	.904	.725	.811	.886	.899	.958	.952
D. Goodness-of-Fit (AIC) $exp[(AIC_{FE} - AIC_{STR})/2]$									
$\ln Y^k_i E^k_j (\widehat{P}^k_j)^{\sigma_k-1}(\widehat{\Pi}^k_i)^{\sigma_k-1}$	.87	.745	.691	.307	.67	.514	.51	.636	.71
$\ln[E^k_{j,t}Y^k_{i,t}]$	.424	.365	.536	.214	.564	.215	.187	.304	.202
E. T-tests									
$(\lvert T\rvert > \lvert t\rvert)$	.672	.506	.062	.04	.028	.971	.547	.63	.002

Dep. Var. $r^k_{ij,t}$	Chemicals	PetrCoal	RbbPlst	Minerals	Metals	Machinery	Electric	Transport	Other
A. Simple R^2	.957	.623	.943	.958	.84	.908	.928	.903	.89
B. Variance Decomposition									
$\ln[(\widehat{P}^k_{j,t})^{\sigma_k-1}(\widehat{\Pi}^k_{i,t})^{\sigma_k-1}]$	.233	.293	.289	.388	.239	.169	.212	.205	.176
$\ln[E^k_{j,t}Y^k_{i,t}]$	.724	.331	.654	.57	.601	.739	.717	.698	.714
C. Composite R^2									
$\psi_t, \phi_i, \varphi_j$	.987	.802	.984	.99	.924	.977	.98	.973	.962
ψ_t	.957	.628	.943	.96	.845	.91	.931	.905	.893
D. Goodness-of-Fit (AIC) $exp[(AIC_{FE} - AIC_{STR})/2]$									
$\ln Y^k_i E^k_j (\widehat{P}^k_j)^{\sigma_k-1}(\widehat{\Pi}^k_i)^{\sigma_k-1}$	.676	.565	.712	.696	.777	.559	.354	.717	.487
$\ln[E^k_{j,t}Y^k_{i,t}]$	.452	.216	.37	.427	.471	.469	.236	.625	.436
E. T-tests									
$(\lvert T\rvert > \lvert t\rvert)$	.237	.966	.636	.409	.998	.444	.108	.65	.56

This table reports the main tests of structural gravity by sector. In panel A, we report simple goodness-of-fit measures based on equation (6). In panel B, we decompose the variance into its multilateral resistance and size components. In panel C, we construct composite goodness-of-fit statistics based on specification (7). In panel D, we report relative AIC probabilities from alternative PPML specifications. Finally, in panel E, we report p-values from one-sample t-tests by sector. See text for details.

Table 2: Smell Tests of Structural Gravity, By Product

A. Dep. Var. $\widehat{\eta}_{i,t}^k + \widehat{\theta}_{j,t}^k$	Food	BevTob	Textiles	Apparel	Leather	Wood	Furniture	Paper	Priniting
$\ln Y_i^k E_j^k (\widehat{P}_j^k)^{\sigma_k - 1} (\widehat{\Pi}_i^k)^{\sigma_k - 1}$	0.926	0.915	0.893	0.788	0.824	0.886	0.923	0.941	0.930
	(0.001)**	(0.005)**	(0.002)**	(0.004)**	(0.004)**	(0.002)**	(0.002)**	(0.001)**	(0.001)**
N	20882	20746	20882	20882	20807	20882	20882	20882	20882
r2	0.962	0.815	0.917	0.738	0.852	0.904	0.900	0.961	0.956
Size vs. MRs									
$\ln(\widehat{P}_j^k)^{\sigma_k - 1} (\widehat{\Pi}_i^k)^{\sigma_k - 1}$	0.811	0.871	0.733	0.689	0.817	0.774	0.823	0.838	0.764
	(0.002)**	(0.007)**	(0.005)**	(0.009)**	(0.011)**	(0.004)**	(0.005)**	(0.003)**	(0.002)**
$\ln Y_i^k E_j^k$	0.943	0.939	0.893	0.804	0.826	0.906	0.930	0.946	0.942
	(0.001)**	(0.005)**	(0.002)**	(0.004)**	(0.004)**	(0.002)**	(0.002)**	(0.001)**	(0.001)**
N	20882	20746	20882	20882	20807	20882	20882	20882	20882
r2	0.967	0.817	0.921	0.751	0.852	0.910	0.901	0.962	0.965
B. Dep. Var. $\widehat{\eta}_{i,t}^k$									
$\ln Y_i^k (\widehat{\Pi}_i^k)^{\sigma_k - 1}$	0.945	0.975	0.935	0.871	0.869	0.913	0.994	0.947	0.934
	(0.008)**	(0.051)**	(0.015)**	(0.032)**	(0.044)**	(0.016)**	(0.014)**	(0.009)**	(0.009)**
N	290	288	290	290	289	290	290	290	290
r2	0.975	0.820	0.938	0.779	0.842	0.944	0.932	0.975	0.973
Size vs. MRs									
$\ln(\widehat{\Pi}_i^k)^{\sigma_k - 1}$	0.834	0.920	0.863	1.008	1.221	0.886	0.991	0.899	0.803
	(0.013)**	(0.068)**	(0.040)**	(0.074)**	(0.089)**	(0.024)**	(0.040)**	(0.024)**	(0.018)**
$\ln Y_i^k$	0.978	1.022	0.936	0.872	0.839	0.922	0.995	0.954	0.967
	(0.008)**	(0.042)**	(0.016)**	(0.032)**	(0.049)**	(0.015)**	(0.017)**	(0.011)**	(0.010)**
N	290	288	290	290	289	290	290	290	290
r2	0.981	0.822	0.939	0.781	0.853	0.945	0.932	0.975	0.979
C. Dep. Var. $\widehat{\theta}_{j,t}^k$									
$\ln E_j^k (\widehat{P}_j^k)^{\sigma_k - 1}$	0.889	0.839	0.827	0.760	0.785	0.842	0.808	0.917	0.932
	(0.016)**	(0.037)**	(0.040)**	(0.034)**	(0.041)**	(0.024)**	(0.034)**	(0.021)**	(0.023)**
N	286	286	286	286	286	286	286	286	286
r2	0.937	0.899	0.847	0.779	0.865	0.869	0.819	0.903	0.899
Size vs. MRs									
$\ln(\widehat{P}_j^k)^{\sigma_k - 1}$	0.690	0.769	0.463	0.525	0.519	0.541	0.437	0.607	0.612
	(0.019)**	(0.048)**	(0.071)**	(0.070)**	(0.063)**	(0.045)**	(0.080)**	(0.074)**	(0.037)**
$\ln E_j^k$	0.887	0.865	0.802	0.785	0.814	0.862	0.795	0.888	0.876
	(0.014)**	(0.036)**	(0.039)**	(0.031)**	(0.042)**	(0.019)**	(0.033)**	(0.023)**	(0.021)**
N	286	286	286	286	286	286	286	286	286
r2	0.953	0.904	0.880	0.800	0.878	0.905	0.859	0.922	0.943

Standard errors in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. This table reports tests of structural gravity by sector. The dependent variables (see panel labels A., B., and C.) are functions of the directional, country-specific fixed effects estimates, obtained from panel PPML gravity estimations. All estimates are obtained with year fixed effects, which are omitted for brevity. The estimator is weighted least squares.

Table 3: Smell Tests of Structural Gravity, By Product

A. Dep. Var. $\widehat{\eta}_{i,t}^k + \widehat{\theta}_{j,t}^k$	Chemicals	PetrCoal	RbbPlst	Minerals	Metals	Machinery	Electric	Transport	Other
$\ln Y_i^k E_j^k (\widehat{P}_j^k)^{\sigma_k-1}(\widehat{\Pi}_i^k)^{\sigma_k-1}$	0.937	0.725	0.927	0.933	0.827	0.883	0.940	0.875	0.873
	(0.001)**	(0.004)**	(0.002)**	(0.001)**	(0.003)**	(0.002)**	(0.002)**	(0.002)**	(0.002)**
N	20882	20657	20882	20882	20882	20882	20882	20882	20882
r2	0.961	0.716	0.955	0.958	0.897	0.926	0.930	0.928	0.910
Size vs. MRs									
$\ln(\widehat{P}_j^k)^{\sigma_k-1}(\widehat{\Pi}_i^k)^{\sigma_k-1}$	0.862	0.587	0.803	0.759	0.617	0.937	0.924	0.891	0.912
	(0.004)**	(0.005)**	(0.003)**	(0.003)**	(0.008)**	(0.006)**	(0.006)**	(0.005)**	(0.009)**
$\ln Y_i^k E_j^k$	0.935	0.775	0.924	0.968	0.817	0.883	0.942	0.875	0.872
	(0.001)**	(0.004)**	(0.002)**	(0.001)**	(0.003)**	(0.002)**	(0.002)**	(0.002)**	(0.002)**
N	20882	20657	20882	20882	20882	20882	20882	20882	20882
r2	0.959	0.729	0.956	0.969	0.906	0.927	0.932	0.929	0.911
B. Dep. Var. $\widehat{\eta}_{i,t}^k$									
$\ln Y_i^k (\widehat{\Pi}_i^k)^{\sigma_k-1}$	0.939	0.714	0.930	0.945	0.836	0.859	0.931	0.857	0.873
	(0.010)**	(0.027)**	(0.018)**	(0.009)**	(0.033)**	(0.018)**	(0.014)**	(0.015)**	(0.020)**
N	290	287	290	290	290	290	290	290	290
r2	0.974	0.772	0.964	0.965	0.911	0.932	0.952	0.935	0.928
Size vs. MRs									
$\ln(\widehat{\Pi}_i^k)^{\sigma_k-1}$	0.964	0.699	0.943	0.780	0.824	1.129	1.026	1.025	1.173
	(0.032)**	(0.041)**	(0.035)**	(0.026)**	(0.086)**	(0.052)**	(0.038)**	(0.041)**	(0.064)**
$\ln Y_i^k$	0.937	0.721	0.929	0.999	0.836	0.854	0.919	0.853	0.856
	(0.011)**	(0.038)**	(0.020)**	(0.012)**	(0.034)**	(0.017)**	(0.017)**	(0.015)**	(0.018)**
N	290	287	290	290	290	290	290	290	290
r2	0.974	0.772	0.964	0.972	0.911	0.938	0.953	0.939	0.935
C. Dep. Var. $\widehat{\theta}_{j,t}^k$									
$\ln E_j^k (\widehat{P}_j^k)^{\sigma_k-1}$	0.923	0.696	0.927	0.910	0.868	0.992	0.917	0.958	0.904
	(0.018)**	(0.058)**	(0.024)**	(0.023)**	(0.041)**	(0.021)**	(0.040)**	(0.022)**	(0.024)**
N	286	286	286	286	286	286	286	286	286
r2	0.921	0.555	0.895	0.918	0.896	0.905	0.802	0.915	0.879
Size vs. MRs									
$\ln(\widehat{P}_j^k)^{\sigma_k-1}$	0.670	0.394	0.513	0.635	0.458	0.432	0.372	0.525	0.420
	(0.053)**	(0.061)**	(0.036)**	(0.027)**	(0.072)**	(0.071)**	(0.112)**	(0.057)**	(0.070)**
$\ln E_j^k$	0.902	0.830	0.856	0.905	0.809	0.914	0.860	0.906	0.883
	(0.018)**	(0.057)**	(0.017)**	(0.016)**	(0.040)**	(0.019)**	(0.044)**	(0.022)**	(0.021)**
N	286	286	286	286	286	286	286	286	286
r2	0.929	0.703	0.943	0.956	0.925	0.927	0.820	0.934	0.895

Standard errors in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. This table reports tests of structural gravity by sector. The dependent variables (see panel labels A., B., and C.) are functions of the directional, country-specific fixed effects estimates, obtained from panel PPML gravity estimations. All estimates are obtained with year fixed effects, which are omitted for brevity. The estimator is weighted least squares.

Table 4: Smell Tests of Structural Gravity, by Year

Dep. Var. $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$	ALL	1990	1994	1998	2002
$\ln Y^k_i E^k_j (\widehat{P}^k_j)^{\sigma_k - 1} (\widehat{\Pi}^k_i)^{\sigma_k - 1}$	0.891	0.907	0.905	0.884	0.877
	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**
cons	7.801	7.740	7.606	7.870	7.934
	(0.006)**	(0.014)**	(0.010)**	(0.013)**	(0.011)**
N	375440	68015	102225	102600	102600
r2	0.938	0.953	0.945	0.936	0.928
Size vs. MRs					
$\ln(\widehat{P}^k_j)^{\sigma_k - 1} (\widehat{\Pi}^k_i)^{\sigma_k - 1}$	0.775	0.815	0.788	0.768	0.749
	(0.001)**	(0.002)**	(0.003)**	(0.002)**	(0.002)**
$\ln Y^k_i E^k_j$	0.898	0.911	0.908	0.893	0.890
	(0.001)**	(0.001)**	(0.001)**	(0.001)**	(0.001)**
_cons	-31.758	-32.504	-32.311	-31.544	-31.271
	(0.033)**	(0.066)**	(0.057)**	(0.071)**	(0.062)**
N	375440	68015	102225	102600	102600
r2	0.940	0.954	0.947	0.938	0.931

Standard errors in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. This table reports tests of structural gravity pooled across all sectors. The first column reports estimates across all years. The next four columns report estimates by year. The dependent variable is always $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_{j,t}$. All estimates are obtained with sector fixed effects, which are omitted for brevity. The estimator is weighted least squares. See text for further details.

Table 5: Testing Structural Gravity, Robustness Analysis

	Main	Time	ProdTime	Country
$\ln Y^k_i E^k_j (\widehat{P}^k_j)^{\sigma_k - 1} (\widehat{\Pi}^k_i)^{\sigma_k - 1}$	0.891	0.891	0.892	0.690
	(0.001)**	(0.001)**	(0.001)**	(0.001)**
cons	7.801	7.700	7.874	9.919
	(0.006)**	(0.007)**	(0.010)**	(0.022)**
Fixed Effects	ψ^k	ψ^k, δ_t	ψ^k_t	$\psi^k_t, \phi_i, \varphi_j$
N	375440	375440	375440	375440
r2	0.938	0.939	0.940	0.958

	Internal	CAAll	CAGoods	CAServices
$\ln Y^k_i E^k_j (\widehat{P}^k_j)^{\sigma_k - 1} (\widehat{\Pi}^k_i)^{\sigma_k - 1}$	0.952	1.187	1.219	1.124
	(0.004)**	(0.005)**	(0.007)**	(0.007)**
cons	7.287	3.067	2.868	-1.940
	(0.037)**	(0.102)**	(0.109)**	(0.024)**
Fixed Effects	ψ^k	ψ^k	ψ^k	ψ^k
N	5142	18746	13832	4914
r2	0.968	0.946	0.869	0.966

Standard errors in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. This table reports robustness tests of structural gravity. The dependent variable is always $\widehat{\eta}^k_{i,t} + \widehat{\theta}^k_j$. The estimator is weighted least squares. See text for further details.

Appendix A: Main Data

This study covers 76 trading partners[18] and 18 commodities aggregated on the basis of the United Nations' 3-digit International Standard Industrial Classification (ISIC) Revision 2.[19]

Bilateral trade flows, measured in thousands of current US dollars, are from CEPII's *Trade, Production and Bilateral Protection Database*[20] (TradeProd) and the United Nation Statistical Division's COMTRADE Database.[21] TradeProd is the primary source. The reason is that TradeProd is based on CEPII's *Base pour l'Analyse du Commerce International* (BACI), which implements a consistent procedure for mapping the CIF (cost, insurance and freight) values reported by the importers in COMTRADE to the FOB (free on board) values reported by the exporters in COMTRADE.[22] To further increase the number of non-missing bilateral trade values, we add the mean of the bilateral trade flows from COMTRADE.[23]

Industrial production data comes from two sources. The primary source is the United Nations' UNIDO Industrial Statistics database, which reports industry level output data at the 3 and 4-digit level of ISIC Code (Revisions 2 and 3). In addition to UNIDO, we

[18] Argentina, Armenia, Australia, Austria, Azerbaijan, Bulgaria, Belgium-Luxembourg, Bolivia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Cyprus, Czech Republic, Germany, Denmark, Ecuador, Egypt, Spain, Estonia, Finland, France, United Kingdom, Greece, Guatemala, Hong Kong, China, Hungary, Indonesia, India, Ireland, Iran, Italy, Jordan, Japan, Kazakhstan, Kenya, Kyrgyz Republic, Korea, Kuwait, Sri Lanka , Lithuania , Latvia, Morocco, Moldova, Mexico, Macedonia, Malta , Mongolia, Mozambique, Mauritius , Malaysia , Netherlands, Norway, Oman, Panama, Philippines, Poland, Portugal, Romania, Russian Federation, Senegal, Singapore , El Salvador, Slovak Republic, Slovenia, Sweden, Trinidad and Tobago, Turkey, Tanzania, Ukraine, Uruguay, United States, Venezuela, South Africa.

[19] The complete United Nations' 3-digit International Standard Industrial Classification consists of 28 sectors. We combine some commodity categories when it is obvious from the data that countries report sectoral output levels in either one disaggregated category or the other. Our commodity categories are: 1 Food; 2 Beverage and Tobacco; 3 Textiles; 4 Apparel; 5 Leather; 6 Wood; 7 Furniture; 8 Paper; 9 Printing; 10 Chemicals; 11 Petroleum and Coal; 12 Rubber and Plastic; 13 Minerals; 14 Metals; 15 Machinery; 16 Electric; 17 Transportation; and, 18. Other. A detailed conversion table between ours and the UN 3-digit ISIC classification is available upon request.

[20] For details regarding this database see Mayer, Paillacar and Zignago (2008).

[21] We access COMTRADE through the World Integrated Trade Solution (WITS) software, http://wits.worldbank.org/witsweb/.

[22] As noted in Anderson and Yotov (2010), in principle, gravity theory calls for valuation of exports at delivered prices. In practice, valuation of exports FOB avoids measurement error arising from poor quality transport cost data. For details regarding BACI see Gaulier and Zignago (2008).

[23] We also experiment by just using the export data from COMTRADE and then assigning missing trade values to the observations when only data on imports are available. Estimation results are very similar.

use CEPII's TradeProd database,[24] as a secondary source.[25] 10.8 percent of the original data were missing after combining the two data sets. As output data are crucial for the calculation of the multilateral resistance indexes, we construct the missing values. First, we interpolate the data to decrease the missing values to 8.6 percent.[26] Then, we extrapolate the rest of the missing values using GDP deflator data, which comes from the World Bank's World Development Indicators (WDI) Database.[27]

We generate internal trade and also expenditure data by combining total shipments data and export data. Internal trade volumes are calculated as

$$X_{ii}^k = Y_i^k - \sum_{j \neq i} X_{ij}^k. \tag{14}$$

For expenditures, we use

$$E_j^k = \sum_{i \neq j} (X_{ij}^k + X_{jj}^k). \tag{15}$$

This procedures may result in negative expenditure and internal trade values, and does so for 1.7% of the internal trade observations and for 0.29 percent of the expenditures. In addition, 4.6 percent of the expenditures were missing.[28] To construct the missing expenditure values, first, we interpolate the data, then, we extrapolate the rest using CPI data from the WDI Database.[29]

To substitute for the negative internal trade and expenditure values, we use the average internal trade to expenditure ratio for each country across all products. This has to be done so that the expenditure shares and shipment shares remain consistent by modifying their values in turn. Specifically, let $K(i)$ denote the set of goods for which, for any country i,

[24]TradeProd uses the OECD STAN Industrial Database as well as UNIDO's IndStat Database.

[25]We experiment with two output variables, based on the main data source, to obtain identical results.

[26]Most of the missing observations are for the early years in the sample (1990-1993) and for the former Soviet republics (e.g. Armenia, Estonia, Lithuania, etc.), which declared independence during the early 90s.

[27]GDP deflator data were not available for Belgium-Luxembourg (BLX). We use Belgium's GDP deflator data to proxy for BLX.

[28]Once again, most of the missing observations are for the early years in the sample (1990-1993) and for the former Soviet republics, which declared independence during the early 90s.

[29]CPI data were not available for Belgium-Luxembourg (BLX). We used Belgium's CPI to proxy for BLX.

$X_{ii}^k > 0$. Aggregate across $k \in K(i)$ to form the aggregate version of (14):

$$Y_i^{'} - \sum_{k \in K(i), j \neq i} X_{ij}^k = X_{ii}^{'}. \tag{16}$$

Similarly form 'aggregate' expenditure

$$E_i^{'} = \sum_{k \in K(i), j \neq i} (X_{ji}^k + X_{ii}^k). \tag{17}$$

From these restricted aggregates, form the average ratio of internal trade to expenditure:

$$s_{ii} = X_{ii}^{'}/E_i^{'}, \forall i. \tag{18}$$

Finally, generate the value of inferred internal trade as

$$X_{ii}^k = s_{ii} E_i^{'}, \forall k \notin K(i). \tag{19}$$

Using the generated values from (19), replace the values of internal trade where (14) gives a non-positive value. Then use (14) again with the new data. For consistency of the data, this means that the original data on Y_i^k must be increased by the inferred value of internal trade from (19).

To construct the distance variable, we employ the methods from Mayer and Zignago (2006).[30] Their approach is appealing because it can be used to calculate consistently both internal distances and bilateral distances.[31] In addition, we follow Eaton and Kortum (2002) to decompose distance effects into four intervals. The distance intervals, in kilometers, are:

[30] Their procedure is based on Head and Mayer (2000), using the following formula to generate weighted distances: $d_{ij} = \sum_{k \in i} \frac{pop_k}{pop_i} \sum_{l \in j} \frac{pop_l}{pop_j} d_{kl}$, where pop_k is the population of agglomeration k in trading partner i, and pop_l is the population of agglomeration l in trading partner j, and d_{kl} is the distance between agglomeration k and agglomeration l, measured in kilometers, and calculated by the Great Circle Distance Formula. All data on latitude, longitude, and population are from the World Gazetteer web site.

[31] In the few instances where we were not able to implement Mayer and Zignago's procedure, we just took the distance between the main cities from the two trading partners.

[0, 3000); [3000, 7000); [7000, 10000); [10000, maximum]. Data on other standard gravity variables such as common language, common border, and colonial ties are from CEPII's *Distances* Database and from Rose (2004).[32] We also generate a set of border dummy variables, which take a value of one for internal trade.

Appendix B: Canadian Data

In order to apply our tests to Canadian trade, we employ a different definition of bilateral trade costs (due to the specifics of Canada's geography and trade), and we use different data. Other than that, the econometric treatment and all other procedures are the same as described in the main text.

Canadian Trade Costs. Following Anderson et.al. (2012), we define bilateral trade costs for a generic sector as:

$$\begin{aligned} t_{ij}^{1-\sigma} &= e^{\gamma_1 DISTANCE_{ij} + \gamma_2 CONTIG\_PR\_PR_{ij} + \gamma_3 CONTIG\_PR\_ST_{ij} +} \times \\ & \quad e^{\gamma_4 SAME\_PROV_{ij} + \gamma_5 CAN + \gamma_6 USA + \gamma_7 ROW}. \end{aligned} \tag{20}$$

Here, $DISTANCE_{ij}$ is the logarithm of bilateral distance between trading partners i and j. $CONTIG\_PR\_PR_{ij}$ takes a value of one when two provinces share a common border and is set to zero otherwise. $CONTIG\_PR\_ST_{ij}$ is equal to one when a Canadian province neighbors a US state. The next four variables are designed to capture regional borders. $SAME\_PROV_{ij}$ takes a value of one for intra-provincial trade, i.e. when $i = j$ and $i \in CA$, and it is equal to zero otherwise. Finally, CAN, USA and ROW are indicator variables designed to capture internal trade(international borders) for each of the three large regions in our sample. For example, USA is equal to one for trade within US, and it is equal to zero otherwise. ROW and CAN are defined similarly. It should be noted that, given the composition of our sample, CAN takes a value of one for inter-provincial trade as well as for intra-

[32] Rose's (2004) original data covers the period up to the year of 2000, so we update some of the variables in order to match the time span investigated in this study.

provincial trade, and it is equal to zero elsewhere. Thus, the estimate on $SAME\_PROV_{ij}$ would capture any deviation of intra-provincial trade above or below inter-provincial trade and should be interpreted accordingly.[33]

Canadian Data. Our data set extends the Canadian goods data from Anderson and Yotov (2010a) and combines it with the Canadian services data from Anderson et.al. (2011). It covers the period 1997-2007[34] for a total of 28 sectors (19 goods and 9 services). The trading partners include 12 provinces and territories, USA and a rest of the world (ROW) aggregate.[35] Without going into details,[36] we note that our sources for trade flows data are Statistics Canada, the United Nation Statistical Division (UNSD) Commodity Trade (COMTRADE) DataBase and the US Bureau of Economic Analysis (BEA). Statistics Canada offers data on provincial output for both goods and services. Manufacturing output data for US and for the rest of the world come from the United Nations' UNIDO Industrial Statistics database and from the World Database of International Trade (BACI), constructed by CEPII. Agricultural and mining output data for US and ROW are from the United Nations Food and Agriculture Organization (FAOSTAT) and from the Energy Information Administration, respectively. Finally, production data for US services come from the BEA and output for the rest of the world is from the GTAP database. Production data limitations allowed us to construct goods multilateral resistances for the period 1997-2003 and services multilateral resistances for the period 2003-2007. This determined the samples for our gravity tests in Table 5.

[33] Such a distinction cannot be made for the aggregate regions (US and ROW).

[34] The first-stage gravity estimates employed in our tests are based on 2-year lags.

[35] The Northwest Territories and Nunavut are combined, even though they are separate since April, 1999.

[36] See the Data Appendix from Anderson and Yotov (2010a) for details on the goods data and Anderson et.al. (2011) for details on the services data.