Economics Bulletin

Volume 33, Issue 3

Gravity with google maps: the border puzzle revisited

Sebastian Benz
Ifo Institute - Leibniz Institute for Economic Research at the University of Munich

Abstract

I calculate road travel times between the capitals of US states and Canadian provinces with Google Maps. With this measure of trade cost I estimate the US-Canada border effect for aggregate and industry-level trade flows in line with the method introduced by Anderson and van Wincoop (2003), as well as using linar fixed effects and Poisson estimation. I find a high degree of heterogeneity in the resulting coefficients.

Citation: Sebastian Benz, (2013) "Gravity with google maps: the border puzzle revisited", *Economics Bulletin*, Vol. 33 No. 3 pp. 2414-2421.
Contact: Sebastian Benz - benz@ifo.de.
Submitted: August 12, 2013. **Published:** September 17, 2013.

1. Introduction

Air distance between country capitals or major economic centers is a commonly used proxy for international trade costs in the academic literature. However, recent research indicates that this proxy might be systematically biased. Djankov et al. (2010) show that transportation to a harbour and waiting time in the ports has an additional negative impact on trade flows. Hummels and Schaur (2013) analyze the fact that exporters circumvent long shipping time by costly air transport for some goods. This sheds some doubt on the validity of the argument that air distance is a sufficiently good proxy for trade costs and that trade costs are equally important for all classes of products.

In this paper I apply these ideas to a well-known setup: trade between US states and Canadian provinces. Using great circle distance for intra-continental trade flows is based on the presumption of air travel. However, according to the US Bureau of Transportation Statistics, road transport accounts for 53.1 percent of the US-Canada trade, followed by rail at 16.2 percent, pipelines at 13.9 percent, vessels at 6.1 percent, and air at 4.4 percent. Consequently, I construct a novel dataset of exact travel time on the road between US states and Canadian provinces. This allows me to analyze whether Anderson and van Wincoop's (2003) solution to McCallum's (1995) border puzzle is robust to such a more precise specification of trade costs. Moreover, I use newer trade data from 2007 to see how coefficients change over time. Finally, I use disaggregated data to check for sectoral heterogeneity in the border effect and to analyze whether travel time or air distance is a more realistic measure in some sectors than in others.

2. Data

Data on intra-US trade comes from the Commodity Flow Survey, data on intra-Canada trade is from Statistics Canada, and data from Canada-US trade is from Industry Canada. All three data sources use different sectoral classification. To make the sectoral trade flows comparable I construct eleven broad categories. These categories are "Agriculture & Food", "Mining, Utilities, and Coal", "Textiles & Leather", "Paper & Wood Product", "Print", "Chemicals & Plastics", "Nonmetallic Minerals", "Metals", "Machinery", "Transportation", and "Furniture". Every reference to sectoral trade flows below means these eleven categories.

The data on road travel time are based on Google Maps. A method to obtain this information with Stata is described in Miles and Ozimek (2011).[1] To give a first impression of the data, I report mean values of air distance and travel time in Table 1. It can easily be seen, that the speed of travel is highest for intra-US pairs, followed by cross-border pairs, and is lowest for intra-Canada pairs. This suggests that road infrastructure in Canada is not as good as in the United States. Thus, the proxy for trade cost used in Anderson and van Wincoop (2003) is systematically biased which may well have influenced their results.

[1]For travel time within a region I follow Anderson and van Wincoop (2003) and others in using a quarter of the travel time to the nearest neighbour.

Table 1. Descriptive Statistics

	US-US	CAN-CAN	US-CAN CAN-US
Air distance (in kilometres)	1717	1948	2278
Google travel time (in minutes)	1295	1857	1878
Ratio (in km/min)	1.33	1.05	1.21

Notes: The table reports mean values of air distance and Google travel time and the ratio thereof, conditional on being a US pair, Canada pair or international pair for all US states except Hawaii and all Canadian provinces

3. Results

In this section I report results from gravity estimation of trade between US states and Canadian provinces. More precisely, the variable on the left-hand side is bilateral trade divided by the product of the two regions' GDP, or the logarithm thereof, depending on the specification. Table 2 starts with data from the year 1993 of trade between 30 states and 10 provinces, the same as used by Anderson and van Wincoop (2003). The 30 US states account for roughly 85 per cent of US GDP. Column (1) is an exact replication of their nonlinear estimation strategy. Column (2) uses the same strategy but great circle distance is replaced by travel time data. While the coefficient on the trade cost parameter is very similar, the estimated coefficients for the border parameter differ from each other.

Columns (3) to (6) show results from a linear fixed effects estimation. Importantly, this allows me to introduce distance and travel time jointly as right-hand side variables. When doing so, the distance coefficient is no longer significantly different from zero. Only when dropping the travel time variable it becomes negative and significant. Moreover, I introduce a contiguity dummy as additional feature of trade cost which is always significantly positive. Replacing the border dummy with two dummies, one for intra-US trade and one for intra-Canada trade reveals that intra-Canada trade flows are significantly larger than cross-border trade flows, whiles the estimated coefficient for intra-US trade is not significantly different from zero.

Columns (7) to (9) display results from a Poisson estimation, following Santos Silva and Tenreyro (2006). Coefficients are qualitatively identical to those from the linear regression. Again, joint use of the distance variable and travel time variable renders the coefficient for distance insignificant. The only difference compared to results from the linear model shows up in the coefficient of the dummy for intra-US trade which is now significantly negative, meaning that intra-US trade flows are significantly smaller than cross-border trade flows. The negative sign is surprising but can probably be explained by the fact that intra-US trade, intra-Canada trade, and cross-border trade are from three different data sources. US data are from the Commodity Flow Survey and then adjusted downwards as outlined in the Appendix of Anderson and van Wincoop (2003). This correction is probably too strong, so that intra-US flows appear to be smaller than

Table 2. 1993 - 30 States and 10 Provinces

	AvW NLS Estimation		Fixed Effects Estimation				Poisson Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln Distance	-0.79***		0.05	0.05	-1.14***		-0.28	-1.59***	
	(0.03)		(0.42)	(0.42)	(0.05)		(0.72)	(0.09)	
ln Traveltime		-0.80***	-1.28***	-1.28***		-1.23***	-1.31*		-1.59***
		(0.03)	(0.44)	(0.44)		(0.05)	(0.70)		(0.09)
Contiguity			0.35***	0.35***	0.44***	0.36***	0.97***	0.92***	0.98***
			(0.09)	(0.09)	(0.10)	(0.09)	(0.25)	(0.24)	(0.24)
Border	-1.65***	-1.75***	-1.97***						
	(0.08)	(0.07)	(0.07)						
US Dummy				-0.33	-0.17	-0.32	-1.81***	-1.72***	-1.83***
				(0.21)	(0.22)	(0.21)	(0.39)	(0.38)	(0.39)
CAN Dummy				4.27***	4.08***	4.27***	3.22***	3.10***	3.24***
				(0.23)	(0.24)	(0.21)	(0.38)	(0.36)	(0.37)
Importer FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Exporter FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Observations	1559	1559	1559	1559	1559	1559	1600	1600	1600
Adjusted R^2	n.a.	n.a.	0.79	0.79	0.79	0.79	0.96	0.96	0.96

Notes: Constant and fixed effects not reported. Robust standard errors in parenthesis. "AvW NLS" is Anderson and van Wincoop's (2003) Nonlinear Least Squares Method. "n.a." means not applicable. ***Significant at 1 percent level, **Significant at 5 percent level, *Significant at 10 percent level.

they are.[2]

The coefficients in Table 3 are estimated with newer data from the year 2007, but again restricting the sample to the 30 states and 10 provinces from Anderson and van Wincoop (2003). All results are qualitatively identical to those with data from 1993. Therefore, I will use the more recent data from 2007 in all following estimations.[3]

Table 4 presents results from an industry-level estimation. I aggregate trade data, which is partly reported on product-level and partly on sector-level, to eleven industries, significantly increasing the number of observations. Exporter and importer fixed effects interacted with industry fixed effects are now used, as proposed by Chen (2004). The intra-US dummy in the linear regressions, which had been insignificant when using aggregate trade flows, is now positive and significant. Moreover, coefficients on distance and travel time are slightly larger in absolute terms compared to the estimation with aggregate data. In the idustry-level Poisson regression contiguity has no significant impact, while coefficients on distance and travel time become slightly larger in absolute terms than in the aggregate estimation and coefficients on the US dummy become absolutely smaller and are hardly significant anymore.

[2]Shipments overestimate internal trade flows because they include shipments of exports from the location of production to a harbour or shipment of imports from a harbour to the location of final consumption. Moreover, some shipments may be counted twice if they are transported from a manufacturing plant to a warehouse and from there to a retailer. Therefore, shipments are scaled down by the ratio of total domestic merchandise trade to total shipments. In 1993 this ratio is 3025/5846 and similar in all other years.

[3]I did the same estimation with data from 1997 and 2002, the other years of the Commodity Flow Survey in the United States. Results are basically the same.

Table 3. 2007 - 30 States and 10 Provinces

	AvW NLS Estimation		Fixed Effects Estimation				Poisson Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln Distance	-0.82***		0.08	0.08	-1.17***		-0.88	-1.53***	
	(0.04)		(0.46)	(0.46)	(0.05)		(0.83)	(0.09)	
ln Traveltime		-0.85***	-1.17***	-1.17***		-1.25***	-0.65		-1.52***
		(0.04)	(0.48)	(0.48)		(0.05)	(0.82)		(0.09)
Contiguity			0.29***	0.29***	0.37***	0.28***	0.78***	0.75***	0.84***
			(0.10)	(0.10)	(0.10)	(0.10)	(0.28)	(0.27)	(0.26)
Border	-1.65***	-1.64***	-1.95***						
	(0.09)	(0.09)	(0.08)						
US Dummy				-0.05	-0.10	-0.06	-1.21***	-1.17***	-1.25***
				(0.21)	(0.21)	(0.21)	(0.34)	(0.32)	(0.34)
CAN Dummy				3.94***	3.77***	3.96***	2.32***	2.27***	2.40***
				(0.25)	(0.21)	(0.24)	(0.31)	(0.30)	(0.31)
Importer FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Exporter FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Observations	1551	1551	1551	1551	1551	1551	1600	1600	1600
Adjusted R^2	n.a.	n.a.	0.77	0.77	0.77	0.77	0.32	0.32	0.32

Notes: Constant and fixed effects not reported. Robust standard errors in parenthesis. "AvW NLS" is Anderson and van Wincoop's (2003) Nonlinear Least Squares Method. "n.a." means not applicable. ***Significant at 1 percent level, **Significant at 5 percent level, *Significant at 10 percent level.

Table 4. 2007 - 30 States and 10 Provinces - Industry-level

	Fixed Effects Estimation				Poisson Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ln Distance	-0.02	-0.02	-1.27***		-0.81	-1.31***	
	(0.21)	(0.21)	(0.02)		(0.59)	(0.05)	
ln Traveltime	-1.33***	-1.33***		-1.36***	-0.51		-1.33***
	(0.22)	(0.22)		(0.03)	(0.60)		(0.05)
Contiguity	0.29***	0.29***	0.38***	0.29***	0.20	0.20	0.23
	(0.05)	(0.05)	(0.05)	(0.05)	(0.16)	(0.16)	(0.15)
Border	-2.37***						
	(0.04)						
US Dummy		0.67**	0.83***	0.67**	-0.39*	-0.35	-0.45*
		(0.31)	(0.31)	(0.31)	(0.23)	(0.23)	(0.23)
CAN Dummy		4.06***	3.87***	4.07***	1.79***	1.76***	1.86***
		(0.32)	(0.32)	(0.31)	(0.22)	(0.22)	(0.22)
Importer x Industry FE	YES	YES	YES	YES	YES	YES	YES
Exporter x Industry FE	YES	YES	YES	YES	YES	YES	YES
Observations	13523	13523	13523	13523	17600	17600	17600
Adjusted R^2	0.76	0.76	0.76	0.76	0.33	0.33	0.33

Notes: Constant and fixed effects not reported. Robust standard errors in parenthesis. "n.a." means not applicable. ***Significant at 1 percent level, **Significant at 5 percent level, *Significant at 10 percent level.

To dig deeper into the sectoral structure of North American trade I split my sample of trade flows into the above mentioned eleven industries. In the following I report estimation results for some industries only. Namely for those which yield results that differ substantially from the estimation with aggregate trade flows. Notably, these sectors are rather small. This is probably no surprise considering that larger sectors have a larger impact on the pattern of aggregate trade flows.

Differences in the results for different sectors may come from a variety of reasons. For example, trade cost coefficients may be significantly larger in a gravity estimation of perishable consumption goods. Border coefficients may be large in industries where we can expect a home bias in comsumption due to cultural traditions. Moreover, coefficients of air distance and traveltime may be very different, depending on whether air transport is important in an industry or whether most transport is on the road.

Table 5 starts with an analysis of trade flows of basic raw materials and utilities. It comprises NAICS industries 21 (Mining), 22 (Utilities), and 324 (Petroleum and Coal Products Manufacturing) and is the sector with the smallest border effect in the nonlinear estimation. As becomes evident from Column (4) onwards, the coefficient of the border dummy is only driven by higher trade flows within Canada than across the border. Surprisingly, contiguity has a negative and significant coefficient when accounting for zero trade flows in the Poisson estimation. The reason might be that extraction of raw materials is concentrated in a small region of neighbouring states, while processing is located somewhere else, so that these neighbouring states have no need to trade with each other. The high concentration of extraction and processing of raw materials is also indicated by the small number of positive trade flows. Moreover, this is the only sector for which the estimated coefficient on travel time is smaller than the one on air distance in the nonlinear specification. An explanation might be that truck transport is less important for this sector, but much of the transport takes place via rail or pipeline.

Table 6 reports results for the textiles and leather industry. This industry accounts for the lowest coefficients on both distance and travel time in the nonlinear specification, while the coefficient for the border dummy is extremely high (second highest of all industries). The strong border effect implies that not only intra-Canada trade but also intra-US trade is significantly higher than cross-border trade. Contrary to trade in raw materials, the contiguity dummy is always positive and significant. Distance and travel time are only significant when using one of the variables.

Table 7 shows estimation results for print products, the sector with the biggest border effect. An explanation might be that preferences have substantial bias for domestic products so that only few cross-border trade flows are observed. All other coefficients are similar to those in the textiles and leather industry. The only difference is that the contiguity dummy is not significant in the linear specification.

4. Conclusion

Using Google Maps I build a dataset of bilateral travel time between the capitals of US states and Canadian provinces which allows me to show that road travel takes longer relative to air distance within Canada and across the border than within the US. With these data I perform gravity estimation of North American trade flows, using the non-linear least square method from Anderson and van Wincoop (2003), as well as linear

Table 5. 2007 - 30 States and 10 Provinces - Mining, Utilities, and Coal

	AvW NLS Estimation		Fixed Effects Estimation				Poisson Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln Distance	-0.61***		0.21	0.21	-1.98***		-2.26	-1.87***	
	(0.19)		(0.85)	(0.85)	(0.10)		(1.76)	(0.17)	
ln Traveltime		-0.54***	-2.34***	-2.34***		-2.11***	0.39		-1.87***
		(0.19)	(0.89)	(0.89)		(0.11)	(1.81)		(0.17)
Contiguity			0.81***	0.82***	0.95***	0.83***	-0.89*	-0.88*	-0.76*
			(0.21)	(0.22)	(0.21)	(0.22)	(0.50)	(0.49)	(0.46)
Border	-1.69***	-1.87***	-1.21***						
	(0.09)	(0.10)	(0.17)						
US Dummy				-0.82	-0.50	-0.79	0.18	0.15	0.06
				(0.51)	(0.49)	(0.49)	(0.53)	(0.49)	(0.53)
CAN Dummy				3.23***	2.88***	3.20***	1.10**	1.12**	1.23**
				(0.55)	(0.52)	(0.52)	(0.51)	(0.52)	(0.56)
Importer FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Exporter FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Observations	994	994	994	994	994	994	1600	1600	1600
Adjusted R^2	n.a.	n.a.	0.69	0.69	0.69	0.69	0.30	0.30	0.30

Notes: Constant and fixed effects not reported. Robust standard errors in parenthesis. ”AvW NLS” is Anderson and van Wincoop’s (2003) Nonlinear Least Squares Method. ”n.a.” means not applicable. ***Significant at 1 percent level, **Significant at 5 percent level, *Significant at 10 percent level.

Table 6. 2007 - 30 States and 10 Provinces - Textiles and Leather

	AvW NLS Estimation		Fixed Effects Estimation				Poisson Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln Distance	-0.40***		0.25	0.25	-0.81***		-0.12	-0.74***	
	(0.07)		(0.71)	(0.71)	(0.08)		(0.60)	(0.05)	
ln Traveltime		-0.41***	-1.14	-1.14		-0.88***	-0.63		-0.76***
		(0.07)	(0.75)	(0.75)		(0.08)	(0.63)		(0.05)
Contiguity			0.35**	0.35**	0.42***	0.36***	0.72***	0.74***	0.73***
			(0.18)	(0.18)	(0.17)	(0.17)	(0.16)	(0.16)	(0.16)
Border	-4.05***	-4.06***	-3.20***						
	(0.13)	(0.13)	(0.14)						
US Dummy				2.85***	2.99***	2.88***	1.10***	1.16***	1.09***
				(0.30)	(0.30)	(0.30)	(0.43)	(0.42)	(0.43)
CAN Dummy				3.56***	3.40***	3.52***	2.04***	1.95**	2.06***
				(0.37)	(0.36)	(0.36)	(0.51)	(0.47)	(0.48)
Importer FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Exporter FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Observations	1174	1174	1174	1174	1174	1174	1600	1600	1600
Adjusted R^2	n.a.	n.a.	0.79	0.79	0.79	0.79	0.13	0.13	0.13

Notes: Constant and fixed effects not reported. Robust standard errors in parenthesis. ”AvW NLS” is Anderson and van Wincoop’s (2003) Nonlinear Least Squares Method. ”n.a.” means not applicable. ***Significant at 1 percent level, **Significant at 5 percent level, *Significant at 10 percent level.

Table 7. 2007 - 30 States and 10 Provinces - Print Products

	AvW NLS Estimation		Fixed Effects Estimation				Poisson Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ln Distance	-0.76***		0.15	0.15	-1.24***		-0.22	-1.19***	
	(0.05)		(0.66)	(0.66)	(0.08)		(0.67)	(0.06)	
ln Traveltime		-0.81***	-1.50**	-1.50		-1.34***	-0.97		-1.20***
		(0.05)	(0.69)	(0.69)		(0.08)	(0.66)		(0.06)
Contiguity			-0.11	-0.11	-0.01	-0.10	0.69***	0.66***	0.70***
			(0.15)	(0.15)	(0.15)	(0.15)	(0.21)	(0.20)	(0.20)
Border	-4.08***	-4.07***	-4.08***						
	(0.12)	(0.12)	(0.13)						
US Dummy				2.97***	3.16***	2.99***	2.06***	2.08***	2.05***
				(0.29)	(0.30)	(0.29)	(0.37)	(0.36)	(0.38)
CAN Dummy				5.20***	4.99***	5.17***	3.73***	3.70***	3.74***
				(0.35)	(0.34)	(0.34)	(0.39)	(0.37)	(0.40)
Importer FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Exporter FE	n.a.	n.a.	YES	YES	YES	YES	YES	YES	YES
Observations	1085	1085	1085	1085	1085	1085	1600	1600	1600
Adjusted R^2	n.a.	n.a.	0.82	0.82	0.82	0.82	0.29	0.29	0.29

Notes: Constant and fixed effects not reported. Robust standard errors in parenthesis. "AvW NLS" is Anderson and van Wincoop's (2003) Nonlinear Least Squares Method. "n.a." means not applicable. ***Significant at 1 percent level, **Significant at 5 percent level, *Significant at 10 percent level.

fixed effects, and Poisson estimation. While coefficients are relatively constant over time from 1993 to 2007, using industry-level trade flows allows me to reveal a high degree of heterogeneity across sectors.

References

Anderson, James E. and Eric van Wincoop (2003) "Gravity with Gravitas: A Solution to the Border Puzzle" *American Economic Review* 93(1), 170-192

Chen, Natalie (2004) "Intra-National versus International Trade in the European Union: Why Do National Borders Matter?" *Journal of International Economics* 63(1), 93-118

Djankov, Simeon, Caroline Freund and Cong S. Pham (2010) "Trading on Time" *The Review of Economics and Statistics* 92(1), 166-173

Hummels, David and Georg Schaur (2013). "Time as a Trade Barrier" *American Economic Review,* forthcoming

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

Miles, Daniel and Adam Ozimek (2011) "Stata utilities for geocoding and generating travel time and travel distance information" *The Stata Journal* 11(1), 106-119

Santos Silva, Joao and Silvana Tenreyro (2006) "The Log of Gravity" *The Review of Economics and Statistics* 88(4), 641-658