

## Intra-national Home Bias: Some Explanations

Russell Hillberry

US International Trade Commission

and Purdue University

David Hummels

Purdue University

March 2002

**Abstract:** Wolf (2000) demonstrates that trade within the U.S. appears substantially impeded by state borders. We revisit this finding with improved data. We show that much intra-national home bias can be explained by wholesaling activity. Shipments by wholesalers are much more localized within states than shipments from manufacturing establishments. Controlling for relative prices and the use of actual, rather than imputed, shipment distances also reduces home bias estimates.

**JEL Classification:** F15 Economic Integration

**Acknowledgements:** We wish to thank, without implication, Dan Kovenock and David Meyer for helpful comments; Cathy Buffington and Arnold Rezneck at the Center for Economic Analysis for data assistance; and Purdue CIBER for financial support.

The views expressed here are those of the authors, and not of the U.S. International Trade Commission or any of its individual Commissioners.

Contact information: David Hummels, Dept of Economics, Krannert School of Management, Purdue University, West Lafayette IN 47907-1310; ph 765 494 4495; fax: 765 496 1778; hummelsd@purdue.edu.

## I. Introduction

Why do political boundaries shape the geographic pattern of trade? This question has attracted considerable attention since McCallum's (1995) finding that Canadian interprovincial trade was twenty-two times larger than province-state trade. While the magnitude of the McCallum estimate is surprising<sup>1</sup>, one can certainly identify reasons why international borders impede trade. Presumably, national borders proxy for a wide range of trading frictions, including tariffs and non-tariff measures imposed intentionally by national governments, as well as costs associated with customs clearance and currency exchange that inevitably arise when shipping goods across differing national jurisdictions.

Such frictions are notably absent in trade between U.S. states, which are constitutionally enjoined from impeding interstate commerce. Yet, it still appears that state borders inhibit trade flows. Using the public sample of the 1993 US Commodity Flow Survey, Wolf (2000) estimated that intra-state trade was more than 4 times larger than trade between states. This leaves us with two questions. One, why do arbitrarily drawn political boundaries like U.S. state borders appear to pose a barrier to trade? Two, is there economic significance to these borders not previously appreciated?

An answer may be found by noting that Wolf's data included shipments originating in both manufacturing and wholesale establishments.<sup>2</sup> One can broadly think of manufacturers and wholesalers as a kind of hub and spoke arrangement. Goods are manufactured in the hub and dispersed, sometimes at great distances, to a number of wholesaling spokes spread throughout the country. The wholesaling spokes then

---

<sup>1</sup> The magnitudes are both surprising and in question. See Anderson and vanWincoop (2001).

<sup>2</sup> This implies that any particular good may be counted twice in the shipments data; once when shipped from the manufacturer, and once when shipped from the wholesaler.

distribute, over very short distances, to retailers. As a result, lumping wholesaling and manufacturing shipments together may provide a misleading picture of spatial frictions.

The reason that manufacturers would choose to employ this hub and spoke system is itself quite interesting and informative about spatial frictions. Put another way, one might ask: if manufacturers find it easy to ship over long distances, why don't wholesalers? And the answer may be that the kinds of geographic frictions wholesalers face are quite different from those faced by manufacturers.

Several possibilities suggest themselves. Wholesaling may be employed in order to efficiently manage inventories and respond rapidly to demand fluctuations. Hub and spokes system may also be used to exploit the relative efficiency of long and short haul transportation modes. That is, large trucks or rail are used for movements between manufacturers and wholesalers, while smaller trucks transport between wholesalers and retailers. In both cases, wholesale shipments are excessively local relative to manufacturing shipments because the cost of some geographic frictions related to distribution rise sharply in stages of the value-chain immediately prior to consumption. Returning to Wolf's result, it is not that state borders matter, per se, but that borders proxy for very short shipment lengths.

It may also be that the political boundaries are of direct interest due to contractual stipulations binding on wholesale shippers. Manufacturers are legally allowed to segment markets by designating explicit geographic boundaries that their wholesalers are not allowed to cross. Such segmentation may be an effective way to engage in resale price maintenance, and state boundaries are an obvious way to divide territory.

All three explanations can be thought of as kinds of geographic frictions, but they are quite different than straightforward transportation cost frictions typically supposed. The question then becomes whether the responsiveness of shipments to geographic frictions depends as much on the nature of the shipper as on the good being shipped. If yes, then this informs us about the nature, as well as the size, of the frictions in question.

To answer these questions we employ a private use sample of the 1997 US Commodity Flow Survey.<sup>3</sup> These data provide two significant benefits relative to the 1993 public use sample used by Wolf. First, we can separate wholesale shipments from shipments by manufacturing establishments in order to distinguish their respective spatial characteristics. Second, we observe actual distances shipped rather than having to impute distances from the physical distribution of a state's population. We find that actual shipment distances within states are much shorter than Wolf's measure of distance suggests, with the effect quite pronounced within states.

We incorporate information on shipment distances and wholesale versus manufacturing establishments in our estimates, and also control for goods' prices in a manner consistent with the Anderson and van Wincoop (2001) critique of the gravity literature. Our results suggest state border effects are still significant, but roughly a third as large as Wolf estimates.

## II. Estimation Approach and Data

We use a gravity equation to estimate the volume of shipments between any state pair  $ij$ , including within state trade. This equation is typically motivated by a model in

---

<sup>3</sup> Hillberry and Hummels (2002) employ this data to link intermediate goods trade with industry agglomeration.

which goods are differentiated by origin and consumers uniquely value each variety.

With CES preferences the value of bilateral trade  $M$ , between origin  $i$  and destination  $j$ , is given by:

$$(1) \quad M_{ij} = Y_i Y_j \left( \frac{P_i}{P_j} \right)^{1-\sigma} t_{ij}^{-\sigma}$$

where  $Y$  denotes income,  $p_i / P_j$  the price of output in region  $i$  relative to the consumer price index in region  $j$ ,  $t_{ij}$  the ad valorem iceberg cost of trade between regions, and  $\sigma$  the elasticity of substitution among varieties of output. Wolf estimates a reduced form of (1) by taking logs and measuring the trade friction ( $t_{ij}$ ) using distance shipped and a dummy variable, *Ownstate*, for whether the flow remains within the originating state ( $i=j$ ). Relative prices are ignored.

$$(2) \quad \ln M_{ij} = a_0 + \beta_1 \ln Y_i + \beta_2 \ln Y_j + \beta_3 \ln DIST_{ij} + \beta_4 OWNSTATE_{ij} + e_{ij}$$

In order to make comparisons directly with Wolf, we use (2) as our baseline regression. Our contributions come primarily in providing significantly improved data on shipment values, and on distance shipped, as detailed below. Anderson and van Wincoop (2001) show that the specification in equation (2) leads to biased estimation because, in equilibrium, the omitted price index term from equation (1) is related to the size and distribution of trade barriers. We incorporate their critique in a parsimonious fashion by including vectors of importing and exporting state fixed effects. This also controls for idiosyncrasies outside of the model such as variation in the share of state output devoted to nontraded services.

The primary data source we use is the raw data file from the 1997 U.S. Commodity Flow Survey (CFS). The CFS is collected every five years by the U.S. Census Bureau, which chooses a stratified random sample of shipments from U.S. mining, manufacturing, and wholesale establishments.<sup>4</sup> Reported shipment characteristics include the shipment's weight, value and commodity classification, an establishment identifier, the shipper's (SIC) industrial classification, the zip code of the shipment's origin and destination and the actual shipping distance between them.

These are the best available data documenting sub-national trade because the data are drawn from stratified random samples of actual shipments. This is in sharp contrast with the Statistics Canada data, which are imputed from multiple data sources. The private use sample provides two improvements relative to the publicly available CFS data used by Wolf (2000) and Anderson and van Wincoop (2001): the ability to include/exclude wholesale shipments and the ability to measure actual shipment distances.

Wholesale shipments are important for reasons addressed above. Estimates of the Ownstate coefficient in (2) can be highly sensitive to the measure of internal state distances because these distances set the baseline against which actual Ownstate trade flows are compared. Wolf relies on a simple measure of internal distance ( $D_{ii}$ ):

$$(3) \quad D_{ii} = 2D_{i,12} \left[ 1 - \frac{Pop_{i,1}}{Pop_{i,1} + Pop_{i,2}} \right]$$

where  $Pop_{i,1}$  and  $Pop_{i,2}$  are the populations of the first and second largest cities in  $i$ , respectively, and  $D_{i,12}$  is the distance between the two largest cities in  $i$ .

---

<sup>4</sup> The procedure samples a set of establishments, then randomly samples shipments for the chosen establishments.

Several authors construct more sophisticated distance measures by linking them directly to a model of spatial activity which treats output as proportional to population and fixed in space.<sup>5</sup> What Wolf and subsequent authors have in common is that their measures overstate actual distances if firms move to be near to sources of idiosyncratic demand. For example, suppose a food processing firm is the primary customer for a particular kind of glass jar. If the glass jar manufacturer moves proximate to its customer within the same state the actual shipping distance of the jars may be minute. Yet, Wolf's measure applies a statewide distance as the measure of frictions applied to this transaction. The problem is most acute for internal state distances, but also applies to distances between adjacent states.

We sidestep the debate on theoretical constructions of distance by using shipment level data on actual distance shipped taken from the Commodity Flow Survey.<sup>6</sup> This draws on Department of Transportation impedance calculations on actual transport miles, that is, mileage that shipments must have traveled given the system of highways and rail lines connecting any two points. This distance can be much larger than straight line miles when highways and rail lines do not permit direct transit. It can also be much smaller than equation (3) if the location of output responds to trade costs.<sup>7</sup>

To see the difference, we compare Wolf's measure of distance to the shipment level data in the 1997 CFS. We regress actual distances for each shipment  $s$  on the distance measure used by Wolf, along with an own state dummy variable, an adjacent state dummy variable and a commodity fixed effect

---

<sup>5</sup> Nitsch (2000) and Helliwell and Verdier (2001) provide the most exhaustive measures of internal distance.

<sup>6</sup> The calculations apply the distance between zip code of origin and zip code of destination. These are still central place distances, but they are applied within a very fine grid.

<sup>7</sup> Direct evidence on this point is provided in Hillberry and Hummels (2001).

$$(4) \ln ACTDIST_{ij}^s = \beta_1 \ln WOLFDIST_{ij} + \beta_2 OWNSTATE_{ij} + \beta_3 ADJ_{ij} + a_k + e_{ij}$$

Table 1. Actual versus imputed distances

	OLS		Fixed Effects	
Wolf distances	0.939 (0.001)	0.821 (0.001)	0.931 (0.001)	0.816 (0.001)
Ownstate		-0.498 (0.004)		-0.488 (0.004)
Adjacent		-0.404 (0.002)		-0.395 (0.002)
Observations	303086	303086	303086	303086
Adj R <sup>2</sup>	0.865	0.878	0.869	0.881

Dependent variable is logged shipment distance reported in CFS. Standard errors in parentheses. All variables are significant at the 1% level.

We find that the Wolf distances slightly overstate actual distances for all pairs, but dramatically overstate distance within own-state and between adjacent state pairs. Actual shipments in-state are half as large as the Wolf measure suggests, while adjacent state shipments are 40 percent smaller. These differences affect our estimates of home bias within states.

### III. Results

We estimate equation (2) as follows. First, we aggregate shipments over commodity categories to yield total bilateral trade flows between states. We regress these shipment values on output in each state, distance and an own-state dummy variable. The columns in Table 2 reflect differences in the dataset employed. Column 1 reports Wolf's (2000) results from the 1993 CFS public use sample. The remaining columns use the 1997 data. Columns 2-4 use Wolf's distance measure. Columns 5-7 use actual shipment distances from the CFS. The other adjustments are excluding wholesale shipments, and

using vectors of origin-state and destination-state fixed effects to control for output levels, relative prices, and other state-level idiosyncracies outside of the model.

Table 2. Aggregate domestic shipments

	1993 CFS (Wolf 2000)	Basic regression	Excluding wholesale shipments	Using origin and destination fixed effects	Basic regression	Excluding wholesale shipments	Using origin and destination fixed effects
$y_i$	1.02	1.04 (0.02)	1.01 (0.02)		1.04 (0.02)	1.01 (0.02)	
$y_j$	0.98	0.98 (0.02)	1.03 (0.02)		0.99 (0.02)	1.05 (0.02)	
$dist_{ij}$ (Wolf measure)	-1.00	-0.96 (0.02)	-0.88 (0.02)	-0.95 (0.02)			
$dist_{ij}$ (observed)					-1.06 (0.02)	-1.00 (0.02)	-1.05 (0.02)
Ownstate	1.48	1.13 (0.14)	0.74 (0.15)	0.53 (0.11)	0.99 (0.13)	0.61 (0.14)	0.44 (0.10)
Observations	2137	2304	2304	2304	2304	2304	2304
Adj $R^2$	0.836	0.825	0.812	0.893	0.848	0.832	0.911

Dependent variable is the logged value of aggregate shipments from state  $i$  to  $j$ . Standard errors in parentheses. All variables significant at the 1% level.

We focus primarily on the Ownstate dummy variable. There are four main findings. First, the estimated Ownstate coefficient is lower in 1997 than in 1993.<sup>8</sup> Second, using actual distances shipped reduces the Ownstate coefficient in all specifications. Third, excluding wholesale shipments considerably reduces the effect of the Ownstate dummy, as well as slightly lowering the distance elasticity. This suggests that wholesale shipments are highly localized. Fourth, using origin and destination fixed effects to control for prices, among other things, reduces the Ownstate coefficient further.

<sup>8</sup> This may simply reflect differences in sample coverage between the private and public use data. Also, the 1997 data includes all state pairs within the continental US, while the 1993 data exclude some pairs.

Our final specification, which incorporates all three adjustments, shows the sizable border coefficient estimated by Wolf to be reduced to one-third its former size.<sup>9</sup>

To check robustness, we also estimated a version of equation (2) separately for each 4 digit SIC category using a dependent variable that first included, then excluded, wholesale shipments. We employed origin-state and destination-state fixed effects, distance shipped, and an own-state dummy as dependent variables. Ownstate coefficients were, on average, twice as large for the shipment regressions that included wholesale shipments. Distance coefficients were also systematically larger for wholesale shipments.

#### IV. Conclusions

By documenting the effect of state borders on internal U.S. trade patterns, Wolf (2000) suggested an important puzzle for the economic geography literature. Home bias in trade among U.S. states is surprising, given the relative absence of obvious trading frictions. We provide three possible explanations for measured home bias in state commodity flows. Taken together, we find home bias one-third as large as Wolf.

First, average distance calculations based on the geographic distribution of population overstate actual distances that shipments travel. This is likely explained by firms locate near to idiosyncratic demand for their output. Second, wholesale shipments are highly localized, and more sensitive to state borders. This is consistent with the view that wholesale shipments serve a substantially different economic function (e.g., inventory holding) than shipments by manufacturers. Wholesale shipments may also be

---

<sup>9</sup> The coefficient implies that the ratio of actual to predicted Ownstate trade is 1.55. Manufacturers' Ownstate shipments are 55% higher than is predicted by the model. Wolf's coefficient estimates imply a ratio of 4.39.

directly affected by state borders if distribution contracts assign exclusive state territories to wholesalers. Finally, controlling for variation in relative prices, as suggested by Anderson and van Wincoop (2001), reduce measured home bias.

References:

Anderson, James and Eric van Wincoop 2001, "Gravity with Gravitas: A Solution to the Border Puzzle," NBER Working Paper 8079.

Helliwell, John F. and Genevieve Verdier 2001, "Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada," *Canadian Journal of Economics* 34, pp. 1024-1041.

Hillberry, Russell and David Hummels 2002, "Explaining Home Bias in Consumption: The Role of Intermediate Input Trade," Purdue University *mimeo*.

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

Nitsch, Volker 2000, "National Borders and International Trade: Evidence from the European Union," *Canadian Journal of Economics* 33, pp. 1091-1095.

Wolf, Holger C. 2000, "Intra-national Home Bias in Trade," *The Review of Economics and Statistics* 82(4), pp. 555-563.