ONLINE APPENDIX

(NOT TO BE PUBLISHED)

WHEN ARE INSTRUMENTS GENERATED FROM GEOGRAPHIC CHARACTERISTICS IN BILATERAL RELATIONSHIPS INVALID? [1]

Sabine Deij[♦], Jakob B. Madsen[*], and Laura Puzzello[†]

[♦]*Melbourne School of Population and Global Health, University of Melbourne*
[*]*Department of Economics, University of Western Australia*
[†]*Department of Economics and SoDa Laboratories, Monash Business School, Monash University*

Table of Contents

[1] We are grateful for the comments and suggestions of Reshad Ahsan, Davin Chor, Ken Clements, Pascalis Raimondos, Mark Razhev, Kevin Staub, Chris Taber, Vladimir Tyazhelnikov, conference participants at the 2017 Melbourne Trade Workshop, 30th PhD Conference in Economics and Business, 13th Australasian Trade Workshop, 2018 Monash Macroeconomics Workshop, and seminar participants at the Graduate Institute Geneva, Tuborg Research Centre, University of Copenhagen and Oz Virtual Econ Research Seminar for helpful comments and suggestions. We are especially grateful to Andy Bernard, Kirill Borusyak, Julia Cajal Grossi, and Andrei Levchenko, four referees and the Editor (Marco Del Negro) for insightful comments and suggestions on earlier versions of the paper. Jakob Madsen acknowledges financial support from the Australian Research Council, grants DP150100061 and DP170100339.

Online Appendix A. Aggregation and Violations of the Exclusion Restriction in other Applications.

We found three examples where the aggregation of predicted values from auxiliary regressions leads to violations of the exclusion restriction. The first example is found in Rajan and Subramanian (2008), which proposes a two-step procedure in the spirit of Frankel and Romer (1999) to generate an instrument for foreign aid and identify the causal effect of aid on income using a cross-section of countries. In the first step they estimate a bilateral aid equation where aid between any two countries depends only on measures of the historic relationship between them. In the second step they sum predicted bilateral aid shares from this equation over *observed* donors only to generate an instrument for aid received by each country. The aggregation choice in this second step implies that the generated instrument captures cross-country variation in both the number of donors and the share of observed aid that is zero. Both these components are likely to be endogenous to recipient countries' income per capita because the net benefit of aid tends to be higher in destinations with better institutions and infrastructure. Thus, the instrument is invalid.

The second example emerges from a strand of the migration literature that generates instruments for immigration or diversity indexes applying the Frankel-Romer approach to bilateral flows of individuals (Felbermayr et al. 2010; Ortega and Peri, 2014; Alesina et al. 2016; and Bove and Elia, 2017). As in the trade literature, we find that there is confusion regarding which predictions from the bilateral migration equation should be included in the relevant instrument. Backed by our results, we advocate that predictions corresponding to all bilateral flows be included, especially when the outcome of interest in the structural equation is income per capita or any index of economic prosperity.

Finally, the last example belongs to the China shock literature. In 2017, one of us was assigned to discuss a draft version of the paper by Amiti et al. (2020) that shows that China's accession to the WTO has been good for Americans after all, due to a drop in product prices. In an exercise instrumental to the analysis, Amiti et al. (2020) use panel data to estimate how expanded imports of inputs due to the WTO accession affected Chinese firms' TFP. Specifically, they first estimate an import demand equation at the firm-input level that includes Chinese tariffs and their interaction with the type of firm as covariates, firm and time fixed effects. From the regression above, after controlling for selection, they predict imported inputs only for the inputs actually purchased by each Chinese firm, which they aggregate to derive a time-varying instrument for imports of inputs at the firm level. This variable is then included among the explanatory variables of Chinese firms' TFP. We found this aggregation choice problematic since the change in the number of a firm's inputs is likely to be positively correlated with its TFP, as predicted in standard product-variety endogenous growth models. Summing predicted firm

level purchases over all possible imported inputs yields an instrument free from this endogeneity concern.

References

Alesina, A., Harnoss, J., & Rapoport, H. (2016). Birthplace Diversity and Economic Prosperity. *Journal of Economic Growth*, *21*(2), 101–138.

Amiti, M., Dai, M., Feenstra, R. C., & Romalis, J. (2020). How Did China's WTO Entry Affect U.S. Prices? *Journal of International Economics*. 126.

Bove, V., & Elia, L. (2017). Migration, Diversity, and Economic Growth. *World Development*, *89*, 227–239.

Felbermayr, G. J., Hiller, S., & Sala, D. (2010). Does Immigration Boost Per Capita Income? *Economics Letters*, *107*(2), 177–179.

Frankel, J. A., & Romer, D. (1999). Does Trade Cause Growth? *American Economic Review*, *89*(3), 379–399.

Ortega, F., & Peri, G. (2014). Openness and Income: The Roles of Trade and Migration. *Journal of International Economics*, *92*(2), 231–251.

Rajan, R. G., & Subramanian, A. (2008). Aid and Growth: What Does the Cross-Country Evidence Really Show? *Review of Economics & Statistics*, 90(4), 643–665.

Online Appendix B. Additional estimates using actual data and the "OLS instrument"

Table OB.1 reports estimates for ten additional specifications of the income equation, each estimated by 2SLS and IV using, respectively, $\hat{T}_i^{Pos}$ and $\hat{T}_i^{All}$ as instruments for trade. These specifications are the same as in Noguer and Siscart (2005).

Data on the percentage of land or population in the tropics, and continents is from the Centre for International Development (Gallup et al. 1999). Legal origin is from La Porta et al. (2008) and, when missing, from the CIA World Factbook. The index of ethno-linguistic fractionalization is from Easterly and Levine (1997). Data on constraint on executive is from the Polity IV Project (2014). Finally, data on corruption and the quality of governance come from the International Country Risk Guide (ICRG) provided by the Political Risk Services Group based on work by Knack and Keefer (1995).

Standard errors for the IV regressions are adjusted following the approach proposed by Frankel and Romer (1999), to account for the fact that instruments depend on the parameters of the bilateral trade equation. Table OB.1 further reports selected results from the first-stage regressions.

Table OB.1 Estimates of the Income Equation using Actual Data and the "OLS instrument": Additional Controls Included

	PANEL A									
	Model (5)		**Model (6)**		**Model (7)**		**Model (8)**		**Model (9)**	
	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$
Income regressions:										
Trade share$_i$	0.693*	1.023***	0.613	1.078***	0.451	0.710**	0.524	0.909***	0.841*	1.110***
	(0.386)	(0.353)	(0.440)	(0.407)	(0.329)	(0.288)	(0.379)	(0.330)	(0.490)	(0.413)
Ln population$_i$	-0.030	0.001	0.067	0.108	0.006	0.028	-0.026	0.010	-0.024	0.017
	(0.077)	(0.081)	(0.085)	(0.088)	(0.066)	(0.066)	(0.073)	(0.076)	(0.097)	(0.090)
Ln area$_i$	0.146**	0.176**	-0.081	-0.033	0.026	0.053	0.060	0.097	0.124*	0.140**
	(0.070)	(0.071)	(0.083)	(0.082)	(0.063)	(0.063)	(0.069)	(0.069)	(0.069)	(0.070)
Latitude$_i$	0.609**	0.568**	0.058	0.010			0.212	0.169	0.296	0.280
	(0.273)	(0.287)	(0.328)	(0.339)			(0.260)	(0.271)	(0.379)	(0.396)
% Population in tropics$_i$	-2.012***	-1.979***			-1.304***	-1.290***	-1.480***	-1.447***	-1.293***	-1.276***
	(0.203)	(0.208)			(0.219)	(0.222)	(0.225)	(0.232)	(0.285)	(0.284)
Distance to equator$_i$					2.483***	2.431***				
					(0.368)	(0.370)				
% Land in tropics$_i$			-1.565***	-1.533***			-0.719***	-0.712***		
			(0.186)	(0.185)			(0.206)	(0.207)		
Sub-Saharan Africa$_i$									-0.865***	-0.839**
									(0.329)	(0.333)
East Asia$_i$									-0.413	-0.529
									(0.394)	(0.374)
Latin America$_i$									-0.123	-0.054
									(0.312)	(0.303)
Observations	98	98	98	98	98	98	98	98	98	98
First-stage regressions:										
$\hat{T}_i^{*}$	8.655***	9.554***	8.853***	9.659***	7.602***	8.487***	8.825***	9.649***	7.364***	8.597***
	(2.216)	(2.302)	(2.211)	(2.294)	(1.935)	(2.091)	(2.222)	(2.327)	(1.787)	(1.831)
Partial R^2	0.310	0.369	0.317	0.374	0.282	0.336	0.315	0.371	0.244	0.328
KP *rk* Wald *F*-stat	15.25	17.22	16.04	17.73	15.43	16.47	15.77	17.19	16.99	22.06

[continues]

Table OB.1 Estimates of the Income Equation using Actual Data and the "OLS instrument": Additional Controls Included (cont.)

	PANEL B									
	Model (10)		Model (11)		Model (12)		Model (13)		Model (14)	
	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$	IV- $\hat{T}_i^{All}$	IV-$\hat{T}_i^{Pos}$
Income regressions:										
Trade share$_i$	0.133	0.356	0.094	0.396	0.782**	0.895***	0.773*	1.099***	0.823**	1.097***
	(0.351)	(0.322)	(0.402)	(0.346)	(0.306)	(0.284)	(0.406)	(0.377)	(0.393)	(0.352)
Ln population$_i$	-0.066	-0.058	-0.043	-0.034	-0.007	0.005	0.032	0.062	0.021	0.051
	(0.053)	(0.054)	(0.058)	(0.058)	(0.077)	(0.077)	(0.078)	(0.083)	(0.087)	(0.091)
Ln area$_i$	0.074	0.108	0.066	0.112	0.132**	0.140**	0.146*	0.178**	0.137**	0.159**
	(0.070)	(0.069)	(0.083)	(0.077)	(0.059)	(0.060)	(0.076)	(0.077)	(0.069)	(0.070)
Latitude$_i$	0.255	0.287	0.450**	0.475**	0.500**	0.483**	0.477*	0.426	0.322	0.258
	(0.189)	(0.198)	(0.221)	(0.224)	(0.209)	(0.213)	(0.278)	(0.289)	(0.323)	(0.335)
% Population in tropics$_i$	-1.499***	-1.527***	-1.622***	-1.647***	-1.268***	-1.261***	-1.715***	-1.661***	-1.944***	-1.915***
	(0.182)	(0.180)	(0.203)	(0.198)	(0.207)	(0.209)	(0.254)	(0.255)	(0.221)	(0.226)
IGRC-Index$_i$	2.425***	2.259***								
	(0.343)	(0.354)								
Corruption$_i$			1.669***	1.520***						
			(0.278)	(0.277)						
Executive constraint$_i$					0.210***	0.210***				
					(0.027)	(0.028)				
Ethno-ling. fract.$_i$							-0.716**	-0.785**		
							(0.319)	(0.319)		
Legal Origin$_i$									0.230*	0.255**
									(0.119)	(0.119)
Observations	90	90	90	90	94	94	95	95	96	96
First-stage regressions:										
$\hat{T}_i^*$	6.812***	7.614***	6.996***	8.003***	8.063***	8.949***	8.786***	9.748***	8.895***	9.813***
	(1.708)	(1.903)	(1.679)	(1.842)	(2.675)	(2.859)	(2.204)	(2.297)	(2.220)	(2.277)
Partial R^2	0.238	0.279	0.228	0.281	0.273	0.328	0.319	0.383	0.334	0.399
KP *rk* Wald *F*-stat	15.90	16.02	17.36	18.88	9.08	9.80	15.89	18.01	16.05	18.58

Notes. The dependent variable is the log of real GDP per capita. Robust standard errors are in parentheses. Standard errors in the income regressions are corrected following Frankel and Romer (1999) to account for the fact that the instruments depend on the parameters of the bilateral trade equation. $\hat{T}_i^{Pos*}$ is the predicted trade openness based only on predictions for positive observed bilateral trade shares. $\hat{T}_i^{All*}$ is the predicted trade openness based on all possible observations, i.e., including predictions for observations of zero or missing bilateral trade shares. The KP *rk* Wald *F*-stat is the Kleibergen-Paap *rk* Wald *F*-statistic for weak instruments. Exogenous variables are included in the first-stage regressions but not shown. *,**, *** Significant at 10, 5 and 1 percent, respectively.

References

Easterly, W., Levine, R., 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics, 112*(4), 1203–1250.

Frankel, J. A., & Romer, D. (1999). Does Trade Cause Growth? *American Economic Review*, *89*(3), 379–399.

Gallup, John & Sachs, Jeffrey & Mellinger, Andrew. (1999). Geography and Economic Development. International Regional Science Review. 22. 179-232.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2008. The Economic Consequences of Legal Origins. *Journal of Economic Literature, 46*(2), 285–332.

Knack, S., Keefer, P. (1995). Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures. *Economics & Politics, 7(3)*, 207-227.

Noguer, M., & Siscart, M. (2005). Trade Raises Income: A Precise and Robust Result. *Journal of International Economics*, *65*(2), 447–460.

Polity IV Project, 2014. Polity IV Project, Political Regime Characteristics and Transitions, 1800-2015. Centre for Systemic Peace, Virginia.

Online Appendix C. Poisson pseudo-maximum-likelihood (PPML) estimates

Table OC.1 reports estimates for the bilateral trade equation using PPML estimation. Because the bilateral trade equation only includes geographic-specific variables that explain bilateral trade in a gravity model, this follows the literature that estimates the gravity equation for trade. PPML estimation allows consistent estimates to be obtained in the presence of errors whose variance depends on the regressors, which occurs when the original gravity equation is log-linearized (Santos Silva and Tenreyro, 2006). Also, PPML estimation provides estimates based on all bilateral trade observations, including zeros. The estimates in Table OC.1 are obtained after all missing observations of bilateral trade shares are replaced by zero. These estimates are comparable to the estimates generated by excluding the missing observations, i.e., the missing observations are not converted to zero, and they are also comparable to the estimates generated if only positive values of the dependent variable are included in the regression. Thus, selection is not an issue in our application. The full set of results are available upon request.

Compared to the estimates in Table 1 of the main text, the estimates in Table OC.1 have the same sign but tend to be smaller in absolute value. However, the instruments generated from PPML predictions, $\hat{T}_{i,ppml}^{Pos*}$ and $\hat{T}_{i,ppml}^{All*}$, correlate almost perfectly and significantly with those obtained from OLS predictions.

Table OC.2 reports the 2SLS and IV results for the four specifications of the income equation from the main text using, respectively, $\hat{T}_{i,ppml}^{Pos}$ and $\hat{T}_{i,ppml}^{All}$ as instruments for trade. Table OC.3 reports the results for the ten additional specifications described in Online Appendix B. Considering the coefficients of trade openness across all fourteen specifications, they are significant at the 10% levels in at least five out of the fourteen cases when the instrument is $\hat{T}_{i,ppml}^{All}$, and in twelve out of fourteen cases when the instrument is $\hat{T}_{i,ppml}^{Pos}$. This is in line with what we find and report in the main text based on the instruments generated from OLS predicted bilateral trade shares.

Table OC.1 Estimation of the Bilateral Trade Equation using PPML

	Variable	Border interaction
Constant	-4.535***	1.397
	(0.722)	(2.814)
Ln distance$_{ij}$	-0.793***	0.182
	(0.059)	(0.446)
Ln population$_i$	-0.307***	-0.062
	(0.054)	(0.165)
Ln population$_j$	0.824***	0.081
	(0.043)	(0.198)
Ln area$_i$	-0.099**	-0.054
	(0.042)	(0.198)
Ln area$_j$	-0.191***	-0.142
	(0.048)	(0.247)
Landlocked$_{ij}$	-0.999***	0.525**
	(0.082)	(0.228)
Observations	15,778	
R^2	0.196	

Note. The dependent variable is τ_{ji}/GDP_i, where all observations of bilateral trade share have a minimum value of zero, i.e., there are no missing observations for the 98 countries and their 161 possible trading partners. The left column reports the coefficient of the variable listed, and the right column shows the coefficient of the interaction between the variable in the first column and border. Robust standard errors are in parentheses. **, *** significant at 5 and 1 percent, respectively.

Table OC.2 Estimates of the Income Equation using Actual Data and the "PPML instrument"

	PANEL A					
	Model (1)			Model (2)		
	OLS	IV- $\hat{T}^{All}_{i,ppml}$	IV- $\hat{T}^{Pos}_{i,ppml}$	OLS	IV- $\hat{T}^{All}_{i,ppml}$	IV- $\hat{T}^{Pos}_{i,ppml}$
Income regressions:						
Trade share$_i$	0.911***	2.620***	2.955***	0.578***	1.040**	1.182***
	(0.306)	(0.792)	(0.838)	(0.204)	(0.421)	(0.363)
Ln population$_i$	0.271***	0.393***	0.417***	0.106	0.143*	0.155**
	(0.102)	(0.137)	(0.147)	(0.072)	(0.079)	(0.079)
Ln area$_i$	-0.087	0.102	0.140	-0.087	-0.037	-0.022
	(0.088)	(0.135)	(0.138)	(0.065)	(0.076)	(0.073)
Distance to equator$_i$				4.158***	4.031***	3.991***
				(0.326)	(0.335)	(0.343)
Obs.	98	98	98	98	98	98
R^2	0.145	-	-	0.600	-	-
First-stage regressions:						
$\hat{T}^{*}_{i,ppml}$	-	1.435***	1.572***	-	1.470***	1.738***
		(0.296)	(0.321)		(0.382)	(0.445)
Partial R^2	-	0.245	0.303	-	0.228	0.301
KP *rk* Wald *F*-stat	-	23.58	23.98	-	14.81	15.27
	PANEL B					
	Model (3)			Model (4)		
	OLS	IV- $\hat{T}^{All}_{i,ppml}$	IV- $\hat{T}^{Pos}_{i,ppml}$	OLS	IV- $\hat{T}^{All}_{i,ppml}$	IV- $\hat{T}^{Pos}_{i,ppml}$
Income regressions:						
Trade share$_i$	0.636***	1.230***	1.551***	0.704***	1.346**	1.486***
	(0.205)	(0.473)	(0.440)	(0.254)	(0.568)	(0.502)
Ln population$_i$	0.072	0.121	0.148*	-0.037	0.059	0.080
	(0.076)	(0.082)	(0.086)	(0.104)	(0.121)	(0.113)
Ln area$_i$	-0.082	-0.017	0.017	0.040	0.083	0.092
	(0.070)	(0.086)	(0.084)	(0.065)	(0.073)	(0.071)
% Land in tropics$_i$	-1.580***	-1.521***	-1.489***			
	(0.167)	(0.170)	(0.174)			
Sub-Saharan Africa$_i$				-1.889***	-1.786***	-1.763***
				(0.206)	(0.219)	(0.214)
East Asia$_i$				-0.626*	-0.887**	-0.943**
				(0.340)	(0.409)	(0.392)
Latin America$_i$				-0.581**	-0.392	-0.351
				(0.221)	(0.258)	(0.241)
Obs.	98	98	98	98	98	98
R^2	0.547	-	-	0.594	-	-
First-stage regressions:						
$\hat{T}^{*}_{i,ppml}$	-	1.477***	1.683***	-	1.265***	1.557***
		(0.350)	(0.394)		(0.287)	(0.333)
Partial R^2	-	0.233	0.300	-	0.194	0.271
KP *rk* Wald *F*-stat	-	17.75	18.20	-	19.41	21.82

Notes. The dependent variable is the log of real GDP per capita reported by PWT Mark 5.6 for the year 1985. Robust standard errors are in parentheses. Standard errors in the IV income regressions are corrected following Frankel and Romer's (1999) approach to account for the fact that the instruments depend on the standard errors of the bilateral trade equation. $\hat{T}^{Pos*}_{i,ppml}$ is the predicted trade openness based only on predictions for positive observed bilateral trade shares. $\hat{T}^{All*}_{i,ppml}$ is the predicted trade openness based on all possible observations, i.e., including predictions for observations of zero or missing bilateral trade shares. The KP *rk* Wald *F*-stat is the Kleibergen-Paap *rk* Wald *F*-statistic for weak instruments. Exogenous variables are included in the first-stage regressions but not shown. *,**, *** indicate significance at 10, 5 and 1 percent, respectively.

Table OC.3 Estimates of the Income Equation using Actual Data and the "PPML instrument": Additional Controls Included

	PANEL A									
	Model (5)		**Model (6)**		**Model (7)**		**Model (8)**		**Model (9)**	
	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$
Income regressions:										
Trade share$_i$	0.452	0.951***	1.233**	1.569***	0.545	0.843***	0.563	1.008***	0.688	1.095**
	(0.487)	(0.369)	(0.500)	(0.479)	(0.400)	(0.321)	(0.450)	(0.352)	(0.591)	(0.475)
Ln population$_i$	-0.052	-0.005	0.122	0.152	0.014	0.039	-0.022	0.019	-0.047	0.015
	(0.080)	(0.078)	(0.092)	(0.100)	(0.068)	(0.067)	(0.076)	(0.076)	(0.111)	(0.097)
Ln area$_i$	0.123*	0.169**	-0.018	0.017	0.036	0.066	0.064	0.106	0.114	0.139**
	(0.074)	(0.072)	(0.086)	(0.084)	(0.068)	(0.065)	(0.073)	(0.071)	(0.070)	(0.070)
Latitude$_i$	0.639**	0.577**	-0.006	-0.041			0.208	0.158	0.305	0.281
	(0.271)	(0.283)	(0.346)	(0.367)			(0.259)	(0.273)	(0.373)	(0.397)
% Population in tropics$_i$	-2.036***	-1.986***			-1.299***	-1.282***	-1.477***	-1.438***	-1.302***	-1.277***
	(0.204)	(0.206)			(0.219)	(0.224)	(0.224)	(0.232)	(0.287)	(0.283)
Distance to equator$_i$					2.464***	2.404***				
					(0.375)	(0.387)				
% Land in tropics$_i$			-1.523***	-1.500***			-0.718***	-0.711***		
			(0.188)	(0.192)			(0.206)	(0.210)		
Sub-Saharan Africa$_i$									-0.879***	-0.840**
									(0.329)	(0.334)
East Asia$_i$									-0.346	-0.523
									(0.424)	(0.401)
Latin America$_i$									-0.163	-0.058
									(0.323)	(0.309)
Observations	98	98	98	98	98	98	98	98	98	98
First-stage regressions:										
$\hat{T}_i^*$	1.638***	1.934***	1.572***	1.838***	1.524***	1.775***	1.655***	1.942***	1.374***	1.652***
	(0.450)	(0.494)	(0.418)	(0.462)	(0.389)	(0.454)	(0.456)	(0.497)	(0.371)	(0.397)
Partial R^2	0.238	0.320	0.238	0.314	0.232	0.303	0.241	0.322	0.202	0.279
KP *rk* Wald *F*-stat	13.25	15.30	14.11	15.85	15.33	15.31	13.15	15.28	13.67	17.35

[continues]

Table OC.3 Estimates of the Income Equation using Actual Data and the "PPML instrument": Additional Controls Included (cont)

	PANEL B									
	Model (10)		Model (11)		Model (12)		Model (13)		Model (14)	
	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$	IV- $\hat{T}_{i,ppml}^{All}$	IV- $\hat{T}_{i,ppml}^{Pos}$
Income regressions:										
Trade share$_i$	0.114	0.456	-0.006	0.460	0.416	0.710**	0.633	1.084***	0.670	1.079***
	(0.396)	(0.301)	(0.485)	(0.330)	(0.431)	(0.314)	(0.466)	(0.374)	(0.449)	(0.347)
Ln population$_i$	-0.067	-0.055	-0.045	-0.032	-0.046	-0.014	0.019	0.061	0.004	0.049
	(0.054)	(0.055)	(0.060)	(0.058)	(0.086)	(0.077)	(0.076)	(0.078)	(0.086)	(0.086)
Ln area$_i$	0.071	0.123**	0.051	0.122*	0.106*	0.127**	0.132*	0.177**	0.124*	0.158**
	(0.069)	(0.060)	(0.087)	(0.068)	(0.059)	(0.058)	(0.078)	(0.077)	(0.070)	(0.071)
Latitude$_i$	0.252	0.302	0.442**	0.480**	0.553***	0.510**	0.499*	0.428	0.357	0.263
	(0.188)	(0.199)	(0.222)	(0.224)	(0.207)	(0.204)	(0.280)	(0.290)	(0.318)	(0.328)
% Population in tropics$_i$	-1.497***	-1.540***	-1.614***	-1.653***	-1.290***	-1.272***	-1.737***	-1.664***	-1.961***	-1.917***
	(0.184)	(0.182)	(0.208)	(0.201)	(0.208)	(0.209)	(0.258)	(0.252)	(0.221)	(0.224)
IGRC-Index$_i$	2.439***	2.184***								
	(0.357)	(0.342)								
Corruption$_i$			1.718***	1.488***						
			(0.311)	(0.287)						
Executive constraint$_i$					0.213***	0.211***				
					(0.026)	(0.027)				
Ethno-ling. fract.$_i$							-0.687**	-0.782**		
							(0.325)	(0.315)		
Legal Origin$_i$									0.217*	0.253**
									(0.117)	(0.115)
Observations	90	90	90	90	94	94	95	95	96	96
First-stage regressions:										
$\hat{T}_i^*$	1.355***	1.592***	1.338***	1.637***	1.493**	1.821***	1.685***	1.999***	1.709***	2.002***
	(0.361)	(0.410)	(0.364)	(0.409)	(0.563)	(0.661)	(0.449)	(0.492)	(0.449)	(0.490)
Partial R^2	0.194	0.256	0.174	0.249	0.195	0.273	0.243	0.334	0.257	0.347
KP *rk* Wald *F*-stat	14.08	15.06	13.54	16.04	7.02	7.58	14.05	16.51	14.49	16.67

Notes. The dependent variable is the log of real GDP per capita reported by PWT Mark 5.6 for the year 1985. Robust standard errors are in parentheses. Standard errors in the income regressions are corrected following Frankel and Romer (1999) approach to account for the fact that the instruments depend on the standard errors of the bilateral trade equation. $\hat{T}_i^{Pos*}$ is the predicted trade openness based only on predictions for positive observed bilateral trade shares. $\hat{T}_i^{All*}$ is the predicted trade openness based on all possible observations, i.e., including predictions for observations of zero or missing bilateral trade shares. The KP *rk* Wald *F*-stat is the Kleibergen-Paap *rk* Wald *F*-statistic for weak instruments. Exogenous variables are included in the first-stage regressions but not shown. *, **, *** indicate significant at 10, 5 and 1 percent, respectively.

References

Frankel, J. A., & Romer, D. (1999). Does Trade Cause Growth? American Economic Review, 89(3), 379–399.

Santos Silva, JMC, and Tenreyro, S. (2006). The log of gravity. The Review of Economics and Statistics, 88(4), 641-658.

Online Appendix D. Full results Monte Carlo Simulations

Tables OD.1 and OD.2 summarize 2SLS estimates for all fourteen specifications of the income equation using $\tilde{T}_i^{All}$ and $\tilde{T}_i^{Pos}$ as instruments for trade. These results are based on 1,000 replications for each model. Online Appendix B describes the additional controls used in Table OD.2.

Table OD.1 Estimates of the Income Equation using Randomized Instruments (1,000 replications)

	Model (1)		**Model (2)**		**Model (3)**		**Model (4)**	
	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$
Income regressions:								
Trade share$_i$	3.319	6.609	0.396	3.523	2.395	4.203	0.814	6.109
	(255.203)	(0.848)	(45.571)	(0.652)	(48.156)	(0.688)	(59.637)	(1.716)
	[10]	[1000]	[4]	[995]	[5]	[1000]	[8]	[987]
	{3}	{999}	{2}	{989}	{2}	{996}	{2}	{969}
Ln population$_i$	0.443	0.677	0.092	0.344	0.217	0.366	-0.021	0.773
	(18.165)	(0.060)	(3.672)	(0.053)	(3.967)	(0.057)	(8.938)	(0.257)
	[115]	[1000]	[4]	[960]	[2]	[942]	[1]	[951]
	{46}	{997}	{0}	{398}	{0}	{160}	{0}	{861}
Ln area$_i$	0.180	0.544	-0.107	0.232	0.109	0.305	0.047	0.402
	(28.271)	(0.094)	(4.937)	(0.071)	(5.228)	(0.075)	(4.001)	(0.115)
	[4]	[560]	[17]	[531]	[13]	[642]	[2]	[780]
	{1}	{24}	{2}	{29}	{2}	{67}	{0}	{191}
Distance to equator$_i$			4.209	3.344				
			(12.606)	(0.180)				
			[670]	[995]				
			{640}	{993}				
% Land in tropics$_i$					-1.406	-1.228		
					(4.752)	(0.068)		
					[667]	[998]		
					{634}	{995}		
Sub-Saharan Africa$_i$							-1.871	-1.021
							(9.571)	(0.275)
							[607]	[820]
							{564}	{718}
East Asia$_i$							-0.670	-2.822
							(24.235)	(0.697)
							[37]	[962]
							{18}	{893}
Latin America$_i$							-0.548	1.010
							(17.552)	(0.505)
							[60]	[203]
							{26}	{0}
Obs.	98	98	98	98	98	98	98	98

Notes. The table reports average values from 1,000 replications. The standard deviation of this average is reported in parentheses. For the estimated coefficients, the number of replications that produce an estimate significant at least at the 10% level is in [square brackets], and the number of replications in which the estimate is significant at least at the 5% level is in {curly brackets}. For each replication, standard errors in the income regressions are corrected following Frankel and Romer (1999) approach, to account for the fact that the instruments depend on the parameters of the bilateral trade equation. The number of times the Kleibergen-Paap *rk* Wald *F*-statistic for weak instruments is greater than 10 is in <angle brackets>.

Table OD.2 Estimates of the Income Equation using Randomized Instruments: Additional Controls Included

	PANEL A									
	Model (5)		**Model (6)**		**Model (7)**		**Model (8)**		**Model (9)**	
	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$
Income regressions:										
Trade share$_i$	3.541	4.102	-0.888	4.322	1.150	3.218	2.532	3.758	-6.162	5.685
	(38.602)	(0.720)	(95.792)	(0.738)	(20.272)	(0.593)	(56.133)	(0.669)	(195.755)	(1.985)
	[8]	[998]	[7]	[998]	[6]	[997]	[9]	[998]	[11]	[986]
	{2}	{992}	{2}	{992}	{1}	{991}	{2}	{992}	{2}	{960}
Ln population$_i$	0.238	0.291	-0.066	0.395	0.065	0.240	0.162	0.276	-1.087	0.712
	(3.628)	(0.068)	(8.478)	(0.065)	(1.714)	(0.050)	(5.250)	(0.063)	(29.725)	(0.301)
	[0]	[0]	[0]	[552]	[0]	[12]	[0]	[0]	[1]	[923]
	{0}	{0}	{0}	{1}	{0}	{0}	{0}	{0}	{0}	{733}
Ln area$_i$	0.408	0.460	-0.236	0.300	0.098	0.310	0.249	0.364	-0.308	0.422
	(3.560)	(0.066)	(9.846)	(0.076)	(2.084)	(0.061)	(5.272)	(0.063)	(12.060)	(0.122)
	[40]	[998]	[11]	[385]	[1]	[993]	[5]	[990]	[26]	[856]
	{13}	{979}	{2}	{11}	{0}	{956}	{0}	{860}	{7}	{339}
Latitude$_i$	0.259	0.190	0.215	-0.329			-0.014	-0.153	0.707	0.012
	(4.748)	(0.089)	(10.002)	(0.077)			(6.339)	(0.076)	(11.482)	(0.116)
	[244]	[0]	[0]	[0]			[0]	[0]	[0]	[0]
	{110}	{0}	{0}	{0}			{0}	{0}	{0}	{0}
% Population in tropics$_i$	-1.726	-1.670			-1.264	-1.146	-1.305	-1.199	-1.726	-0.993
	(3.873)	(0.072)			(1.160)	(0.034)	(4.888)	(0.058)	(12.123)	(0.123)
	[719]	[998]			[717]	[992]	[677]	[975]	[664]	[371]
	{685}	{998}			{684}	{981}	{642}	{908}	{611}	{148}
Distance to equator$_i$					2.342	1.925				
					(4.086)	(0.120)				
					[611]	[844]				
					{562}	{558}				
% Land in tropics$_i$			-1.666	-1.314			-0.684	-0.663		
			(6.462)	(0.050)			(0.975)	(0.012)		
			[731]	[998]			[691]	[634]		
			{705}	{998}			{627}	{168}		

[continues]

Table OD.2 Estimates of the Income Equation using Randomized Instruments: Additional Controls Included (cont.)

	Model (5)		**Model (6)**		**Model (7)**		**Model (8)**		**Model (9)**	
	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$
Income regressions (cont.):										
Sub-Saharan Africa$_i$									-1.537	-0.400
									(18.797)	(0.191)
									[462]	[4]
									{351}	{0}
East Asia$_i$									2.621	-2.511
									(84.792)	(0.860)
									[5]	[964]
									{0}	{882}
Latin America$_i$									-1.924	1.123
									(50.353)	(0.511)
									[0]	[135]
									{0}	{0}
Observations	98	98	98	98	98	98	98	98	98	98
First-stage regressions:										
$\tilde{T}_i^*$	5.591	28.950	5.127	29.581	5.492	31.760	5.563	29.846	5.487	21.384
	(27.723)	(5.795)	(28.213)	(5.646)	(27.969)	(6.582)	(28.165)	(5.951)	(25.398)	(4.898)
	[115]	[997]	[115]	[997]	[126]	[995]	[112]	[997]	[136]	[980]
	{52}	{987}	{49}	{992}	{51}	{979}	{52}	{986}	{63}	{943}
Partial R^2	0.012	0.109	0.012	0.115	0.012	0.117	0.012	0.112	0.013	0.072
	(0.017)	(0.031)	(0.016)	(0.031)	(0.017)	(0.033)	(0.017)	(0.031)	(0.017)	(0.027)
KP *rk* Wald *F*-stat	1.160	10.917	1.140	11.458	1.163	9.632	1.153	10.818	1.230	9.846
	<1>	<646>	<0>	<706>	<1>	<468>	<1>	<633>	<2>	<505>

[continues]

Table OD.2 Estimates of the Income Equation using Randomized Instruments: Additional Controls Included (cont.)

	PANEL B									
	Model (10)		**Model (11)**		**Model (12)**		**Model (13)**		**Model (14)**	
	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$
Income regressions:										
Trade share$_i$	-0.375	2.140	1.384	-4.425	0.869	2.898	-2.270	3.840	1.490	3.936
	(35.116)	(47.475)	(40.107)	(273.753)	(59.465)	(0.585)	(114.702)	(0.703)	(27.250)	(0.683)
	[2]	[507]	[1]	[891]	[11]	[994]	[7]	[998]	[8]	[998]
	{0}	{203}	{0}	{795}	{3}	{958}	{2}	{991}	{2}	{992}
Ln population$_i$	-0.084	0.005	-0.006	-0.172	0.003	0.219	-0.249	0.316	0.094	0.363
	(1.247)	(1.685)	(1.145)	(7.817)	(6.333)	(0.062)	(10.610)	(0.065)	(2.993)	(0.075)
	[0]	[0]	[0]	[0]	[1]	[0]	[0]	[5]	[0]	[119]
	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}	{0}
Ln area$_i$	-0.003	0.379	0.262	-0.621	0.138	0.284	-0.154	0.449	0.191	0.391
	(5.329)	(7.204)	(6.095)	(41.604)	(4.270)	(0.042)	(11.323)	(0.069)	(2.224)	(0.056)
	[6]	[515]	[8]	[857]	[75]	[985]	[32]	[998]	[32]	[996]
	{0}	{240}	{2}	{726}	{21}	{866}	{9}	{981}	{7}	{959}
Latitude$_i$	0.180	0.548	0.556	0.078	0.487	0.191	0.957	-0.007	0.167	-0.399
	(5.139)	(6.947)	(3.301)	(22.534)	(8.678)	(0.085)	(18.102)	(0.111)	(6.308)	(0.158)
	[0]	[0]	[194]	[18]	[218]	[0]	[66]	[0]	[3]	[0]
	{0}	{0}	{32}	{1}	{106}	{0}	{7}	{0}	{0}	{0}
% Population in tropics$_i$	-1.435	-1.752	-1.730	-1.245	-1.262	-1.139	-2.213	-1.212	-1.872	-1.609
	(4.427)	(5.985)	(3.346)	(22.841)	(3.608)	(0.036)	(18.792)	(0.115)	(2.933)	(0.074)
	[663]	[907]	[725]	[986]	[736]	[998]	[610]	[976]	[711]	[998]
	{623}	{880}	{699}	{979}	{714}	{994}	{569}	{948}	{690}	{998}
IGRC-Index$_i$	2.804	0.929								
	(26.171)	(35.382)								
	[297]	[14]								
	{219}	{6}								
Corruption$_i$			1.033	3.897						
			(19.774)	(134.970)						
			[257]	[9]						
			{188}	{2}						

[continues]

Table OD.2 Estimates of the Income Equation using Randomized Instruments: Additional Controls Included (cont.)

	Model (10)		**Model (11)**		**Model (12)**		**Model (13)**		**Model (14)**	
	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$	IV- $\tilde{T}_i^{All}$	IV- $\tilde{T}_i^{Pos}$
Income regressions (cont.):										
Executive constraint$_i$					0.210	0.196				
					(0.395)	(0.004)				
					[769]	[999]				
					{745}	{998}				
Ethno-ling. fract.$_i$							-0.071	-1.367		
							(24.315)	(0.149)		
							[154]	[999]		
							{75}	{994}		
Legal Origin$_i$									0.290	0.508
									(2.432)	(0.061)
									[76]	[999]
									{23}	{995}
Observations	90	90	90	90	94	94	95	95	96	96
First-stage regressions:										
$\tilde{T}_i^*$	1.621	15.267	2.210	21.142	5.574	29.212	5.577	30.037	5.981	29.452
	(25.598)	(6.173)	(26.425)	(6.332)	(26.837)	(6.560)	(29.502)	(6.177)	(28.562)	(5.841)
	[124]	[509]	[127]	[886]	[114]	[976]	[118]	[996]	[121]	[997]
	{62}	{253}	{63}	{789}	{46}	{895}	{56}	{983}	{51}	{983}
Partial R^2	0.013	0.031	0.013	0.058	0.013	0.101	0.013	0.116	0.013	0.116
	(0.018)	(0.021)	(0.018)	(0.027)	(0.018)	(0.032)	(0.017)	(0.033)	(0.017)	(0.032)
KP *rk* Wald *F*-stat	1.178	2.927	1.170	6.202	1.152	6.039	1.171	11.219	1.176	10.620
	<1>	<3>	<1>	<76>	<1>	<21>	<3>	<669>	<2>	<611>

Notes. The table reports average values from 1,000 replications. The standard deviation of this average is reported in parentheses. For the estimated coefficients, the number of replications that produce an estimate significant at least at the 10% level is in [square brackets], and the number of replications in which the estimate is significant at least at the 5% level is in {curly brackets}. For each replication, standard errors in the income regressions are corrected following Frankel and Romer's (1999) approach to account for the fact that the instruments depend on the parameters of the bilateral trade equation. The number of times the Kleibergen-Paap *rk* Wald *F*-statistic for weak instruments is greater than 10 is in <angle brackets>.

Online Appendix E. Trade Costs, Trade and Income

E.1 Number of Trading Partners, Trade Costs and Income Per Capita

To gain more insight into the determinants of number of trade partners, N_i, we examine how various measures of trade costs affect the number of a country's trade partners. To achieve this, we collect data on countries' regulation costs of firm entry from Djankov et al. (2002) and the quality of infrastructure from Limão and Venables (2001).[2] Following Helpman et al. (2008) we use firm entry regulation costs as measures of the fixed costs faced by firms exporting to or from other countries. The data underlying our variables are based on firm entry regulation costs as a percentage of GDP per capita as well as the number of days and legal procedures that are required for an entrepreneur to legally start operating a business. Specifically, we consider two indicators of high fixed-costs of trade: 1) I(*High Regulation Cost*)$_i$, which takes the value of 1 if country *i*'s relative costs are above the cross-country median; and 2) I(*High # of days and procedures*)$_i$, which equals 1 if country *i*'s required number of days and legal procedures are above the median, and zero otherwise. For the quality of infrastructure, we use two indexes constructed by Limão and Venables (2001). The first index, '*Own Infrastructure*', is estimated as the average of road density, rail density, number of telephone lines per capita raised to the power of -0.3. The other index, '*Transit Infrastructure*', applies to landlocked countries only and it is the average infrastructure index of the transit countries that a country needs to pass through to reach the sea. For both infrastructure indexes, a higher value indicates worse infrastructure.

Table OE.1. Number of partners, trading costs and income per capita

	N_i	N_i	N_i
	(1)	(2)	(3)
I(*High Regulation Cost*)$_i$	-0.3740***	-0.1831***	-0.0698
	(0.0712)	(0.0653)	(0.0708)
I(*High # of Days and*	-0.1563**	-0.0974	-0.0661
Procedures)$_i$	(0.0774)	(0.0589)	(0.0633)
Own Infrastructure$_i$		-0.2333***	-0.1671***
		(0.0413)	(0.0441)
Transit Infrastructure$_i$		-0.1356***	-0.1025*
		(0.0591)	(0.0564)
Log *Income per Capita*$_i$			0.1186***
			(0.0350)
Obs.	92	88	88

Notes. Robust standard errors are reported in parentheses *, **, *** indicate significance at 10, 5 and 1 percent, respectively.

[2] Dictated by data availability, we use regulation costs for 1999 and infrastructure data for 1990.

We estimate the relationship between the number of trading partners, trade costs and income using a negative binomial model to account for the count nature of the dependent variable. The estimation results are shown in Table OE.1.

The results in column (1) show that high trading costs are associated with a small number of trading partners, as implied by the significantly negative coefficient of the number of days to start a business and legal procedures. The model by Helpman et al. (2008) offers a framework to rationalize this pattern. Extending the model of Melitz (2003) to include fixed costs of exporting and bounded productivity distributions, Helpman et al. (2008) show that some countries do not trade with each other because the firms are not sufficiently productive to penetrate each other's markets. In this framework, destinations with lower fixed costs of exporting are, *ceteris paribus*, more likely to trade with other countries.

Turning to Column (2) in Table OE.1, it is evident that the number of trading partners is a decreasing function of underprovided infrastructure. These coefficient estimates become smaller and less significant when we include a country's income per capita as a control due to the high negative correlation between income per capita and trading costs. Income per capita correlates positively with N_i, as already shown in Figure 1 of the main text. Overall, these results indicate that trade costs are a contributing factor to the positive relationship between the number of trade partners and per capita income.

E.2 Missing versus Zero Bilateral Trade

In this subsection, we first show that zero and missing bilateral trade flows are non-random events determined by similar factors, including income per capita. Second, we bring some evidence that missing bilateral shares in our sample are likely to be unobserved zeros.

To investigate the determinants of zeros and missing values, we create two indicator variables: The first indicator takes the value of 1 for positive values of trade and 0 for *zero* bilateral trade (thus, de facto excluding the missing values). The second indicator takes the value of 1 for positive values of trade and 0 for *missing* trade values (thus, de facto excluding the observed zeros). We then estimate a probit model for each of these indicator variables following the specification of Helpman et al. (2008, Table I). We source data on common legal systems, common language, common currency, regional trade agreement between countries *i* and *j*, and World Trade Organization (WTO) membership from Head et al. (2010), while data on island status, and common religion[3] are taken from Helpman et al. (2008).

[3] Helpman et al. (2008) construct the variable common religion as follows: (% Protestants in reporter country X % Protestants in partner country) + (% Catholics in reporter country X % Catholics in partner country) + (% Muslims in reporter X % Muslims in partner country).

Table OE.2 Probability of positive bilateral trade versus zero or missing values

	Prob ($\tau_{ij}\neq0$) excluding missing values	Prob($\tau_{ij}\neq.$) excluding zeros	Prob ($\tau_{ij}\neq0$) excluding missing values	Prob($\tau_{ij}\neq.$) excluding zeros
	(1)	(2)	(3)	(4)
Real GDP per capita$_i$ (log)	0.410***	0.455***		
	(0.014)	(0.017)		
Distance$_{ij}$ (log)	-0.266***	-0.384***	-0.287***	-0.381***
	(0.023)	(0.031)	(0.026)	(0.034)
Border$_{ij}$	0.512***	0.336	0.552***	0.388*
	(0.134)	(0.204)	(0.140)	(0.203)
Number of islands$_{ij}$	-0.067**	-0.194***	-0.085**	-0.523***
	(0.031)	(0.036)	(0.042)	(0.047)
Sum landlocked$_{ij}$	-0.225***	-0.613***	-0.469***	-0.970***
	(0.028)	(0.029)	(0.041)	(0.044)
Common legal system$_{ij}$	-0.182***	-0.206***	-0.165***	-0.211***
	(0.031)	(0.035)	(0.034)	(0.040)
Language$_{ij}$	0.118***	0.623***	0.184***	0.735***
	(0.041)	(0.055)	(0.045)	(0.062)
Common currency$_{ij}$	-0.164	-0.465***	-0.067	-0.259
	(0.130)	(0.168)	(0.138)	(0.180)
Regional trade agreement$_{ij}$	0.298	0.516	0.312	0.486
	(0.182)	(0.456)	(0.230)	(0.580)
Common religion$_{ij}$	0.380***	0.278***	0.475***	0.291***
	(0.057)	(0.067)	(0.065)	(0.081)
WTO_d1 (=1 if both NOT in WTO)	-0.509***	-0.672***	-0.675***	-1.075***
	(0.047)	(0.047)	(0.054)	(0.058)
WTO_d2 (=1 if both in WTO)	0.411***	0.721***	0.648***	1.133***
	(0.030)	(0.036)	(0.037)	(0.047)
Observations	12,004	11,606	11,149+	11,606
Reporter country FE	No	No	Yes	Yes
Pseudo R^2	0.149	0.289	0.207	0.413

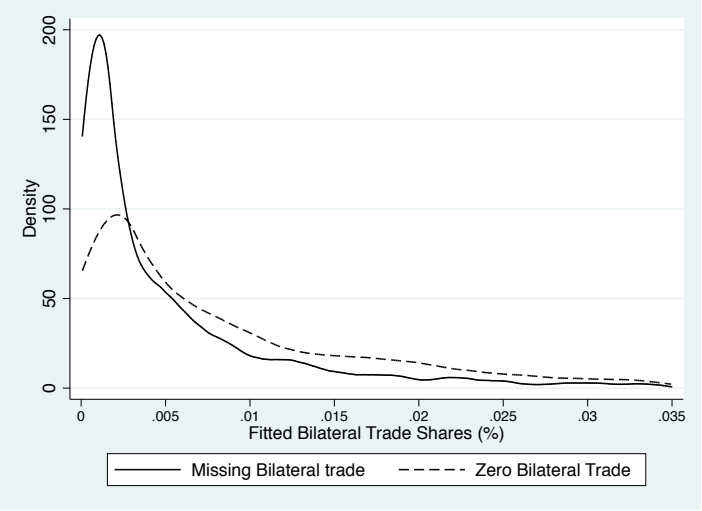
Notes. Probit estimates. Country *i* is the reporter country. Robust standard errors in parentheses. Colonial ties are omitted from the regressions because they cannot be identified when the missing observations are excluded. All the variables have similar signs as in Helpman et al. (2008), except common legal system. +The number of observations is lower than in column (1) because six reporter countries trade with all the partners for whom the data are available. *, **, *** indicate significance at 10, 5 and 1 percent, respectively.

Two key results emerge from the estimates of these probit models, which are reported in Table OE.2: First, the coefficients of income are significantly positive and, in terms of economic and statistical significance, are close to each other in the two regressions (columns (1) and (2)). This suggests that the probability of trade is positively related to per capita income regardless of whether the unobserved trade values are missing or zeros. Stated differently, the probability of reporting missing or zero bilateral trade is significantly higher for poor than for rich countries. Second, the coefficients of the covariates are quite similar in the two cases, even more so when reporting-country heterogeneity is controlled for through

reporting-country fixed effects (columns (3)-(4)) (note that the reporting country's per capita income drops out when these fixed effects are included in the regression). These results suggest that the probability of reporting zero or missing bilateral trade is determined by the same factors that are included in the model: per capita income, geographical and institutional factors.

To shed further light on the missing-zero distinction, the distributions of predicted bilateral trade shares from the bilateral trade model (Eq. (2)) is generated for missing and zero observations, respectively, are displayed in Figure OE.1 The relative position of both lines shows that the predicted values for missing observations are systematically lower than the predicted values for zero trade observations, suggesting that missing values are likely to be zero or really small values. The evidence in this sub-section suggests that treating zeros and missing bilateral shares in the same way is without loss of generality in our application.

Figure OE.1 Distribution of fitted bilateral trade shares for missing and zero bilateral trade



Note. The distributions are truncated at 0.035 for legibility. Note that more than 90% of all observations in each distribution are plotted in the figure.

References

Djankov, S., La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. (2002). The Regulation of Entry. *Quarterly Journal of Economics, 117*(1), 1–37.

Frankel, J. A., & Romer, D. (1999). Does Trade Cause Growth? *American Economic Review*, *89*(3), 379–399.

Head, K., Mayer, T., & Ries, J. (2010). The Erosion of Colonial Trade Linkages after Independence. *Journal of International Economics, 81*(1), 1–14.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics, 123*(2), 441-487.

Limão, N., & Venables, A. J. (2001). Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. *The World Bank Economic Review*, *15*(3), 451–479.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, *71*(6), 1695–1725.

Online Appendix F. Additional estimates using $\widehat{T}_i^{Pos*}$ while controlling for number of partners

Table OF.1 reports estimates for ten additional specifications of the income equation each estimated by 2SLS/IV using $\hat{T}_i^{Pos*}$ as the instrument for trade and controlling for the number of partners, N_i. More information on the data used can be found in Online Appendix B.

Table OF.1 Estimates of the Income Equation using $\widehat{T}_i^{Pos*}$ while controlling for the number of trading partners: additional controls included

	Model (5)		**Model (6)**		**Model (7)**		**Model (8)**		**Model (9)**	
	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$
Income regressions:										
Trade share$_i$	1.023***	-0.097	1.078***	-0.154	0.710**	-0.240	0.909***	-0.114	1.110***	0.082
	(0.353)	(0.313)	(0.407)	(0.353)	(0.288)	(0.336)	(0.330)	(0.314)	(0.413)	(0.318)
N_i		0.020***		0.023***		0.019***		0.020***		0.020***
		(0.002)		(0.003)		(0.002)		(0.002)		(0.002)
Observations	98	98	98	98	98	98	98	98	98	98
First-stage regressions:										
$\hat{T}_i^{Pos*}$	9.554***	8.893***	9.659***	8.931***	8.487***	7.585***	9.649***	8.923***	8.597***	8.225***
	(2.302)	(2.490)	(2.294)	(2.477)	(2.091)	(2.032)	(2.327)	(2.488)	(1.831)	(2.018)
Partial R^2	0.369	0.287	0.374	0.290	0.336	0.252	0.371	0.290	0.328	0.278
KP *rk* Wald *F*-stat	17.22	12.75	17.73	13.00	16.47	13.93	17.19	12.87	22.06	16.61

[continues]

Table OF.1 Estimates of the Income Equation using $\hat{T}_i^{Pos*}$ while controlling for the number of trading partners: additional controls included (cont.)

	Model (10)		**Model (11)**		**Model (12)**		**Model (13)**		**Model (14)**	
	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$	IV- $\hat{T}_i^{Pos*}$
Income regressions:										
Trade share$_i$	0.356	-0.241	0.396	-0.411	0.895***	0.124	1.099***	0.015	1.097***	-0.006
	(0.322)	(0.311)	(0.346)	(0.366)	(0.284)	(0.298)	(0.377)	(0.326)	(0.352)	(0.313)
N_i		0.013***		0.017***		0.014***		0.019***		0.020***
		(0.003)		(0.003)		(0.002)		(0.002)		(0.002)
Observations	90	90	90	90	94	94	95	95	96	96
First-stage regressions:										
$\hat{T}_i^{Pos*}$	7.614***	7.691***	8.003***	7.598***	8.949***	8.288***	9.748***	9.096***	9.813***	9.281***
	(1.903)	(2.111)	(1.842)	(2.038)	(2.859)	(2.925)	(2.297)	(2.480)	(2.277)	(2.482)
Partial R^2	0.279	0.254	0.281	0.233	0.328	0.247	0.383	0.294	0.399	0.313
KP *rk* Wald *F*-stat	16.02	13.28	18.88	13.90	9.797	8.031	18.01	13.45	18.58	13.98

Notes. The dependent variable is the log of real GDP per capita reported by PWT Mark 5.6 for the year 1985. Robust standard errors are in parentheses. Standard errors in the income regressions are corrected following Frankel and Romer (1999) approach, to account for the fact that the instruments depend on the parameters of the bilateral trade equation. $\hat{T}_i^{Pos*}$ is the predicted trade openness based only on predictions for observations of actual positive bilateral trade. N_i is the number of trading partners of country *i*. Model (1) controls for land area and population (in logs). Models (2), (3) and (4) include the distance to the equator, percentage of land in the tropics and continental dummies, respectively, as additional control variables. The KP *rk* Wald *F*-stat is the Kleibergen-Paap *rk* Wald *F*-statistic for weak instruments. Exogenous variables are included in the first-stage regressions but not shown. *, **, *** indicate significance at 10, 5 and 1 percent, respectively.