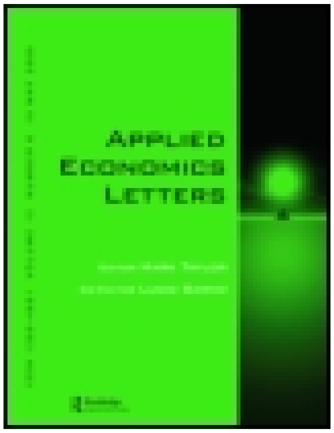
2-2015

A Simple Solution to the Distance Puzzle: Balanced Data and Poisson Estimation

Roger White
Whittier College, rwhite1@whittier.edu

David Buehler
Pennsylvania State University - Harrisburg, dlb74@psu.edu



Applied Economics Letters
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rael20

A simple solution to the distance puzzle: balanced data and Poisson estimation

David Buehler[a] & Roger White[b]

[a] School of Business Administration, Penn State University - Harrisburg, Middletown, PA 17057, USA

[b] Department of Economics, Whittier College, Whittier, CA 90506, USA

Published online: 29 Sep 2014.

Applied Economics Letters, 2014
http://dx.doi.org/10.1080/13504851.2014.959650

Routledge
Taylor & Francis Group

A simple solution to the distance puzzle: balanced data and Poisson estimation

David Buehler[a,*] and Roger White[b]

[a]*School of Business Administration, Penn State University – Harrisburg, Middletown, PA 17057, USA*
[b]*Department of Economics, Whittier College, Whittier, CA 90506, USA*

We propose the use of a balanced panel data set and Poisson regression as a solution to the distance puzzle. Employing annual data for the period 1972–2010, we confirm the existence of the puzzle by applying OLS regression to both an unbalanced panel data set and a narrowly defined balanced panel. We find that Poisson regression remedies the distance puzzle, producing a constant trend for the distance coefficient when the unbalanced panel is examined and a positive trend for the balanced data. The findings confirm the common intuition that the influence of transport costs on trade flows has decreased over time.

Keywords: distance puzzle; gravity model; imports; OLS; Poisson regression

JEL Classification: F14; F60

I. Introduction

As the workhorse for empirical studies of international trade the gravity model is nearly ubiquitous in the literature. The model is a proven and powerful tool for analysing trade flows between countries; however, a lingering by-product of the expansive and extensive use of the gravity model is what has become known as the ‘distance puzzle’. Since the distance variable serves as a proxy for, among other things, transportation costs, it may be expected that transport-related technological advances and the increased prevalence of services in international trade flows have resulted in time-specific distance coefficients that decrease in magnitude (or, possibly, that remain constant) over time (Cairncross, 1997). Several authors have, however, found that coefficients on the distance variable, when iterated annually, become increasingly negative, a phenomenon that seems counter-intuitive.[1]

The literature on the distance puzzle focuses on four main explanations: econometric methods, omitted variable bias, sample selection and composition effects (Carrère *et al.*, 2013).[2] The solution we

*Corresponding author. E-mail: dlb74@psu.edu

[1] See Disdier and Head (2008) and Carrère *et al.* (2013) for more extensive analyses of the topic. The puzzle is also called the ‘missing globalization puzzle’ by Coe *et al.* (2007) and Siliverstovs and Schumacher (2008).

[2] Another vein of literature focuses on the average distance of trade and the intensive and extensive margins of trade. See, for example, Carrère and Schiff (2005) and Berthelon and Freund (2008).

Table 1. Prior studies of the distance puzzle

Author(s) (Year)	Reference period	No. of countries	Balanced data?	Techniques used[a]	Puzzle solved?
Brun *et al.* (2005)	1962–1996	130	No	GLS and HT	Yes
Carrère and Schiff (2005)	1962–2000	150	No[b]	N/A	Mixed
Felbermayr and Kohler (2006)	1970–1990	175	No[c]	Tobit	Yes
Coe *et al.* (2007)	1975–2000	73	No[d]	Log-linear and nonlinear	Mixed
Siliverstovs and Schumacher (2008)	1970–2000	22	No[e]	OLS and BV-OLS	Mostly
Berthelon and Freund (2008)	1985–2005	100	Yes[f]/No	OLS	No
Boulhol and De Serres (2010)	1970–2005	32 (areas)[g]	No	OLS and PPML	No
Lin and Sim (2012)	1950–1999	175	No[h]	OLS and PPML	No
Yotov (2012)	1965–2005	93	No	OLS and PPML	Yes
Carrère *et al.* (2013)	1970–2006	124	No	OLS, HT, Tobit and PPML	Mixed

[a]GLS, generalized least squares; PPML, Poisson pseudo-maximum likelihood; HT, Hausman and Taylor estimator; BV-OLS, bonus vetus ordinary least squares.
[b]Mirror estimates used for missing data.
[c]Zeros input for missing trade values.
[d]Zero trade flows included.
[e]A minimal value is used to replace zero trade values.
[f]A balanced sample is used when industry-level data are examined.
[g]Some areas include multiple countries.
[h]Zeros input for missing trade values during robustness checks.

propose considers econometric methodology, omitted variable bias/specification issues and sample selection. We do not address composition effects, as Berthelon and Freund (2008) found that 'compositional shifts do not explain the increase in the importance of distance' (p. 319). Table 1 summarizes the literature. Prior works have considered a variety of reference periods, employed several different estimation techniques and typically examined cross-sectional or unbalanced panel data sets. The results of these previous studies, in terms of 'solving' the puzzle, are quite varied, ranging from resolving the puzzle to partial resolution to no resolution at all. Thus, we consider the distance puzzle an open empirical question.

We apply the OLS and Poisson estimation techniques to both an unbalanced panel data set and a narrowly defined balanced panel. The unbalanced data set includes all trading partner pairs for which data are available for the years 1972–2010. The balanced panel includes only those country pairs for which positive trade flows occurred in *every* year of our reference period, meaning any country pair for which data are either missing or reported as zero is excluded from the balanced panel.[3] This method of balancing the data also simultaneously addresses sample selection and the treatment of missing/zero trade values as explanations of the distance puzzle. Within this balanced sample, only the countries' transport costs relative to themselves evolve over time, which is what is expected to decline. For the unbalanced sample, trade costs may vary more as trade occurs between more countries, some of which may have higher trade costs and, thus, drive average trade costs up. By defining the balanced sample as such, we focus on the effects of distance on the intensive margin of trade, while the unbalanced data set includes changes at both the intensive and extensive margins. In doing so, we further examine the role of transport-related technological advances, which is central to the distance puzzle.

[3] The unbalanced panel data set includes 210 countries, while the balanced panel data set includes 65 countries. Using 1972 as the initial year in our reference period produces the largest number of observations in our balanced data set. Use of any earlier year results in fewer observations as the number of countries for which data are available decreases. Beginning the reference period more recently increases the number of country pairs; however, the number of observations in the data set decreases due to the inclusion of fewer annual observations.

II. Empirical Specification, Data and Variable Construction

Equation 1 represents our empirical specification. Our choice of explanatory variables follows that of Anderson and Van Wincoop (2003), who discuss both the theoretical foundations of the gravity model and the proper specification of the model. This specification addresses omitted variable bias as a possible explanation of the distance puzzle.

The dependent variable series M_{ijt} is the value of imports traded from origin country i to destination country j during year t (United Nations (UNCTAD), 2014).[4] We follow Felbermayr and Kohler (2006) and examine imports since they constitute a tax base and, thus, relative to export values are more likely to be accurately recorded.

$$\begin{aligned}\ln M_{ijt} = {} & \alpha_0 + \beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta_3 \ln REM_{it} \\ & + \beta_4 \ln REM_{jt} + \beta_5\, BORDER_{ij} \\ & + \sum_{t=1972}^{2010} \left[\beta_{6,t}\left(\ln DIST_{ij} \times \Omega_t\right)\right] + \beta_{\Omega_t}\Omega_t \\ & + \beta_{O_i} O_i + \beta_{D_j} D_j + \varepsilon_{ijt} \end{aligned} \quad (1)$$

Our variable of primary interest $DIST_{ij}$ is a population-weighted measure of the geodesic distance between trading partners (CEPII, 2014). Y_{it} and Y_{jt} are the GDP values of the origin and destination countries, respectively (World Bank, 2014). REM_{jt} is a measure of country j's economic remoteness (Head and Ries, 1998; CEPII, 2014; and World Bank, 2014). It is constructed as $1/\sum_{k=1}^{K}\left[(GDP_k/GDP_w)/DIST_{jk}\right]$, where GDP_w represents gross global product and k identifies potential trading partners for country j other than country i.[5] An analogous remoteness variable is constructed for country i. All monetary values are in year 2005 US Dollars. Representing another facet of geography-related trading costs, $BORDER_{ij}$ is a dummy variable that is equal to one if trading partners share a common border (CEPII, 2014). Lastly, we include time (Ω_t) and country (origin (O_i) and destination (D_j)) fixed effect terms to control for year- and trading partner-specific determinants of trade that are not captured by our explanatory variables. Descriptive statistics for the unbalanced and balanced data sets are presented in Table 2.

III. Estimation Results

We begin our analysis by confirming the existence of the distance puzzle. Employing annual data for the period from 1972 through 2010, we apply standard OLS regression to both an unbalanced panel data set and our balanced panel. Results are presented in Table 3. We find that the year-specific distance

Table 2. Descriptive statistics

Variable	Abbrev.	Exp. sign	Balanced	Unbalanced
Real imports	$RIMP_{ijt}$	n.a.	1708.34 (8020.47)	390.05 (3495.99)
Distance	$DIST_{ij}$	–	7559.76 (4665.91)	7431.69 (4498.07)
Real GDP (origin)	$RGDP_{it}$	+	643 955.54 (1 522 300.45)	304 220.56 (1 055 099.36)
Real GDP (destination)	$RGDP_{jt}$	+	450 929.37 (1 244 059.81)	246 260.99 (940 960.55)
Remoteness (origin)	REM_{it}	+	4851.77 (2600.82)	5439.24 (2554.48)
Remoteness (destination)	REM_{jt}	+	4052.86 (2553.57)	5167.78 (2623.89)
Border	$BORDER_{ij}$	+	0.0368 (0.1883)	0.0237 (0.1520)

Notes: SDs are in parentheses. Sample sizes: balanced, $N = 79{,}482$ and unbalanced, $N = 473{,}942$. Real imports and Real GDP values are in millions of US dollars.

[4] The dependent variable is not log-transformed when the Poisson estimation technique is employed.
[5] Internal distance, when $k = j$, is calculated as $0.4 \times \sqrt{LANDMASS_j}$ (Head and Mayer, 2000).

Table 3. Estimation results, OLS and Poisson (balanced and unbalanced samples)

	(a) OLS (balanced)		(b) Poisson (balanced)		(c) OLS (unbalanced)		(d) Poisson (unbalanced)	
$\ln RGDP_{it}$	1.1653***	(0.0433)	1.1056***	(0.0692)	1.1262***	(0.0247)	1.2689***	(0.0652)
$\ln RGDP_{jt}$	1.5459***	(0.0354)	1.4491***	(0.0528)	1.2328***	(0.0221)	1.4429***	(0.0521)
$\ln REM_{it}$	0.6322***	(0.1013)	1.5896***	(0.1418)	−0.1121	(0.0722)	1.5972***	(0.1270)
$\ln REM_{jt}$	−0.3686***	(0.1068)	0.9346***	(0.1381)	−0.6920***	(0.0803)	1.1686***	(0.1209)
$BORDER_{ij}$	0.1525***	(0.0264)	0.5858***	(0.0221)	0.6643***	(0.0237)	0.6896***	(0.0218)
$\ln DIST_{ij1972}$	−1.1018***	(0.0375)	−0.7203***	(0.0252)	−1.3189***	(0.0357)	−0.6793***	(0.0240)
$\ln DIST_{ij1973}$	−1.0957***	(0.0362)	−0.7204***	(0.0249)	−1.3671***	(0.0354)	−0.6795***	(0.0233)
$\ln DIST_{ij1974}$	−1.0782***	(0.0362)	−0.6951***	(0.0252)	−1.3897***	(0.0339)	−0.6330***	(0.0240)
$\ln DIST_{ij1975}$	−1.0621***	(0.0361)	−0.7182***	(0.0247)	−1.3685***	(0.0332)	−0.6386***	(0.0244)
$\ln DIST_{ij1976}$	−1.1032***	(0.0339)	−0.7246***	(0.0240)	−1.3920***	(0.0313)	−0.6399***	(0.0225)
$\ln DIST_{ij1977}$	−1.1062***	(0.0336)	−0.7213***	(0.0225)	−1.4453***	(0.0319)	−0.6424***	(0.0213)
$\ln DIST_{ij1978}$	−1.0984***	(0.0344)	−0.7056***	(0.0221)	−1.4714***	(0.0318)	−0.6336***	(0.0205)
$\ln DIST_{ij1979}$	−1.0679***	(0.0315)	−0.7172***	(0.0228)	−1.4494***	(0.0310)	−0.6485***	(0.0220)
$\ln DIST_{ij1980}$	−1.0797***	(0.0317)	−0.6959***	(0.0242)	−1.4992***	(0.0299)	−0.6178***	(0.0232)
$\ln DIST_{ij1981}$	−1.0735***	(0.0320)	−0.6529***	(0.0241)	−1.5092***	(0.0307)	−0.5691***	(0.0224)
$\ln DIST_{ij1982}$	−1.1236***	(0.0324)	−0.6630***	(0.0242)	−1.5049***	(0.0301)	−0.5892***	(0.0225)
$\ln DIST_{ij1983}$	−1.1261***	(0.0334)	−0.6658***	(0.0256)	−1.5164***	(0.0297)	−0.6014***	(0.0229)
$\ln DIST_{ij1984}$	−1.1238***	(0.0333)	−0.6397***	(0.0308)	−1.4506***	(0.0303)	−0.5895***	(0.0267)
$\ln DIST_{ij1985}$	−1.0962***	(0.0317)	−0.6461***	(0.0343)	−1.4675***	(0.0303)	−0.6091***	(0.0299)
$\ln DIST_{ij1986}$	−1.1101***	(0.0295)	−0.6778***	(0.0349)	−1.4519***	(0.0289)	−0.6497***	(0.0311)
$\ln DIST_{ij1987}$	−1.1548***	(0.0298)	−0.6951***	(0.0313)	−1.4431***	(0.0287)	−0.6622***	(0.0280)
$\ln DIST_{ij1988}$	−1.1026***	(0.0290)	−0.6816***	(0.0286)	−1.4466***	(0.0278)	−0.6547***	(0.0263)
$\ln DIST_{ij1989}$	−1.1034***	(0.0291)	−0.6722***	(0.0272)	−1.4677***	(0.0272)	−0.6410***	(0.0251)
$\ln DIST_{ij1990}$	−1.1435***	(0.0279)	−0.7001***	(0.0254)	−1.4461***	(0.0258)	−0.6695***	(0.0239)
$\ln DIST_{ij1991}$	−1.1345***	(0.0269)	−0.6932***	(0.0251)	−1.4248***	(0.0258)	−0.6662***	(0.0240)
$\ln DIST_{ij1992}$	−1.1470***	(0.0276)	−0.6948***	(0.0245)	−1.3361***	(0.0237)	−0.6799***	(0.0241)
$\ln DIST_{ij1993}$	−1.1267***	(0.0273)	−0.6462***	(0.0261)	−1.4222***	(0.0230)	−0.6324***	(0.0254)
$\ln DIST_{ij1994}$	−1.1233***	(0.0272)	−0.6506***	(0.0261)	−1.4313***	(0.0218)	−0.6386***	(0.0256)
$\ln DIST_{ij1995}$	−1.1199***	(0.0263)	−0.6619***	(0.0246)	−1.5145***	(0.0202)	−0.6544***	(0.0242)
$\ln DIST_{ij1996}$	−1.1093***	(0.0272)	−0.6629***	(0.0246)	−1.5054***	(0.0195)	−0.6600***	(0.0244)
$\ln DIST_{ij1997}$	−1.1073***	(0.0273)	−0.6516***	(0.0248)	−1.5301***	(0.0193)	−0.6511***	(0.0250)
$\ln DIST_{ij1998}$	−1.1392***	(0.0273)	−0.6656***	(0.0251)	−1.5298***	(0.0187)	−0.6714***	(0.0259)
$\ln DIST_{ij1999}$	−1.1335***	(0.0268)	−0.6552***	(0.0260)	−1.5486***	(0.0180)	−0.6593***	(0.0264)
$\ln DIST_{ij2000}$	−1.1363***	(0.0279)	−0.6359***	(0.0264)	−1.6020***	(0.0182)	−0.6155***	(0.0269)
$\ln DIST_{ij2001}$	−1.1384***	(0.0280)	−0.6464***	(0.0260)	−1.5935***	(0.0188)	−0.6262***	(0.0266)
$\ln DIST_{ij2002}$	−1.1650***	(0.0283)	−0.6571***	(0.0254)	−1.6015***	(0.0193)	−0.6400***	(0.0265)
$\ln DIST_{ij2003}$	−1.1777***	(0.0280)	−0.6796***	(0.0240)	−1.6359***	(0.0187)	−0.6592***	(0.0257)
$\ln DIST_{ij2004}$	−1.1767***	(0.0278)	−0.6835***	(0.0242)	−1.6621***	(0.0199)	−0.6567***	(0.0254)
$\ln DIST_{ij2005}$	−1.1625***	(0.0274)	−0.6799***	(0.0251)	−1.6761***	(0.0216)	−0.6465***	(0.0256)
$\ln DIST_{ij2006}$	−1.1394***	(0.0269)	−0.6806***	(0.0247)	−1.6596***	(0.0206)	−0.6454***	(0.0254)
$\ln DIST_{ij2007}$	−1.1711***	(0.0279)	−0.6936***	(0.0238)	−1.6749***	(0.0205)	−0.6529***	(0.0248)
$\ln DIST_{ij2008}$	−1.1590***	(0.0280)	−0.6922***	(0.0234)	−1.6908***	(0.0219)	−0.6420***	(0.0248)
$\ln DIST_{ij2009}$	−1.1622***	(0.0277)	−0.6832***	(0.0237)	−1.6860***	(0.0212)	−0.6406***	(0.0261)
$\ln DIST_{ij2010}$	−1.1714***	(0.0282)	−0.6641***	(0.0231)	−1.6495***	(0.0236)	−0.6052***	(0.0262)
Constant	−45.024***	(2.7162)	−65.836***	(3.9846)	−20.888***	(1.5694)	−64.152***	(3.3489)
N	79 482		79 482		473 942		473 942	
F-statistic	1665.09***		.		3151.32***		.	
χ^2	.		300 080***		.		766 151***	
R^2	0.8094		.		0.7200		.	
Pseudo R^2	.		0.9335		.		0.9260	

Notes: Robust SEs are in parentheses. All estimations include time, exporter and importer fixed effects. Fixed effect coefficients not reported due to space constraints.
"***"denotes significance from zero at the 1% level.

variable coefficients decrease in value over time (i.e., become more negative in value) and do so at a statistically significant rate. In all estimations, the year-specific distance coefficient values are negative and significantly different from zero. The estimated coefficients are generally consistent with the findings described in Disdier and Head (2008) and in other studies that have used panel estimation (e.g., Coe *et al.*, 2007 and Carrère *et al.*, 2013).

Looking briefly at the control variables in the specification, we see that the coefficients on the GDP variables and on the common border variables are positive, as anticipated, and that all are significantly different from zero. The signs and significance of the estimated coefficients on the economic remoteness variables vary depending on the estimation technique. We observe the *a priori* expectation of positive coefficient when the Poisson technique is employed. Results are mixed when OLS is used. The consistency of coefficient signs and the pattern of significance reported are taken as evidence of the appropriateness of the Poisson estimation technique.

Turning our focus to the series of distance coefficients, results obtained from application of OLS, presented in columns (a) and (c) of Table 3, confirm the existence of the distance puzzle. Fig. 1 depicts the time paths of the distance coefficients and the corresponding linear time trends. Regressing each set of annual distance coefficients to produce the time trends results in slope coefficients that are negative and significantly different from zero. More specifically, the slope coefficients that correspond with the application of OLS to the balanced panel and to the unbalanced panel are −0.0023 (p = 0.000) and −0.0078 (p = 0.000), respectively. These findings are taken as confirmation of the existence of the distance puzzle and are similar to the trends reported from cross-sectional estimation by Yotov (2012).

Attempting to resolve the puzzle, we follow Silva and Tenreyro (2006) and apply the Poisson regression technique to the same two data sets. Estimation of Equation 1 produces the sets of distance coefficients that are presented in columns (b) and (d) of Table 3, for the balanced and unbalanced data sets, respectively. The slope coefficient of the trend line associated with the unbalanced panel is equal −0.0002 (p = 0.535). That the slope coefficient is insignificant is taken to indicate that the application of the Poisson technique partially remedies the distance puzzle and is consistent with the findings that nonlinear estimation outperforms log-linear estimation (Silva and Tenreyro, 2006 and Coe *et al.*, 2007). The slope coefficient of the trend line associated with the balanced panel is positive and significantly different from zero: 0.001 (p = 0.003). Thus, application of the Poisson technique to a balanced panel data set provides results that are consistent with the resolution of the distance puzzle.

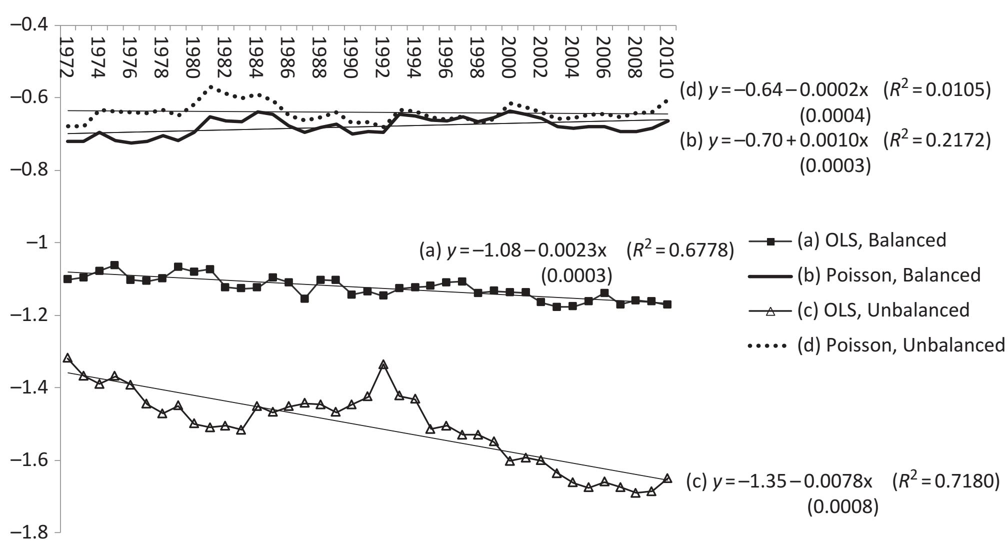


Fig. 1. Estimated distance coefficients and fitted time trends

IV. Conclusion

The distance puzzle has been an anomaly in the empirical trade literature. Using two data sets, one an unbalanced panel that includes all country pairs who engaged in trade in any year during the period 1972–2010 and the other a balanced panel that includes only those country pairs that traded in every year during the reference period, we apply both the OLS and the Poisson estimation techniques to the preferred gravity specification of Anderson and van Wincoop (2003) in a way that allows us to directly examine the evolution of estimated distance coefficients. These methods address methodology, sample selection, omitted variable bias and specification issues as explanations of the distance puzzle. OLS estimation confirms the existence of the distance puzzle in our data sets, while application of the Poisson technique remedies the puzzle. Accordingly, we propose the application of the Poisson estimation technique and the use of a balanced panel data set as a simple solution to the distance puzzle.

References

Anderson, J. E. and van Wincoop, E. (2003) Gravity with gravitas: a solution to the border puzzle, *American Economic Review*, **93**, 190–2.

Berthelon, M. and Freund, C. (2008) On the conservation of distance in international trade, *Journal of International Economics*, **75**, 310–20. doi:10.1016/j.jinteco.2007.12.005

Boulhol, H. and de Serres, A. (2010) Have developed countries escaped the curse of distance? *Journal of Economic Geography*, **10**, 113–39. doi:10.1093/jeg/lbp015

Brun, J.-F., Carrère, C., Guillaumont, P., *et al.* (2005) Has distance died? Evidence from a panel gravity model, *The World Bank Economic Review*, **19**, 99–120. doi:10.1093/wber/lhi004

Cairncross, E. (1997) *The Death of Distance: How the Communications Revolution is Changing Our Lives*. Harvard Business School Press, Boston, MA.

Carrère, C., De Melo, J. and Wilson, J. (2013) The distance puzzle and low-income countries: an update, *Journal of Economic Surveys*, **27**, 717–42.

Carrère, C. and Schiff, M. (2005) On the geography of trade: distance is alive and well, *Revue Economique*, **56**, 1249–74. doi:10.3917/reco.566.1249

Centre d'etudes prospectives et d'informations internationales (CEPII) (2014) Gravity Dataset. Available at http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=8 (accessed May 2014).

Coe, D., Subramanian, A. and Tamirisa, N. (2007) The missing globalization puzzle: evidence of the declining importance of distance, *IMF Staff Papers*, **54**, 34–58. doi:10.1057/palgrave.imfsp.9450003

Disdier, A.-C. and Head, K. (2008) The puzzling persistence of the distance effect on bilateral trade, *The Review of Economics and Statistics*, **90**, 37–48. doi:10.1162/rest.90.1.37

Felbermayr, G. J. and Kohler, W. (2006) Exploring the intensive and extensive margins of world trade, *Weltwirtschaftliches Archive (Review of World Economics)*, **142**, 642–74. doi:10.1007/s10290-006-0087-3

Head, K. and Mayer, T. (2000) Non-Europe: the magnitude and causes of market fragmentation in Europe, *Weltwirtschaftliches Archive (Review of World Economics*), **136**, 285–314.

Head, K. and Ries, J. (1998) Immigration and trade creation: econometric evidence from Canada, *The Canadian Journal of Economics*, **31**, 47–62. doi:10.2307/136376

Lin, F. and Sim, N. C. (2012) Death of distance and the distance puzzle, *Economics Letters*, **116**, 225–8. doi:10.1016/j.econlet.2012.03.004

Siliverstovs, B. and Schumacher, D. (2008) Disaggregated trade flows and the missing globalization puzzle, *CEPII Research Center*, **115**, 141–64.

Silva, J. M. C. S. and Tenreyro, S. (2006) The log of gravity, *Review of Economics and Statistics*, **88**, 641–58.

United Nations (UNCTAD) (2014) United Nations Commodity Trade Statistics Database. Available at http://comtrade.un.org/db/ (accessed January 2014).

World Bank (2014) *World Development Indicators*, Development Data Group of the World Bank's International Economics Department, Washington, DC. Available at http://databank.worldbank.org (accessed May 2014).

Yotov, Y. V. (2012) A simple solution to the distance puzzle in international trade, *Economics Letters*, **117**, 794–8. doi:10.1016/j.econlet.2012.08.032