

Gravity with Gravitas: A Solution to the Border Puzzle

Author(s): James E. Anderson and Eric van Wincoop

Source: *The American Economic Review*, Vol. 93, No. 1 (Mar., 2003), pp. 170-192

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/3132167>

Accessed: 07-09-2016 16:02 UTC

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at  
<http://about.jstor.org/terms>



*American Economic Association* is collaborating with JSTOR to digitize, preserve and extend access to *The American Economic Review*

# Gravity with Gravititas: A Solution to the Border Puzzle

By JAMES E. ANDERSON AND ERIC VAN WINCOOP\*

*Gravity equations have been widely used to infer trade flow effects of various institutional arrangements. We show that estimated gravity equations do not have a theoretical foundation. This implies both that estimation suffers from omitted variables bias and that comparative statics analysis is unfounded. We develop a method that (i) consistently and efficiently estimates a theoretical gravity equation and (ii) correctly calculates the comparative statics of trade frictions. We apply the method to solve the famous McCallum border puzzle. Applying our method, we find that national borders reduce trade between industrialized countries by moderate amounts of 20–50 percent. (JEL F10, F15)*

The gravity equation is one of the most empirically successful in economics. It relates bilateral trade flows to GDP, distance, and other factors that affect trade barriers. It has been widely used to infer trade flow effects of institutions such as customs unions, exchange-rate mechanisms, ethnic ties, linguistic identity, and international borders. Contrary to what is often stated, the empirical gravity equations do not have a theoretical foundation. The theory, first developed by Anderson (1979), tells us that after controlling for size, trade between two regions is decreasing in their bilateral trade barrier relative to the *average* barrier of the two regions to trade with all their partners. Intuitively, the more resistant to trade with all others a region is, the more it is pushed to trade with a given bilateral partner. We will refer to the theoretically appropriate average trade barrier as “multilateral resistance.” The empirical gravity literature either does not include any form of multilateral resistance in the analysis or in-

cludes an atheoretic “remoteness” variable related to distance to all bilateral partners. The remoteness index does not capture any of the other trade barriers that are the focus of the analysis. Moreover, even if distance were the only bilateral barrier, its functional form in the remoteness index is at odds with the theory.<sup>1</sup>

The lack of theoretical foundation of empirical gravity equations has two important implications. First, estimation results are biased due to omitted variables. Second, and perhaps even more important, one cannot conduct comparative statics exercises, even though this is generally the purpose of estimating gravity equations.<sup>2</sup> In order to conduct a comparative statics exercise, such as asking what the effects are of removing certain trade barriers, one has to be able to solve the general-equilibrium model before and after the removal of trade barriers. In this paper we will (i) develop a method that consistently and efficiently estimates a theoretical gravity equation, (ii) use the estimated general-equilibrium gravity model to conduct comparative statics exercises of the ef-

\* Anderson: Department of Economics, Boston College, Chestnut Hill, MA 02467 (e-mail: james.anderson@bc.edu), and NBER; van Wincoop: Department of Economics, University of Virginia, 116 Rouss Hall, Charlottesville, VA 22904 (e-mail: vanwincoop@virginia.edu), and NBER. We would like to thank two referees, Carolyn Evans, Robert Feenstra, Jim Harrigan, John Helliwell, Russell Hillberry, David Hummels, Andy Rose, and Kei-Mu Yi for helpful comments. We also thank seminar participants at Boston College, Brandeis University, Harvard University, Princeton University, Tilburg University, the University of California at Davis, the University of Colorado, the University of North Carolina, the University of Virginia, the University of Wisconsin, Vanderbilt University, and the 2000 NBER ITI Fall meeting for helpful comments.

<sup>1</sup> Jeffrey H. Bergstrand (1985, 1989) acknowledges the multilateral resistance term and deals with its time-series implications, but is unable to deal with the cross-section aspects which are crucial for proper treatment of bilateral trade barriers. Anderson and Douglas Marcouiller (2002) use a Törnqvist approximation to the multilateral resistance term which handles the cross-section variation of bilateral barriers.

<sup>2</sup> Recently, some authors (e.g., David Hummels, 1999) control for multilateral resistance in estimation with fixed effects, but cannot consistently do comparative statics on this basis.

fect of trade barriers on trade flows, and (iii) apply the theoretical gravity model to resolve the “border puzzle.”

One of the most celebrated inferences from the gravity literature is John McCallum's (1995) finding that the U.S.–Canadian border led to 1988 trade between Canadian provinces that is a factor 22 (2,200 percent) times trade between U.S. states and Canadian provinces. Maurice Obstfeld and Kenneth Rogoff (2001) pose it as one of their six puzzles of open economy macroeconomics. John F. Helliwell and McCallum (1995) document its violation of economists' prior beliefs. Gene Grossman (1998) says it is an unexpected result, even more surprising than Daniel Treffer's (1995) “mystery of the missing trade.” A rapidly growing literature is aimed at measuring and understanding trade border effects.<sup>3</sup> So far none of the subsequent research has explained McCallum's finding. We solve the border puzzle in this paper by applying the theory of the gravity equation seriously both to estimation and to the general-equilibrium comparative statics of borders.

The first step in solving the border puzzle is to estimate the gravity equation correctly based on the theory. In doing so we aim to stay as close as possible to McCallum's (1995) gravity equation, in which bilateral trade flows between two regions depend on the output of both regions, their bilateral distance, and whether they are separated by a border. The theory modifies McCallum's equation only by adding the multilateral resistance variables. The second step in solving the border puzzle is to conduct the general-equilibrium comparative statics exercise of removing the U.S.–Canada border barrier in order to determine the effect of the border on trade flows. The primary concern of policy makers and macroeconomic analysts is the impact of borders on *international* trade. McCallum's regression model (and the subsequent literature following him) cannot validly be used

to infer such border effects.<sup>4</sup> In contrast, our theoretically grounded approach can be used to compute the impact of borders both on *intra*national trade (within a country) and international trade. Applying our approach to 1993 data, we find that borders reduce trade between the United States and Canada by 44 percent, while reducing trade among other industrialized countries by 29 percent. While not negligible, we consider these to be plausibly moderate impacts of borders on international trade.

Two factors contribute to making McCallum's *ceteris paribus* ratio of interprovincial to province–state trade so large. First, his estimate is based on a regression with omitted variables, the multilateral resistance terms. Estimating McCallum's regression for 1993 data we find a ratio of 16.4, while our calculation based on asymptotically unbiased structural estimation and the computed general-equilibrium comparative statics of border removal implies a ratio of 10.7. Second, the magnitude of both ratios largely reflects the small size of the Canadian economy. If we estimate McCallum's regression with U.S. data, we find that trade between states is only a factor 1.5 times trade between states and provinces. The intuition is simple in the context of the model. Even a moderate barrier between Canada and the rest of the world has a large effect on multilateral resistance of the provinces because Canada it is a small open economy that trades a lot with the rest of the world (particularly the United States). This significantly raises interprovincial trade, by a factor 6 based on our estimated model. In contrast, the multilateral resistance of U.S. states is much less affected by a border barrier since it does not affect the barrier between a state and the rest of the large U.S. economy. Therefore trade between the states is not much increased by border barriers.

To a large extent the contribution of this paper is methodological. Our specification can be applied in many different contexts in which various aspects of implicit trade barriers are the focus. Gravity equations similar to McCallum's have been estimated to determine the impact of trade unions,<sup>5</sup> monetary

<sup>3</sup> See Hans Messinger (1993), Helliwell and McCallum (1995), Helliwell (1996, 1997, 1998), Shang-Jin Wei (1996), Russell Hillberry (1998, 1999, 2001), Michael A. Anderson and Stephen L. S. Smith (1999a, b), Jon Haveman and Hummels (1999), Hummels (1999), Natalie A. Chen (2000), Carolyn L. Evans (2000a, b), Holger Wolf (2000), Keith Head and John Ries (2001), Helliwell and Genevieve Verdier (2001), and Hillberry and Hummels (2002).

<sup>4</sup> McCallum cautiously did not claim that his estimated factor 22 implied that removal of the border would raise Canada–U.S. trade relative to within-Canada trade by 2,200 percent.

<sup>5</sup> See Jeffrey Frankel et al. (1998).

unions,<sup>6</sup> different languages, adjacency, and a variety of other factors; all can be improved with our methods. Authors have, like McCallum, often hesitated to draw comparative static inferences from their estimates. Using our methods, they can. Gravity equations have also been applied to migration flows, equity flows, and FDI flows.<sup>7</sup> Here there is no received theory to apply, consistently or not, but our results suggest the fruitfulness of theoretical foundations.

The remainder of the paper is organized as follows. In Section I we will provide some results based on McCallum's gravity equation. The main new aspect of this section is that we also report the results from the U.S. perspective, comparing interstate trade to state–province trade. In Section II we derive the theoretical gravity equation. The main innovation here is to rewrite it in a simple symmetric form, relating bilateral trade to size, bilateral trade barriers, and multilateral resistance variables. Section III discusses the procedure for estimating the theoretical gravity equation, both for a two-country version of the model, consisting of the United States and Canada, and for a multicountry version that also includes all other industrialized countries. The results are discussed in Section IV. Section V performs sensitivity analysis, and the final section concludes.

### I. The McCallum Gravity Equation

McCallum (1995) estimated the following equation:

$$(1) \quad \ln x_{ij} = \alpha_1 + \alpha_2 \ln y_i + \alpha_3 \ln y_j \\ + \alpha_4 \ln d_{ij} + \alpha_5 \delta_{ij} + \varepsilon_{ij}.$$

Here  $x_{ij}$  is exports from region  $i$  to region  $j$ ,  $y_i$  and  $y_j$  are gross domestic production in regions

$i$  and  $j$ ,  $d_{ij}$  is the distance between regions  $i$  and  $j$ , and  $\delta_{ij}$  is a dummy variable equal to one for interprovincial trade and zero for state–province trade. For the year 1988 McCallum estimated this equation using data for all 10 provinces and for 30 states that account for 90 percent of U.S.–Canada trade. In this section we will also report results when estimating equation (1) from the U.S. perspective. In that case the dummy variable is one for interstate trade and zero for state–province trade. We also report results when pooling all data, in which case there are two dummy variables. The first is one for interprovincial trade and zero otherwise, while the second is one for interstate trade and zero otherwise.

The data are discussed in Appendix A. Without going into detail here, a couple of comments are useful. The interprovincial and state–province trade data are from different divisions of Statistics Canada, while the interstate trade data are from the Commodity Flow Survey conducted by the Bureau of the Census. We follow McCallum by applying adjustment factors to the original data in order to make them as closely comparable as possible. All results reported below are for the year 1993, for which the interstate data are available. We follow McCallum and others by using data for only 30 states.

The results from estimating (1) are reported in Table 1. The first three columns report results for, respectively, (i) state–province and interprovincial trade, (ii) state–province and interstate trade, (iii) state–province, interprovincial, and interstate trade. In the latter case there are separate border dummies for within-U.S. trade and within-Canada trade. The final three columns report the same results after imposing unitary coefficients on the GDP variables. This makes comparison with our theoretically based gravity equation results easier because the theory imposes unitary coefficients.

*Border–Canada* is the exponential of the Canadian dummy variable coefficient,  $\alpha_5$ , which gives us the effect of the border on the ratio of interprovincial trade to state–province trade after controlling for distance and size. Similarly, *Border–U.S.* is the exponential of the coefficient on the U.S. dummy variable, which gives the effect of the border on the ratio of interstate trade to state–province trade after controlling for distance and size.

Four conclusions can be reached from the

<sup>6</sup> Andrew K. Rose (2000) finds that trade among countries in a monetary union is three times the size of trade among countries that are not in a monetary union, holding other trade costs constant. Rose and van Wincoop (2001) apply the theory developed in this paper to compute the effect of monetary unions on bilateral trade.

<sup>7</sup> The first application to migration flows dates from the nineteenth-century writings by Ernst G. Ravenstein (1885). For a more recent application see Helliwell (1997). Richard Portes and Helene Rey (1998) applied a gravity equation to bilateral equity flows. Paul Brenton et al. (1999) apply the gravity equation to FDI flows.

TABLE 1—McCALLUM REGRESSIONS

Data	McCallum regressions			Unitary income elasticities		
	(i) CA–CA CA–US	(ii) US–US CA–US	(iii) US–US CA–CA CA–US	(iv) CA–CA CA–US	(v) US–US CA–US	(vi) US–US CA–CA CA–US
Independent variable						
ln $y_i$	1.22 (0.04)	1.13 (0.03)	1.13 (0.03)	1	1	1
ln $y_j$	0.98 (0.03)	0.98 (0.02)	0.97 (0.02)	1	1	1
ln $d_{ij}$	–1.35 (0.07)	–1.08 (0.04)	–1.11 (0.04)	–1.35 (0.07)	–1.09 (0.04)	–1.12 (0.03)
<i>Dummy–Canada</i>	2.80 (0.12)		2.75 (0.12)	2.63 (0.11)		2.66 (0.12)
<i>Dummy–U.S.</i>		0.41 (0.05)	0.40 (0.05)		0.49 (0.06)	0.48 (0.06)
<i>Border–Canada</i>	16.4 (2.0)		15.7 (1.9)	13.8 (1.6)		14.2 (1.6)
<i>Border–U.S.</i>		1.50 (0.08)	1.49 (0.08)		1.63 (0.09)	1.62 (0.09)
$\bar{R}^2$	0.76	0.85	0.85	0.53	0.47	0.55
Remoteness variables added						
<i>Border–Canada</i>	16.3 (2.0)		15.6 (1.9)	14.7 (1.7)		15.0 (1.8)
<i>Border–U.S.</i>		1.38 (0.07)	1.38 (0.07)		1.42 (0.08)	1.42 (0.08)
$\bar{R}^2$	0.77	0.86	0.86	0.55	0.50	0.57

Notes: The table reports the results of estimating a McCallum gravity equation for the year 1993 for 30 U.S. states and 10 Canadian provinces. In all regressions the dependent variable is the log of exports from region  $i$  to region  $j$ . The independent variables are defined as follows:  $y_i$  and  $y_j$  are gross domestic production in regions  $i$  and  $j$ ;  $d_{ij}$  is the distance between regions  $i$  and  $j$ ; *Dummy–Canada* and *Dummy–U.S.* are dummy variables that are one when both regions are located in respectively Canada and the United States, and zero otherwise. The first three columns report results based on nonunitary income elasticities (as in the original McCallum regressions), while the last three columns assume unitary income elasticities. Results are reported for three different sets of data: (i) state–province and interprovincial trade, (ii) state–province and interstate trade, (iii) state–province, interprovincial, and interstate trade. The border coefficients *Border–U.S.* and *Border–Canada* are the exponentials of the coefficients on the respective dummy variables. The final three rows report the border coefficients and  $\bar{R}^2$  when the remoteness indices (3) are added. Robust standard errors are in parentheses.

table. First, we confirm a very large border coefficient for Canada. The first column shows that, after controlling for distance and size, interprovincial trade is 16.4 times state–province trade. This is only somewhat lower than the border effect of 22 that McCallum estimated based on 1988 data. Second, the U.S. border coefficient is much smaller. The second column tells us that interstate trade is a factor 1.50 times state–province trade after controlling for distance and size. We will show below that this large difference between the Canadian and U.S. border coefficients is exactly what the theory predicts. Third, these border coefficients are very similar when pooling all the data. Finally, the border coefficients are also similar

when unitary income coefficients are imposed. With pooled data and unitary income coefficients (last column), the Canadian border coefficient is 14.2 and the U.S. border coefficient is 1.62.

The bottom of the table reports results when remoteness variables are added. We use the definition of remoteness that has been commonly used in the literature following McCallum’s paper. The regression then becomes

(2)  $\ln x_{ij} = \alpha_1 + \alpha_2 \ln y_i + \alpha_3 \ln y_j + \alpha_4 \ln d_{ij}$

$+ \alpha_5 \ln REM_i + \alpha_6 \ln REM_j$

$+ \alpha_7 \delta_{ij} + \varepsilon_{ij}$

where the remoteness of region  $i$  is

$$(3) \quad REM_i = \sum_{m \neq j} d_{im}/y_m.$$

This variable is intended to reflect the average distance of region  $i$  from all trading partners other than  $j$ . Although these remoteness variables are commonly used in the literature, we will show in the next section that they are entirely disconnected from the theory. Table 1 shows that adding remoteness indices for both regions changes the border coefficient estimates very little and also has very little additional explanatory power based on the adjusted  $R^2$ .

## II. The Gravity Model

The empirical literature cited above pays no more than lip service to theoretical justification. We show in this section how taking the existing gravity theory seriously provides a different model to estimate with a much more useful interpretation.

Anderson (1979) presented a theoretical foundation for the gravity model based on constant elasticity of substitution (CES) preferences and goods that are differentiated by region of origin. Subsequent extensions (Bergstrand, 1989, 1990; Alan V. Deardoff, 1998) have preserved the CES preference structure and added monopolistic competition or a Heckscher-Ohlin structure to explain specialization. A contribution of this paper is our manipulation of the CES expenditure system to derive an operational gravity model with an elegantly simple form. On this basis we derive a decomposition of trade resistance into three intuitive components: (i) the bilateral trade barrier between region  $i$  and region  $j$ , (ii)  $i$ 's resistance to trade with all regions, and (iii)  $j$ 's resistance to trade with all regions.

The first building block of the gravity model is that all goods are differentiated by place of origin. We assume that each region is specialized in the production of only one good.<sup>8</sup> The supply of each good is fixed.

<sup>8</sup> With this assumption we suppress finer classifications of goods. Our purpose is to reveal resistance to trade on average, with special reference to the proper treatment of international borders. Resistance to trade does differ among goods, so there is something to be learned from disaggregation.

The second building block is identical, homothetic preferences, approximated by a CES utility function. If  $c_{ij}$  is consumption by region  $j$  consumers of goods from region  $i$ , consumers in region  $j$  maximize

$$(4) \quad \left( \sum_i \beta_i^{(1-\sigma)/\sigma} c_{ij}^{(\sigma-1)/\sigma} \right)^{\sigma/(\sigma-1)}$$

subject to the budget constraint

$$(5) \quad \sum_i p_{ij} c_{ij} = y_j.$$

Here  $\sigma$  is the elasticity of substitution between all goods,  $\beta_i$  is a positive distribution parameter,  $y_j$  is the nominal income of region  $j$  residents, and  $p_{ij}$  is the price of region  $i$  goods for region  $j$  consumers. Prices differ between locations due to trade costs that are not directly observable, and the main objective of the empirical work is to identify these costs. Let  $p_i$  denote the exporter's supply price, net of trade costs, and let  $t_{ij}$  be the trade cost factor between  $i$  and  $j$ . Then  $p_{ij} = p_i t_{ij}$ .

We assume that the trade costs are borne by the exporter. We have in mind information costs, design costs, and various legal and regulatory costs as well as transport costs. The new empirical literature on the export behavior of firms (Mark Roberts and James Tybout, 1997; Andrew Bernard and Joachim Wagner, 2001) emphasizes the large costs facing exporters. Formally, we assume that for each good shipped from  $i$  to  $j$  the exporter incurs export costs equal to  $t_{ij} - 1$  of country  $i$  goods. The exporter passes on these trade costs to the importer. The nominal value of exports from  $i$  to  $j$  ( $j$ 's payments to  $i$ ) is  $x_{ij} = p_{ij} c_{ij}$ , the sum of the value of production at the origin,  $p_i c_{ij}$  and the trade cost  $(t_{ij} - 1)p_i c_{ij}$  that the exporter passes on to the importer. Total income of region  $i$  is therefore  $y_i = \sum_j x_{ij}$ .<sup>9</sup>

<sup>9</sup> The model is essentially the same when adopting the "iceberg melting" structure of the economic geography literature, whereby a fraction  $(t_{ij} - 1)/t_{ij}$  of goods shipped is lost in transport. The only small difference is that observed free on board (f.o.b.) trade data do not include transportation costs, while they do include costs that are borne by the exporter and passed on to the importer. When transportation costs are the only trade costs, the observed f.o.b. trade flows

The nominal demand for region  $i$  goods by region  $j$  consumers satisfying maximization of (4) subject to (5) is

$$(6) \quad x_{ij} = \left( \frac{\beta_i p_i t_{ij}}{P_j} \right)^{(1-\sigma)} y_j,$$

where  $P_j$  is the consumer price index of  $j$ , given by

$$(7) \quad P_j = \left[ \sum_i (\beta_i p_i t_{ij})^{1-\sigma} \right]^{1/(1-\sigma)}.$$

The general-equilibrium structure of the model imposes market clearance, which implies:

$$\begin{aligned} (8) \quad y_i &= \sum_j x_{ij} \\ &= \sum_j (\beta_i t_{ij} p_i / P_j)^{1-\sigma} y_j \\ &= (\beta_i p_i)^{1-\sigma} \sum_j (t_{ij} / P_j)^{1-\sigma} y_j, \quad \forall i. \end{aligned}$$

To derive the gravity equation, Deardorff (1998) followed Anderson (1979) in using market clearance (8) to solve for the coefficients  $\{\beta_i\}$  while imposing the choice of units such that all supply prices  $p_i$  are equal to one and then substituting into the import demand equation.<sup>10</sup> Because we are interested in the general-equilibrium determination of prices and in comparative statics where these will change, we apply the same technique to solve for the scaled prices  $\{\beta_i p_i\}$  from the market-clearing conditions (8) and substitute them in the demand equation (6). Define world nominal income by  $y^W \equiv \sum_j y_j$  and income shares by  $\theta_j \equiv y_j / y^W$ . The technique yields

are equal to  $p_i c_{ij}$ . The same is the case when the costs are borne by the importer. While we believe that most trade costs are borne by the exporter, particularly for U.S.–Canada trade where formal import barriers are very low, this is not critical to the findings of the paper; the results would be similar when assuming that observed trade flows are equal to  $p_i c_{ij}$ .

<sup>10</sup> Deardorff simplified by abstracting from the multiple goods classes which Anderson allowed in his Appendix on the CES case.

$$(9) \quad x_{ij} = \frac{y_i y_j}{y^W} \left( \frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma}$$

where

$$(10) \quad \Pi_i \equiv \left( \sum_j (t_{ij} / P_j)^{1-\sigma} \theta_j \right)^{1/(1-\sigma)}.$$

Substituting the equilibrium scaled prices into (7), we obtain

$$(11) \quad P_j = \left( \sum_i (t_{ij} / \Pi_i)^{1-\sigma} \theta_i \right)^{1/(1-\sigma)}.$$

Taken together, (10) and (11) can be solved for all  $\Pi_i$ 's and  $P_i$ 's in terms of income shares  $\{\theta_i\}$ , bilateral trade barriers  $\{t_{ij}\}$  and  $\sigma$ .

We achieve a very useful simplification by assuming that the trade barriers are symmetric, that is,  $t_{ij} = t_{ji}$ .<sup>11</sup> Under symmetry it is easily verified that a solution to (10)–(11) is  $\Pi_i = P_i$  with:

$$(12) \quad P_j^{1-\sigma} = \sum_i P_i^{\sigma-1} \theta_i t_{ij}^{1-\sigma} \quad \forall j.$$

This provides an implicit solution to the price indices as a function of all bilateral trade barriers and income shares.<sup>12</sup> The gravity equation then becomes

$$(13) \quad x_{ij} = \frac{y_i y_j}{y^W} \left( \frac{t_{ij}}{P_i P_j} \right)^{1-\sigma}.$$

<sup>11</sup> There are many equilibria with asymmetric barriers that lead to the same equilibrium trade flows as with symmetric barriers, so that empirically they are impossible to distinguish. In particular, if  $\lambda_i$  and  $\lambda_j$  are region-specific constants, multiplying  $t_{ij}$  by  $\lambda_j / \lambda_i$ ,  $\forall i, j$  leads to the same equilibrium trade flows [ $P_i$  is multiplied by  $\lambda_i$  and  $P_j$  is multiplied by  $\lambda_j$  in (8)]. The product of the trade barriers in different directions remains the same though. If the  $\lambda$ 's are country specific, but differ across countries, we have introduced asymmetric border barriers across countries, while the product of border barriers remains the same. We can therefore interpret the border barriers we estimate in this paper as an average of the barriers in both directions. Our analysis suggests that inferential identification of the asymmetry is problematic.

<sup>12</sup> The solution for the equilibrium price indexes from (12) can be shown to be unique. If we denote by  $\bar{P}_i = \bar{\Pi}_i$  the solution to (12), the general solution to (10)–(11) is  $P_i = \lambda \bar{P}_i$  and  $\Pi_i = \bar{\Pi}_i / \lambda$  for any nonzero  $\lambda$ . The solution (12) therefore implicitly adopts a particular normalization.

Our basic gravity model is (13) subject to (12). Equation (13) significantly simplifies expressions derived by Anderson (1979) and Dear-dorff (1998), while our simultaneous use of the market-clearing constraints to obtain the equilibrium price indexes in (12) is a significant innovation that will allow us to estimate the gravity equation and therefore make it operational.

We will refer to the price indices  $\{P_i\}$  as "multilateral resistance" variables as they depend on all bilateral resistances  $\{t_{ij}\}$ , including those not directly involving  $i$ . A rise in trade barriers with all trading partners will raise the index. For example, in the absence of trade barriers (all  $t_{ij} = 1$ ) it follows immediately from (12) that all price indices are equal to 1. Below we will show that a marginal increase in cross-country trade barriers will raise all price indices above 1.

While the  $P_i$  are consumer price indices in the model, that would not be a proper interpretation of these indices more generally. One can derive exactly the same gravity equation and solution to the  $P_i$  when trade costs are nonpecuniary. An example is home bias in preferences, whereby  $c_{ij}$  in the utility function is replaced by  $c_{ij}/t_{ij}$ . In that case  $P_i$  no longer represents the consumer price index and the border barrier includes home bias.

The gravity equation tells us that bilateral trade, after controlling for size, depends on the bilateral trade barrier between  $i$  and  $j$ , relative to the product of their multilateral resistance indices. It is easy to see why higher multilateral resistance of the importer  $j$  raises its trade with  $i$ . For a given bilateral barrier between  $i$  and  $j$ , higher barriers between  $j$  and its other trading partners will reduce the relative price of goods from  $i$  and raise imports from  $i$ . Higher multilateral resistance of the exporter  $i$  also raises trade. Higher trade barriers faced by an exporter will lower the demand for its goods and therefore its supply price  $p_i$ . For a given bilateral barrier between  $i$  and  $j$ , this raises the level of trade between them.

The gravity model (13), subject to (12), implies that bilateral trade is homogeneous of degree zero in trade costs, where these include the costs of shipping within a region,  $t_{ii}$ . This follows because the equilibrium multilateral resistances  $P_i$  are homogeneous of degree  $1/2$  in the trade costs. The economics behind the formal result is that the constant vector of real products must be distributed despite higher trade costs.

The rise in trade costs is offset by the fall in supply prices [they are homogeneous of degree minus  $1/2$  in trade costs, based on (7) and the homogeneity of the equilibrium multilateral resistances] required to achieve shipment of the same volume. The invariance of trade to uniform decreases in trade costs may offer a clue as to why the usual gravity model estimation has not found trade becoming less sensitive to distance over time (Barry Eichengreen and Douglas A. Irwin, 1998).

The key implication of the theoretical gravity equation is that trade between regions is determined by *relative* trade barriers. Trade between two regions depends on the bilateral barrier between them relative to average trade barriers that both regions face with all their trading partners. This insight has many implications for the impact of trade barriers on trade flows. Here we will focus on one important set of implications related to the size of countries because they are useful in interpreting the findings in Section IV. Consider the simple thought experiment of a uniform rise in border barriers between all countries. For simplicity we assume that each region  $i$  is a frictionless country. We will discuss three general-equilibrium comparative static implications of this experiment, which are listed below.

**IMPLICATION 1:** *Trade barriers reduce size-adjusted trade between large countries more than between small countries.*

**IMPLICATION 2:** *Trade barriers raise size-adjusted trade within small countries more than within large countries.*

**IMPLICATION 3:** *Trade barriers raise the ratio of size-adjusted trade within country 1 relative to size-adjusted trade between countries 1 and 2 by more the smaller is country 1 and the larger is country 2.*

The experiment amounts to a marginal increase in trade barriers *across* all countries, so  $dt_{ij} = dt$ ,  $i \neq j$ ;  $dt_{ii} = 0$ . Frictionless initial equilibrium implies  $t_{ij} = 1 \forall i, j \Rightarrow P_i = 1$ . Differentiating (12) at  $t_{ij} = 1$ ,  $\forall i, j$  yields<sup>13</sup>

<sup>13</sup> To obtain this expression we differentiate totally at  $t_{ij} = 1 = P_i$  to obtain



$$(14) \quad dP_i = \left( \frac{1}{2} - \theta_i + \frac{1}{2} \sum_k \theta_k^2 \right) dt.$$

Thus a uniform increase in trade barriers raises multilateral resistance more for a small country than a large country.<sup>14</sup> In a two-country example, where the small country's income is 10 percent of the total, a 20-percent trade barrier raises the price index of the large country by 0.2 percent, while raising the price index of the small country by 16 percent. This is not unlike the U.S.–Canada example to which the model will be applied later. For a very large country multilateral resistance is not much affected because increased trade barriers do not apply to trade within the country. For a small country trade is more important and trade barriers therefore have a bigger effect on multilateral resistance.

Equation (14) implies that the level of trade between countries  $i$  and  $j$ , after controlling for size, changes by

(15)

$$d\left(x_{ij} \frac{y^w}{y_i y_j}\right) = -(\sigma - 1) \left[ \theta_i + \theta_j - \sum_k \theta_k^2 \right] dt.$$

This implies that trade between large countries drops more than trade between small countries (Implication 1). While two small countries face a larger bilateral trade barrier, they face the same increase in trade barriers with almost the entire world. Bilateral trade depends on the relative trade resistance  $t_{ij}/P_i P_j$ . Since multilateral trade resistance rises much more for small countries than for large countries, relative trade re-

sistance rises less for small countries, so that their bilateral trade drops less.<sup>15</sup>

Equation (14) also implies that trade within a country  $i$ , after controlling for size, increases by

(16)

$$d\left(x_{ii} \frac{y^w}{y_i y_i}\right) = (\sigma - 1) \left[ 1 - 2\theta_i + \sum_k \theta_k^2 \right] dt.$$

Therefore trade within a small country increases more than trade within a large country (Implication 2). A rise in multilateral resistance implies a drop in relative resistance  $t_{ii}/P_i P_i$  for intranational trade. The drop is larger for small countries that face a bigger increase in multilateral resistance.

Implication 3 follows from the previous two. After controlling for size, trade within country  $i$  relative to trade between countries  $i$  and  $j$  rises by

$$(17) \quad d\left(\frac{x_{ii}/y_i y_i}{x_{ij}/y_i y_j}\right) = (\sigma - 1)[1 - \theta_i + \theta_j] dt.$$

The increase is larger the smaller  $i$  and the bigger  $j$ . We already knew from Implication 2 that intranational trade rises most for small countries. From Implication 1 we also know that for a given small country international trade drops most with large countries.

The implications relating to size are much more general than the specifics of the model might suggest. Consider the following example without any reference to gravity equations and multilateral resistance variables. A small economy with two regions and a large economy with 100 regions engage in international trade. All regions have the same GDP. What matters here is not the number of regions, but the relative size of the two economies as measured by total GDP. We only introduce regions in this example because it is illustrative in the context of the U.S. states and Canadian provinces that are the focus of the empirical analysis. Under borderless

$$dP_j = \sum_i \theta_i dt_{ij} - \sum_i \theta_i dP_i + \frac{1}{1 - \sigma} \sum_i d\theta_i.$$

$\sum_i d\theta_i = 0$ , since the sum of the shares is equal to one. Multiplying each equation by  $\theta_j$  and summing using  $dt_{ij} = dt$ ,  $i \neq j$ ,  $dt_{ii} = 0$ , we solve for  $\sum \theta_j dP_j = (1 - \sum \theta_j^2)dt/2$  and thus  $dP_i = (\frac{1}{2} - \theta_i + \sum \theta_j^2/2)dt$ .

<sup>14</sup> Country size is determined by the endowment of the goods. It can be shown that at the frictionless equilibrium, a rise in country  $i$ 's endowment will lower its supply price  $p_i$ , raise all other supply prices, and with  $\sigma > 1$  this will raise  $\theta_i$  and lower the other income shares. Thus we treat  $\theta_i$  as an exogenous variable for the purposes of talking about country size.

<sup>15</sup> As is immediately clear from (15), trade between two small countries can even rise after a uniform increase in trade barriers. This is because the pre-barrier prices  $p_i$  drop more in small countries than in large countries as small countries are more affected by a drop in foreign demand. This makes it more attractive for small countries to trade with each other than with large countries.

trade, all regions sell one unit of one good to all 102 regions (including themselves). Now impose a barrier between the small and the large country, reducing trade between the two countries by 20 percent. Region 1 in the small country then reduces its exports to the large country by 20. It sells ten more goods to itself and ten more goods to region 2 in the small country. Trade between the two regions in the small country rises by a factor 11, while trade between two regions in the large country rises by a factor of only 1.004 (an illustration of Implication 2 above). This shows that even a small drop in international trade can lead to a very large increase in trade within a small country. Trade between the two regions in the small country is now 13.75 times trade between regions in both countries, while trade between two regions in the large country is only 1.255 times trade between regions in the two countries (an illustration of Implication 3).

The final step in our theoretical development of the gravity equation is to model the unobservable trade cost factor  $t_{ij}$ . We follow other authors in hypothesizing that  $t_{ij}$  is a loglinear function of observables, bilateral distance  $d_{ij}$ , and whether there is an international border between  $i$  and  $j$ :

$$(18) \quad t_{ij} = b_{ij} d_{ij}^{\rho},$$

$b_{ij} = 1$  if regions  $i$  and  $j$  are located in the same country. Otherwise  $b_{ij}$  is equal to one plus the tariff equivalent of the border barrier between the countries in which the regions are located. Other investigators have added other factors related to trade barriers, such as adjacency and linguistic identity. We have chosen the trade costs specification (18) to stay as close as possible to McCallum's (1995) equation, so that we can keep the focus on the multilateral resistance indices that are absent from McCallum's analysis.

We can now compare the theoretical gravity equation with that estimated in the empirical literature. The theory implies that

$$(19) \quad \begin{aligned} \ln x_{ij} = & k + \ln y_i + \ln y_j + (1 - \sigma)\rho \ln d_{ij} \\ & + (1 - \sigma)\ln b_{ij} - (1 - \sigma)\ln P_i \\ & - (1 - \sigma)\ln P_j \end{aligned}$$

where  $k$  is a constant. The key difference between (20) and equation (1) estimated by McCallum is the two price index terms. The omitted multilateral resistance variables are functions of all bilateral trade barriers  $t_{ij}$  through (12), which in turn are a function of  $d_{ij}$  and  $b_{ij}$  through the trade cost equation (18). Since the multilateral resistance terms are therefore correlated with  $d_{ij}$  and  $b_{ij}$ , they create omitted variable bias when the coefficient of the distance and border variables is interpreted as  $(1 - \sigma)\rho$  and  $(1 - \sigma)\ln b_{ij}$ . Our multilateral resistance variables bear some resemblance to "remoteness" indexes such as (3) that have been included in gravity equation estimates subsequent to McCallum's paper. But the latter do not include border barriers and even without border barriers the functional form is entirely disconnected from the theory. Finally, our multilateral resistance variables as equilibrium constructs are functions of all bilateral resistances in the solution to (12).

A small difference between the theory and the empirical literature is that the theoretical gravity equation imposes unitary income elasticities. Anderson (1979) provided a rationale for earlier (and subsequent) empirical gravity work that estimates nonunitary income elasticities. He allowed for nontraded goods and assumed a reduced-form function of the expenditure share falling on traded goods as a function of total income. We already found in Section I that imposing unitary income elasticities has little effect on McCallum's border estimates. We will therefore impose unitary income elasticities in most of the analysis, leaving an extension to nonunitary elasticities to sensitivity analysis.

### III. Estimation

We implement the theory both in the context of a two-country model, consisting of the United States and Canada, and a multicountry model that also includes other industrialized countries. The latter approach is obviously more realistic as it takes into account that the United States and Canada also trade with other countries. It has the additional advantage that it delivers an estimate of the impact of border barriers on trade among the other industrialized countries. We first discuss the two-country model and then the multicountry model.

### A. Two-Country Model

In the two-country model we estimate the gravity equation for trade flows among the same 30 states and 10 provinces as in McCallum (1995). We do not include in the sample the other 21 regions (20 states plus the District of Columbia), which account for about 15 percent of U.S. GDP, and trade flows internal to a state or province. However, in order to compute the multilateral resistance variables for the regions in our sample, we do need to use information on size and distance associated with the other 21 regions and we also need to use information on the distances within regions. We simplify by aggregating the other 21 regions into one region, defining the distance between this region and region  $i$  in our sample as the GDP weighted average of the distance between  $i$  and each of the 21 regions that make up the new region. There is no obvious way to compute distances internal to a region. Fortunately, as we will show in Section V, our results are not very sensitive to assumptions about internal distance. We use the proxy developed by Wei (1996), which is one-fourth the distance of a region's capital from the nearest capital of another region.<sup>16</sup>

In the two-country model  $b_{ij} = b^{1-\delta_{ij}}$ , where  $b - 1$  represents the tariff-equivalent U.S.–Canada border barrier and  $\delta_{ij}$  is the same dummy variable as in Section I, equal to one if  $i$  and  $j$  are in the same country and zero otherwise.

We estimate a stochastic form of (13):

$$(20) \quad \ln z_{ij} \equiv \ln \left( \frac{x_{ij}}{y_i y_j} \right) \\ = k + a_1 \ln d_{ij} + a_2 (1 - \delta_{ij}) \\ - \ln P_i^{1-\sigma} - \ln P_j^{1-\sigma} + \varepsilon_{ij}$$

where  $a_1 = (1 - \sigma)\rho$  and  $a_2 = (1 - \sigma)\ln b$ . To stay as close as possible to McCallum's (1995) regression we have simply added an

error term to the logarithmic form of the gravity equation, which one can think of as reflecting measurement error in trade. Apart from the unitary income elasticities, the only difference with McCallum (1995) is the presence of the two multilateral resistance terms.

The multilateral resistance terms are not observables. As discussed above, the price indices in general cannot be interpreted as consumer price levels.<sup>17</sup> The observables in our model are distances, borders, and income shares. Using the 41 goods market-equilibrium conditions (12) and the trade cost function (18), we can solve for the vector of the  $P_i^{1-\sigma}$  as an implicit function of observables and model parameters  $a_1$  and  $a_2$ :

$$(21) \quad P_j^{1-\sigma} = \sum_i P_i^{\sigma-1} \theta_i e^{a_1 \ln d_{ij} + a_2 (1 - \delta_{ij})} \\ j = 1, \dots, 41.$$

After substituting the implicit solutions for the  $P_i^{1-\sigma}$  in (21), the gravity equation to be estimated becomes:

$$(22) \quad \ln \mathbf{z} = h(\mathbf{d}, \boldsymbol{\delta}, \boldsymbol{\theta}; k, a_1, a_2) + \boldsymbol{\varepsilon}$$

where  $\mathbf{z}$ ,  $\mathbf{d}$ ,  $\boldsymbol{\delta}$ ,  $\boldsymbol{\theta}$ , and  $\boldsymbol{\varepsilon}$  are vectors that contain all the elements of the corresponding variables with subscripts, and  $h(\cdot)$  is the right-hand side of (20) after substituting the equilibrium  $P_i^{1-\sigma}$  and  $P_j^{1-\sigma}$ .

The right-hand side is now written explicitly as a function of observables. We estimate (22) with nonlinear least squares, minimizing the sum of squared errors. For any set of parameters the error terms of the regression can only be computed after first solving for 41 equations (21). The estimated parameters are  $k$ ,  $a_1$ , and

<sup>16</sup> For the region obtained from the aggregation of the 21 regions, we compute internal distance as  $\sum_{i=1}^{21} \sum_{j=1}^{21} s_i s_j d_{ij}$ , where  $s_i$  is the ratio of GDP in region  $i$  to total GDP of the 21-region area.

<sup>17</sup> Even if one assumes that the price indices are consumer price levels, which would require that all trade costs are pecuniary costs, there are still many measurement problems that makes them unobservable for our purposes. Non-traded goods, which are not present in our model, play a key role in explaining differences in price levels across countries and regions. In the short to medium run, nominal exchange rates also have a significant impact on the ratio of price levels across countries. Moreover, while comparable price-level data are available for countries, this is not the case for states and provinces.

$a_2$ .<sup>18</sup> The substitution elasticity  $\sigma$  cannot be estimated separately as it enters in multiplicative form with the trade cost parameters  $\rho$  and  $\ln b$  in  $a_1$  and  $a_2$ .<sup>19</sup>

Our estimator is unbiased if  $\varepsilon$  is uncorrelated with the derivatives of  $\mathbf{h}$  with respect to  $\mathbf{d}$ ,  $\delta$ , and  $\theta$ . This is not a problem when we interpret  $\varepsilon_{ij}$  simply as measurement error associated with bilateral trade, as we have done. Errors can enter the model in many other ways of course, about which the theory has little to say. In particular, it is possible that the trade cost function (18) is misspecified in that other factors than just distance and borders matter, or the functional form is incorrect. One could add an error term to the trade cost function to capture this. If this error term is correlated with  $\mathbf{d}$  or  $\delta$ , our estimates will be biased. But this is a standard omitted variables problem that is not specific to the presence of multilateral resistance terms. We have chosen the trade cost function to stay as close as possible to McCallum's (1995) specification. If an error term in the trade cost function is uncorrelated with  $\mathbf{d}$  and  $\delta$ , there is still the problem that the error term affects equilibrium prices and therefore income shares  $\theta$ , which affect the multilateral resistance terms. In practice the bias resulting from this is very small though. As we will report below, even if we take the dramatic step of entirely removing the U.S.–Canada border, practically none of the resulting changes in the  $P_i^{1-\sigma}$  are associated with changes in income shares.

An alternative to the estimation method described above is to replace the multilateral resistance terms with country-specific dummies. This leads to consistent estimates of model parameters. Hummels (1999) has done so for a gravity equation using disaggregated U.S. import data. The main advantage is simplicity as

ordinary least squares can be used. Another advantage is that we do not need to make any assumptions about distances internal to states and provinces, which are needed to compute the structural multilateral resistance terms and are difficult to measure. Rose and van Wincoop (2001) use this estimator when applying the method in this paper to determine the effect on trade of monetary unions. We need to emphasize though that the fixed-effects estimator is less efficient than the nonlinear least-squares estimator discussed above, which uses information on the full structure of the model. The simple fixed-effects estimator is not necessarily more robust to specification error. For example, if the trade cost function is misspecified, either in terms of functional form or set of variables, both estimators are biased to the extent that the specification error is correlated with distance or the border dummy.

For comparative statics analysis, such as removing the U.S.–Canada border, the structural model can be used with either method of estimation. We use the fixed-effects estimator in sensitivity analysis reported in Table 6, giving similar results.

### B. Multicountry Model

In the multicountry model the world consists of all industrialized countries, a total of 22 countries.<sup>20</sup> In that case there are 61 regions in our analysis: 30 states, the rest of the United States, 10 provinces, and 20 other countries. We will often refer to the 20 additional countries as ROW (rest of the world). In this expanded environment we assume that border barriers  $b_{ij}$  may differ for U.S.–Canada trade, US–ROW trade, Canada–ROW trade, and ROW–ROW trade. We define these respectively as  $b_{US,CA}$ ,  $b_{US,ROW}$ ,  $b_{CA,ROW}$ , and  $b_{ROW,ROW}$ .

For consistency with the estimation method in the two-country model, and given our focus on the U.S.–Canada border effect, we will continue to estimate the parameters by minimizing the sum of the squared residuals for the 30 states and 10 provinces. But there are now three ad-

<sup>18</sup> Computationally, we solve

$$\begin{aligned} \min_{k, a_1, a_2} \sum_i \sum_{j \neq i} [\ln z_{ij} - k - a_1 \ln d_{ij} - a_2 (1 - \delta_{ij}) \\ + \ln P_i^{1-\sigma} + \ln P_j^{1-\sigma}]^2 \\ \text{subject to } P_j^{1-\sigma} = \sum_i P_i^{\sigma-1} \theta_i e^{a_1 \ln d_{ij} + a_2 (1 - \delta_{ij})} \forall j. \end{aligned}$$

<sup>19</sup> As Hummels (1999) has shown, identification of  $\sigma$  is possible in applications where elements of  $t_{ij}$  are directly observable, as with tariffs.

<sup>20</sup> Those are the United States, Canada, Australia, Japan, New Zealand, Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

ditional parameters that affect the multilateral resistance variables of the states and provinces:  $(1 - \sigma)\ln b_{US,ROW}$ ,  $(1 - \sigma)\ln b_{CA,ROW}$ , and  $(1 - \sigma)\ln b_{ROW,ROW}$ . We impose three constraints in order to obtain estimates for these parameters. The constraints set the average of the residuals for US–ROW trade, CA–ROW trade, and ROW–ROW trade equal to zero.<sup>21</sup> Formally,

$$\sum_{j \in ROW} (\varepsilon_{US,j} + \varepsilon_{j,US}) = 0$$

$$\sum_{j \in ROW} (\varepsilon_{CA,j} + \varepsilon_{j,CA}) = 0$$

$$\sum_{i,j \in ROW, i \neq j} \varepsilon_{ij} = 0.$$

Since we have data on trade only between the ROW countries and all of the United States, the residuals  $\varepsilon_{US,j}$  and  $\varepsilon_{j,US}$  are defined as the log of bilateral trade between the United States and country  $j$  minus the log of predicted trade, where the latter is obtained by summing over the model's predicted trade between  $j$  and all U.S. regions. The same is done for trade between Canada and countries in ROW.<sup>22</sup>

<sup>21</sup> Apart from consistency with the two-country estimation method, there are two reasons why we prefer this estimation method as opposed to minimizing the sum of all squared residuals, including those of the ROW countries. First, border barriers are likely to be different across country pairs for the 20 other industrialized countries. Neither estimation method allows us to identify all these barriers separately, but the method we chose is less sensitive to such differences as we only use information on the average error terms involving the ROW countries. Second, the alternative method of minimizing the sum of all squared residuals has weaker finite sample properties. The US–ROW barrier has a much greater impact on US–ROW trade than on trade among the states and provinces, but US–ROW observations are only 2 percent of the sample. If there is only weak spurious correlation between the 1,511 error terms for trade among states and provinces and the partial derivatives of the corresponding multilateral resistance terms with respect to the US–ROW barrier, it could significantly affect the estimate of that barrier.

<sup>22</sup> Data on exports from individual states to ROW countries do exist (see Robert C. Feenstra, 1997), but this is based on information about the location of the exporter, which is often not the location of the plant where the goods are produced. The International Trade Division and the Input Output Divisions of Statistics Canada both report data on trade between provinces and the rest of the world. The data from the IO Division are considered more reliable, but

## IV. Results

Our goal in this section is threefold. First, we report results from estimating the theoretical gravity equation. Second, we use the estimated gravity equation to determine the impact of national borders on trade flows. This is done by computing the change in bilateral trade flows after removing the border barriers. Finally, we use the estimated gravity equation to account for the estimated McCallum border parameters. This procedure illustrates the role of the multilateral resistance variables in generating a much smaller McCallum border parameter for the United States than for Canada as well as the effect of the omitted variable bias in McCallum's procedure.

### A. Parameter Estimates

Table 2 reports the bilateral trade resistance parameter estimates. The estimate of the U.S.–Canada border barrier is very similar in both the two-country model and the multicountry model. In the multicountry model the border barrier estimates are also strikingly similar across country pairs. The barrier between the United States and Canada is only slightly lower than between the other 20 industrialized countries, the majority of which is trade among European Union countries. The only border barrier that is a bit higher than the others is between Canada and the ROW countries.

As discussed above, we can estimate only  $(1 - \sigma)\ln b$ . We would need to make an assumption about the elasticity of substitution  $\sigma$  in order to obtain an estimate of  $b - 1$ , the ad valorem tariff equivalent of the border barrier. The model is of course highly stylized in that there is only one elasticity. In reality some goods may be perfect substitutes, with an infinite elasticity, while others are weak substitutes. Hummels (1999) obtains estimates for the elasticity of substitution within industries. The results depend on the disaggregation of the industries. The average elasticity is respectively 4.8, 5.6, and 6.9 for 1-digit, 2-digit, and 3-digit

only the IT division reports trade with individual countries. The differences between the total export and import numbers reported by both divisions are often very large (almost a factor 8 difference for imports by Prince Edward Island).

TABLE 2—ESTIMATION RESULTS

		Two-country model	Multicountry model
Parameters	$(1 - \sigma)\rho$	-0.79 (0.03)	-0.82 (0.03)
	$(1 - \sigma)\ln b_{US,CA}$	-1.65 (0.08)	-1.59 (0.08)
	$(1 - \sigma)\ln b_{US,ROW}$		-1.68 (0.07)
	$(1 - \sigma)\ln b_{CA,ROW}$		-2.31 (0.08)
	$(1 - \sigma)\ln b_{ROW,ROW}$		-1.66 (0.06)
Average error terms:	US-US	0.06	0.06
	CA-CA	-0.17	-0.02
	US-CA	-0.05	-0.04

*Notes:* The table reports parameter estimates from the two-country model and the multicountry model. Robust standard errors are in parentheses. The table also reports average error terms for interstate, interprovincial, and state-province trade.

industries. For further levels of disaggregation the elasticities could be much higher, with some goods close to perfect substitutes.<sup>23</sup> It is therefore hard to come up with an appropriate average elasticity. To give a sense of the numbers though, the estimate of  $-1.58$  for  $(1 - \sigma)\ln b_{US,CA}$  in the multicountry model implies a tariff equivalent of respectively 48, 19, and 9 percent if the average elasticity is 5, 10, and 20.

The last three rows of Table 2 report the average error terms for interstate, interprovincial, and state-province trade. Particularly for the multicountry model they are close to zero. The average percentage difference between actual trade and predicted trade in the multicountry model is respectively 6,  $-2$ , and  $-4$  percent for interstate, interprovincial, and state-province trade. The largest error term in the two-country model is for interprovincial trade, where on average actual trade is 17 percent lower than predicted trade.<sup>24</sup>

<sup>23</sup> For example, for a highly homogeneous commodity such as silver bullion, Feenstra (1994) estimates a 42.9 elasticity of substitution among varieties imported from 15 different countries.

<sup>24</sup> The  $\bar{R}^2$  is respectively 0.43 and 0.45 for the two-country and multicountry model, which is somewhat lower than the 0.55 for the McCallum equation with unitary elasticities (last column Table 1). This is not a test of the theory though because McCallum's equation is not theoretically grounded. It also does not imply that multilateral resistance

### B. The Impact of the Border on Bilateral Trade

We now turn to the general-equilibrium comparative static implications of the estimated border barriers for bilateral trade flows. We will calculate the ratio of trade flows with border barriers to that under the borderless trade implied by our model estimates. Appendix B discusses how we compute the equilibrium after removing all border barriers while maintaining distance frictions. It turns out that we need to know the elasticity  $\sigma$  in order to solve for the free trade equilibrium. This is because the new income shares  $\theta_i$  depend on relative prices, which depend on  $\sigma$ . We set  $\sigma = 5$ , but we will show in the sensitivity analysis section that results are almost identical for other elasticities. The elasticity  $\sigma$  plays no role other than to affect the equilibrium income shares a little.

In what follows we define the “average” of trade variables and (transforms of the) multilateral resistance variables as the exponential of

does not matter; the dummies in McCallum's equation capture the average difference in multilateral resistance of states and provinces. With a higher estimate of internal distance, the  $\bar{R}^2$  from the structural model becomes quite close to that in the McCallum equation. It turns out though that internal distance has little effect on our key results (Section V).

TABLE 3—AVERAGE OF  $P^{1-\sigma}$ 

	US	Canada	ROW
Two-country model			
With border barrier (BB)	0.77 (0.03)	2.45 (0.12)	
Borderless trade (NB)	0.75 (0.03)	1.18 (0.01)	
Ratio (BB/NB)	1.02 (0.00)	2.08 (0.08)	
Multicountry model			
With border barrier (BB)	1.55 (0.01)	4.67 (0.09)	2.97 (0.07)
Borderless trade (NB)	1.39 (0.00)	1.91 (0.04)	1.54 (0.01)
Ratio (BB/NB)	1.12 (0.01)	2.44 (0.09)	1.93 (0.06)

Notes: The table reports the average of  $P_i^{\sigma-1}$ , where the average is defined as the exponential of the average logarithm. For the United States the average is taken over the 30 states in the sample, for Canada over the 10 provinces, and for ROW over the other 20 industrialized countries.

the average logarithm of these variables, consistent with McCallum (1995).<sup>25</sup>

The multilateral resistance variables are critical to understanding the impact of border barriers on bilateral trade and understanding what accounts for the McCallum border parameters. Defining regions in the United States, Canada, and ROW as three sets, Table 3 reports the average transform of multilateral resistance  $P^{\sigma-1}$  for regions in each of these sets. The results are shown both with the estimated border barrier and under borderless trade. As discussed in Section II, based on the model we would expect border barriers to lead to a larger increase of multilateral resistance in small countries than in large countries. This is exactly what we see in Table 3.  $P^{\sigma-1}$  rises by 12 percent for U.S. states, while it rises by a factor 2.44 for Canadian provinces.<sup>26</sup> The number is intermediate for ROW countries, whose size is also

intermediate. The Canadian border creates a barrier between provinces and most of its potential trading partners, while states face no border barriers with the rest of the large U.S. economy. Multilateral resistance therefore rises much more for provinces than for states.

Even under borderless trade  $P_i^{1-\sigma}$  is substantially higher for provinces than for states. Distances are somewhat larger on average between the United States and Canada than within them. This affects multilateral resistance for provinces more than for states as most potential trading partners of the provinces are outside their country, while for the states they are inside the country. This is again the result of the small size of the Canadian economy.

Table 3 reports the transforms  $P_i^{1-\sigma}$  of the multilateral resistance indices because they matter for trade levels. It is worthwhile pointing out that the  $P_i$  themselves, which are a measure of average trade barriers faced by regions, rise much less as a result of borders. For  $\sigma = 5$ ,  $P_i$  rises on average by 3 percent for states and 25 percent for Canadian provinces. For higher  $\sigma$  it is even smaller.

Table 4 reports the impact of border barriers on bilateral trade flows among and within each of the three sets of regions (US, CA, ROW). Size is controlled for by multiplying the bilateral trade numbers by  $y^W/(y_i y_j)$ . Letting a tilde denote borderless trade, the ratio of average trade between regions in sets  $h$  and  $k$  ( $h, k = \text{US, CA, ROW}$ ) with and without border barriers is

$$(23) \quad b_{hk}^{1-\sigma} \left( \frac{P_h^{\sigma-1}}{\bar{P}_h^{\sigma-1}} \right) \left( \frac{P_k^{\sigma-1}}{\bar{P}_k^{\sigma-1}} \right)$$

where  $P_h^{\sigma-1}$  refers to the average of regions in that set. We can therefore break down the impact of border barriers on trade into the impact of the bilateral border barrier and the impact of border barriers on multilateral resistance of regions in both sets. To the extent that border barriers raise average trade barriers faced by an importer and an exporter (multilateral resistance), it dampens the negative impact of the bilateral border barrier on trade between the two

<sup>25</sup> McCallum's border effect is the difference between the average logarithm of bilateral trade among regions in the same country and the average logarithm of bilateral trade of regions in different countries. This is converted back to levels by taking the exponential. Among a set of regions, bilateral trade between two regions is therefore considered to be average when the logarithm of bilateral trade is average within the set.

<sup>26</sup> Very little of the change in  $P^{\sigma-1}$  is associated with a change in income shares  $\theta_i$ . The change in income shares

alone would lower  $P^{\sigma-1}$  for Canadian provinces by 0.4 percent and raise it for states by 0.8 percent.

TABLE 4—IMPACT OF BORDER BARRIERS ON BILATERAL TRADE

	US-US	CA-CA	US-CA	US-ROW	CA-ROW	ROW-ROW
Two-country model						
Ratio BB/NB	1.05 (0.01)	4.31 (0.34)	0.41 (0.02)			
Due to bilateral resistance	1.0 (0.0)	1.0 (0.0)	0.19 (0.01)			
Due to multilateral resistance	1.05 (0.01)	4.31 (0.34)	2.13 (0.09)			
Multicountry model						
Ratio BB/NB	1.25 (0.02)	5.96 (0.42)	0.56 (0.03)	0.40 (0.01)	0.46 (0.01)	0.71 (0.02)
Due to bilateral resistance	1.0 (0.0)	1.0 (0.0)	0.20 (0.02)	0.19 (0.01)	0.10 (0.01)	0.19 (0.01)
Due to multilateral resistance	1.25 (0.02)	5.96 (0.42)	2.72 (0.12)	2.15 (0.09)	4.70 (0.31)	3.71 (0.25)

*Notes:* The table reports the ratio of trade with the estimated border barriers (BB) to that under borderless trade (NB). This ratio is broken down into the impact of border barriers on trade through bilateral resistance ( $t_{ij}^{1-\sigma}$ ) and through multilateral resistance ( $P_i^{\sigma-1}P_j^{\sigma-1}$ ).

countries. In what follows we will focus on the numbers for the more realistic multicountry model.

Implication 2 of the theory that cross-country trade barriers raise trade within a country more for small than for large countries is strongly confirmed in Table 4. The table reports a spectacular factor 6 increase in interprovincial trade due to borders, while interstate trade rises by only 25 percent. The larger increase in multilateral resistance of the provinces leads to a bigger drop in relative trade resistance  $t_{ij}/P_iP_j$  for trade within Canada than within the United States, explaining the large increase in interprovincial trade.

Table 4 also reports that borders reduce trade between the United States and Canada to a fraction 0.56 of that under borderless trade, or by 44 percent. Trade among ROW countries is reduced by 29 percent. The bilateral border barrier itself implies an 80-percent drop in trade between states and provinces, but increased multilateral resistance, particularly for provinces, raises state-province trade by a factor 2.72. While U.S. goods have become more expensive for Canada due to the border barrier, the goods of almost all trading partners of the provinces have become more expensive. This significantly moderates the negative impact on U.S.–Canada trade.

It may seem somewhat surprising that trade between the ROW countries drops somewhat

less than between the United States and Canada, particularly because the estimates in Table 2 imply a slightly lower U.S.–Canada border barrier. But it can be understood in the context of Implication 1 from the theory that border barriers have a bigger effect on trade between countries the larger their size. For the same border barriers, U.S.–Canada trade would have dropped much less if the United States were a much smaller country. This also explains why trade between the United States and the ROW countries drops somewhat more than between the United States and Canada. Canada is even smaller than the average ROW country. Based on size alone one would expect trade between Canada and the ROW countries to drop less than between Canada and the United States, but this is not the case as a result of the higher trade barrier between Canada and the ROW countries.

### C. Intranational Trade Relative to International Trade

McCallum aimed to measure the impact of borders on intranational trade (within Canada) to international trade (between the United States and Canada). In this subsection we will show that the large McCallum border parameter for Canada is due to a combination of (i) the relative small size of the Canadian economy and (ii) omitted variables bias.



TABLE 5—IMPACT BORDER ON INTRANATIONAL TRADE RELATIVE TO INTERNATIONAL TRADE

	Two-country model		Multicountry model	
	Canada	US	Canada	US
Theoretically consistent estimate	10.5 (1.16)	2.56 (0.13)	10.7 (1.06)	2.24 (0.12)
McCallum parameter implied by theory	16.5 (1.63)	1.64 (0.09)	14.8 (1.32)	1.63 (0.10)

*Notes:* The first row of the table reports the theoretically consistent estimate of the impact of border barriers on intranational trade relative to international trade for both Canada and the United States. The second row reports the McCallum border parameter implied by the model, which provides a biased estimate of the impact of borders on the ratio of intranational to international trade.

The impact of border barriers on intranational trade relative to international trade follows immediately from Table 4 and is reported in the first row of Table 5. The multicountry model implies that national borders lead to trade between provinces that is a factor 10.7 larger than between states and provinces. In contrast, border barriers raise trade between states by only a factor 2.24 relative to trade between states and provinces. This is exactly as anticipated by Implication 3 of the theory. It is the result of the relatively small size of Canada, leading to a factor 6 increase in trade between the provinces. The small change in trade between U.S. states leads to a correspondingly much smaller increase in intranational to international trade for the United States.

This is only part of the explanation for the large McCallum border parameter for Canada. The other part is the result of omitted variables bias in two distinct senses: estimation and computation. By estimation bias we mean the ordinary econometric omitted variables bias. By computation bias we mean the erroneous comparative statics which arise from a reduced-form calculation which omits terms. In order to analyze the omitted variables bias, rewrite the theoretical gravity equation as

$$(24) \quad \ln x_{ij} = k + \ln y_i + \ln y_j + \rho(1 - \sigma)\ln d_{ij} + R_{ij} + \varepsilon_{ij}$$

where

$$R_{ij} = (1 - \sigma)\ln b_{ij} - (1 - \sigma)\ln P_i - (1 - \sigma)\ln P_j.$$

$R_{ij}$  measures the sum of all trade resistance terms with the exception of the bilateral distance term. McCallum estimated (24), but replaced  $R_{ij}$  with a dummy variable that is 1 for interprovincial trade and 0 for state–province trade. In the absence of the multilateral resistance terms this would yield unbiased estimates of  $(1 - \sigma)\rho$  and  $(1 - \sigma)\ln b$ . But since the omitted multilateral resistance terms are correlated with both distance and the border dummy, McCallum's regression does not yield an unbiased estimate of either  $(1 - \sigma)\rho$  or  $(1 - \sigma)\ln b$ . Next, consider computation bias. Assume for the moment that McCallum had correctly estimated the parameter  $(1 - \sigma)\rho$  multiplying bilateral distance. In that case McCallum's border effect can still not be interpreted as the effect of borders on the ratio of interprovincial trade relative to state–province trade. In the context of the theory we can then interpret McCallum's border parameter for Canada as an estimator of the average of  $R_{ij}$  for interprovincial trade minus the average for state–province trade, and similarly for the United States.<sup>27</sup> Taking the exponential for comparison with McCallum's headline number, we get (following the notation of Section I)

$$(25) \quad \text{Border}_{\text{Canada}} = (b_{\text{US,CA}})^{\sigma-1} \frac{P_{\text{CA}}^{\sigma-1}}{P_{\text{US}}^{\sigma-1}}.$$

<sup>27</sup> We will take the average over all trade pairs, even though for a few state–province pairs and state–state pairs no trade data exist. Taking the average only over pairs for which trade data exist leads to almost identical numbers.

Similarly, for the United States we get

$$(26) \quad \text{Border}_{US} = (b_{US,CA})^{\sigma-1} \frac{P_{US}^{\sigma-1}}{P_{CA}^{\sigma-1}}.$$

The theoretical McCallum border parameters implied by (25)–(26) are reported in the second row of Table 5. For the multicountry model the border parameters are 14.8 for Canada and 1.63 for the United States. This corresponds closely to the 14.2 and 1.62 parameters reported in the last column of Table 1 when estimating McCallum's regression with unitary income coefficients. The much higher Canadian (transform of) multilateral resistance term,  $P_{CA}^{\sigma-1}$ , than the U.S. multilateral resistance term,  $P_{US}^{\sigma-1}$ , blows up the border effect for Canada, while dampening it with the same factor for the United States.

A comparison of rows 1 and 2 of Table 5 shows that McCallum's measure for Canada overstates our consistent estimate of the impact of borders on intranational trade relative to international trade. The reason is that in the correct measure of the impact of borders on intranational relative to international trade, the multilateral resistance terms in (25) and (26) are replaced by the ratio of multilateral resistance with border barriers relative to that without border barriers; the comparative static experiment of taking away the borders must include its effect on multilateral resistance. McCallum's measure would have implied a border parameter larger than 1 for Canada even in the absence of border barriers because of the higher multilateral resistance of provinces than states due to distance alone.

The difference between the two rows in Table 5 illustrates the omitted variables bias in McCallum's results due to comparative statics alone as we have used the parameter estimates from the theoretical model to compute (25) and (26). It turns out that almost all of the bias resulting from omitted variables is associated with comparative statics as opposed to a biased estimate of the distance coefficient  $(1 - \sigma)\rho$ . If we reestimate McCallum's regression in the last column of Table 1 after imposing the distance coefficient obtained from estimating the theoretical gravity equation, the resulting McCallum border coefficient changes only slightly from 14.2 to 14.7.

There is also a literature that has estimated the impact of borders on domestic trade relative

to international trade for a wide range of other OECD countries. This literature is based on McCallum-type regressions, often with atheoretical remoteness variables added, using international trade data combined with an estimate of total domestic trade in each of the countries. The findings from this literature can be compared to the theory. Based on the estimated multicountry model, international trade among the ROW countries drops to a fraction 0.71 of that under free trade, while intranational trade rises on average by a factor 3.8. This implies a factor 5.4 ( $3.8/0.71$ ) increase in intranational trade relative to international trade, which falls within the range of estimates of about 2.5 to 10 that have been reported in the empirical literature. For example, Helliwell (1998) reports a factor 5.7 for 1992 data, estimating (3) with the atheoretical remoteness variables (3) included. Our findings suggest that the trade home bias reported in this literature is primarily a result of the large increase in intranational trade. International trade drops by only 29 percent as a result of borders. Intranational trade rises so much for the same reason that interprovincial trade rises so much in Canada. Most countries are relatively small as a fraction of the world economy.

## V. Sensitivity Analysis

Table 6 reports the results from a variety of sensitivity analysis. In order to save space we report only the key variables of interest, the impact of borders on trade, and the McCallum border parameter. For comparison we report in column (i) results from the base regression.

Column (ii) assumes a higher elasticity of substitution  $\sigma = 10$  (in the benchmark  $\sigma = 5$ ). This has no impact on the nonlinear least-squares estimator itself but, as discussed in Section IV, subsection B, it affects somewhat the equilibrium when removing the border barrier. It is clear though from column (ii) that the difference is negligible. The same is the case when we lower  $\sigma$  to 2 or raise it 20 (not reported). The insensitivity to  $\sigma$  is encouraging as there is little agreement about the precise magnitude of this parameter.

Columns (iii) and (iv) report results when we respectively double and halve our measure of distance internal to states, provinces, and the other industrialized countries. While we have

TABLE 6—SENSITIVITY ANALYSIS

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)	(ix)
Two-country model									
Trade (BB/NB)									
US–US	1.05 (0.01)	1.05 (0.01)	1.05 (0.01)	1.05 (0.01)	1.04 (0.01)	1.04 (0.00)	1.03 (0.00)	1.05 (0.01)	
CA–CA	4.31 (0.34)	4.26 (0.34)	4.31 (0.30)	4.41 (0.39)	3.92 (0.53)	3.82 (0.29)	4.37 (0.38)	3.76 (0.23)	
US–CA	0.41 (0.02)	0.41 (0.02)	0.41 (0.02)	0.41 (0.02)	0.50 (0.03)	0.50 (0.02)	0.54 (0.02)	0.43 (0.02)	
McCallum parameter									
US	1.64 (0.09)	1.64 (0.09)	1.58 (0.08)	1.70 (0.09)	1.37 (0.11)	1.03 (0.05)	1.05 (0.06)	1.33 (0.07)	
CA	16.5 (1.63)	16.5 (1.63)	17.1 (1.64)	16.1 (1.65)	12.2 (1.93)	15.4 (1.50)	14.9 (1.63)	16.4 (1.56)	
Multicountry model									
Trade (BB/NB)									
US–US	1.25 (0.02)	1.26 (0.02)	1.21 (0.02)	1.29 (0.02)	1.22 (0.04)	1.19 (0.02)	1.18 (0.02)		1.14 (0.02)
CA–CA	5.96 (0.42)	5.93 (0.42)	5.90 (0.37)	6.21 (0.49)	5.07 (0.66)	5.03 (0.36)	5.21 (0.40)		4.44 (0.39)
US–CA	0.56 (0.03)	0.56 (0.03)	0.54 (0.03)	0.57 (0.03)	0.63 (0.05)	0.62 (0.03)	0.63 (0.03)		0.51 (0.03)
ROW–ROW	0.71 (0.02)	0.70 (0.02)	0.71 (0.02)	0.72 (0.02)	0.83 (0.06)	0.50 (0.02)	0.87 (0.04)		0.76 (0.04)
McCallum parameter									
US	1.63 (0.10)	1.63 (0.10)	1.56 (0.09)	1.69 (0.11)	1.38 (0.12)	1.13 (0.06)	1.19 (0.08)		1.56 (0.09)
CA	14.8 (1.32)	14.8 (1.32)	15.3 (1.34)	14.5 (1.33)	11.1 (1.57)	13.6 (1.37)	12.9 (1.29)		12.4 (1.25)

*Notes:* The table reports sensitivity analysis with regards to the ratio of trade with border barriers to trade without border barriers and with regards to the McCallum border parameters implied by the model. Column (i) repeats results from the benchmark regression. Column (ii) assumes  $\sigma = 10$  (in the benchmark  $\sigma = 5$ ). Columns (iii) and (iv) report results when respectively doubling and halving distances internal to regions and countries. Column (v) reports results based on a regression that does not use interstate data. Columns (vi) and (vii) report results when income  $y$  is replaced by  $x^\alpha y$  with  $x$  respectively income  $y$  and per capita income  $y/N$ .  $x^\alpha$  represents the fraction spent on tradables in a region or country. Column (viii) reports for the two-country case results based on fixed-effects estimation. The final column reports for the multicountry case results when minimizing the sum of all squared error terms, including those involving ROW countries.

used the proxy by Wei (1996) that has been commonly used in the literature, this is only a rough estimate. The correct measure depends a lot on a region's geography.<sup>28</sup> Helliwell (1998) finds that results are very sensitive to internal distance when applying a McCallum gravity equation to international and intranational trade of OECD countries. Halving internal distances reduces the border effect by about half, while doubling internal distances more than doubles it. In contrast, columns (iii) and (iv) of Table 6 show that doubling or halving internal distances

has very little effect on our results. A big advantage of the U.S.–Canada data set is that the intranational trade data are for interstate trade and interprovincial trade. It is relatively easy to measure distances between states and between provinces. We do not use data on trade internal to states and provinces, for which distance is hard to measure. In our regression internal distance matters only to the extent that it affects multilateral resistance.

Column (v) reports results when we do not use data on interstate trade. The reason for doing so is that McCallum did not use interstate trade data and we do not want to leave the impression that the interstate data set is critical to our findings. The results reported in column (v) are somewhat different from those based on

<sup>28</sup> For example, it is possible that most trade takes place within one industrial area, in which case the appropriate measure of internal distance could be close to zero.

the benchmark regression, but they are qualitatively identical. The reported impact of borders on trade levels is not statistically different from that reported under the benchmark regression. If anything it reinforces our key finding of a moderate impact of borders on international trade. The multicountry model tells us that U.S.–Canada trade is reduced by 37 percent as a result of border barriers, while trade among other industrialized countries is reduced by 17 percent.

Columns (vi) and (vii) report results when allowing for nonunitary income elasticities. Anderson (1979) allowed for nonunitary income elasticities by modeling the fraction spent on tradable goods. We have used total GDP, from hereon  $Y$ , as an estimate of tradables output  $y$  in the model. But in reality GDP also includes nontradables. Anderson (1979) assumed that a fraction  $\phi$  of total income is spent on tradables, so that spending on tradables is  $\phi Y$ . Because of balanced trade, output of tradables must also be  $\phi Y$ . Anderson allowed  $\phi$  to be a function of both  $Y$  and  $N$  (population). Column (vi) reports results when  $\phi = Y^\alpha$ , so that bilateral trade is equal to  $Y_i^{1+\alpha} Y_j^{1+\alpha}$  times the trade resistance terms. While this introduces nonunitary income elasticities, as in McCallum (1995), we should stress that there is no clear theoretical foundation for specifying the fraction spent on tradables as  $Y^\alpha$ . Column (vii) reports results when  $\phi = (Y/N)^\alpha$ . In that case bilateral trade is equal to  $Y_i^{1+\alpha} Y_j^{1+\alpha} N_i^{-\alpha} N_j^{-\alpha}$  times the trade resistance terms. This assumption has somewhat more solid theoretical grounding. The well-known Balassa-Samuelson effect tells us that regions with higher productivity in the tradables sector will have a higher relative price of nontradables, which should raise the fraction spent on tradables. To the extent that  $Y/N$  proxies for productivity in the tradables sector, one might expect  $\alpha$  to be positive. This is indeed what we find in the estimation.<sup>29</sup> The results reported in Table 6, while they change somewhat from the base regression, are still qualitatively the same. If anything, we find that the impact of borders on interna-

tional trade is reduced somewhat further when  $\phi$  is a function of  $Y/N$ .

Column (viii) reports results based on fixed-effects estimation, replacing the transforms of multilateral resistance terms with region dummies. This is feasible only in the two-country model. It again has very little effect on the trade results. It is worthwhile pointing out that while the border parameter  $(1 - \sigma)\ln b$  does not change much, the distance parameter  $(1 - \sigma)\rho$  drops from  $-0.79$  in the structural model to  $-1.25$ . This suggests that internal distances are larger than assumed when estimating the structural model. Raising the benchmark internal distances leads to a more negative estimate of  $(1 - \sigma)\rho$  in the structural model. It also raises the adjusted  $R^2$  and leads to a higher correlation between the region dummies and the theoretical multilateral resistance terms. Fixed costs of transportation may provide a justification for higher internal distances.

As a final form of sensitivity analysis, in the last column of Table 6 we report results for the multicountry model when we estimate all parameters by minimizing the sum of all squared residuals, including the ROW–ROW, US–ROW, and CA–ROW residuals. As discussed above, an important reason for not doing so in the first place is that this estimation procedure has weaker finite sample properties, primarily because there are relatively few US–ROW and CA–ROW observations. One implausible finding, not reported in Table 6, is that the US–ROW barrier now becomes lower than the US–CA border barrier, with  $(1 - \sigma)\ln b$  of respectively  $-0.88$  and  $-1.48$ . Nonetheless the results reported in Table 6, the impact of borders on trade and the McCallum parameters, remain quite close to those under the benchmark regression.

Overall we can therefore conclude that the results from the benchmark regression are robust to a wide range of sensitivity analysis.

## VI. Conclusion

Although commonly estimated gravity equations generally have a very good fit to the data, we have shown that they are not theoretically grounded. This leads to biased estimation, incorrect comparative statics analysis, and generally a lack of understanding of what is driving the results. In this paper we have developed a

<sup>29</sup> For example, in the multicountry model we find  $\alpha = 1.07$ , and similar for the two-country model. When  $\psi = Y^\alpha$ , we find that  $\alpha$  is about 0.3 for both models.

method that consistently and efficiently estimates a theoretical gravity equation. We have applied the method to solve the border puzzle. We find that borders reduce bilateral national trade levels by plausible though substantial magnitudes. The results of previous studies that imply enormous border effects are explicable in terms of our model: (i) they considered the effect of the border on the ratio *intranational* to *international* trade; (ii) this border effect is inherently large for small countries; and (iii) omitted variables biased the estimated border effect upwards. The approach can easily be applied to determine the effect of many other institutions on bilateral trade flows.

The methodology we developed is based on existing gravity theory, which makes a variety of simplifying assumptions that need to be generalized in future research. One important drawback of the existing theory is that all countries import all varieties of a good from all countries that produce the good. Haveman and Hummels (1999) find that this violates available evidence that many goods are imported from only one or two producers. They suggest extensions of the model that involve homogeneous goods, differences in preferences, and fixed costs.<sup>30</sup> Another limitation of the model is the assumption of an endowment economy. Border barriers can also affect trade through their impact on the production structure. Hillberry and Hummels (2002) discuss the effect of borders on production location of intermediate goods producers, while Kei-Mu Yi (2003) analyzes the effect of tariffs on trade in the context of vertical specialization. We believe that these are all fruitful directions for future research. We suspect though that the key aspect of the gravity model, the dependence of trade on bilateral and multilateral resistance, will hold up under a wide range of generalizations.

#### APPENDIX A: THE DATA

The paper uses data on trade, distances, GDP, and population for states, provinces, and 20 other industrialized countries. Before turning to a detailed discussion of the trade data, we describe the sources of the other data first. Great

circle distances are computed using the longitude and latitude of states, provinces, and other countries, obtained from the web site <http://www.indo.com/distance/>. GDP data are from Statistics Canada for the provinces, the Bureau of Economic Analysis for the states, and from the IMF's International Financial Statistics for the 20 other industrialized countries. Population data are from the Bureau of the Census for the states. For provinces and the other industrialized countries the source is the same as for GDP.

The paper combines four trade data sets: interprovincial merchandise trade from the Input-Output Division of Statistics Canada; province-trade merchandise trade from the International Trade Division of Statistics Canada; interstate commodity flows from the Commodity Flow Survey by the U.S. Census; and merchandise trade among the other industrialized countries from the IMF's Direction of Trade Statistics. It should be said from the outset that these data sets use concepts that are different from each other and adjustments are necessary in order to make them more compatible.

McCallum (1995) combines the first two data sources listed above. The IO Division, which collects the interprovincial trade data, also collects data on trade between each province and the rest of the world. Those data net out exports and imports that are en route to and from other provinces. The trade data from the IT Division, on the other hand, are based on customs data, for which the original source and final destination of shipments are not known. There is a nice discussion of these issues in Anderson and Smith (1999a, b). Because the data of the IO Division are more reliable, McCallum multiplies the state-province trade flows from the IT Division by the ratio of trade of each province with the rest of the world from the IO and IT sources. Helliwell (1998) makes the same adjustment, but at the more detailed level of 27 individual industries. In this paper we use the data with the more detailed adjustment by Helliwell. Data are available for all 90 interprovincial pairs, while they are available for 589 of the 600 state-province pairs (bilateral flows between 10 provinces and 30 states).

For the year 1993 the Commodity Flow Survey (CFS) by the U.S. Census Bureau provides data on within-state and cross-state shipments. The data set and methodology are described in

<sup>30</sup> Evans (2000a) and Hillberry (2001) analyze the impact of borders when there are fixed costs.

detail in the Bureau of Transportation web site <http://www.bts.gov/ntda/cfs/>. The data consists of shipments by domestic establishments in manufacturing, wholesale, mining, and selected retail establishments. The survey covers 200,000 representative establishments out of a total of about 800,000. Four times per year, during a two-week period, the surveyed establishments were asked to report the value and volume of shipments, as well as the origin and destination addresses. There are three important differences between these shipments data and the merchandise trade data. First, while merchandise trade data measure only shipments from source to final user, the commodity flow data include all shipments. For example, a product may be shipped from a manufacturing plant to a warehouse and from there to a retailer. Second, goods that are intended for exports, but are first shipped domestically (e.g., to a harbor), are included in domestic shipments. Similarly, goods that are imported are measured once they are shipped from the port of entry to another domestic destination. Third, while the Commodity Flow Survey provides extensive coverage of the manufacturing sector, which is by far the most important goods-producing sector, it excludes agriculture and part of mining.

As a result of these inconsistencies, an adjustment is made to the CFS data. The CFS data are scaled down by the ratio of total domestic merchandise trade to total domestic shipments from the CFS. Following Helliwell (1997, 1998) and Wei (1996), total domestic merchandise trade is approximated as gross output in mostly goods-producing sectors, minus merchandise exports. The goods-producing sectors are defined as the sum of agriculture, mining, and manufacturing. Using this methodology, total domestic U.S. merchandise trade was \$3,025 billion in 1993, while shipments in the CFS total to \$5,846 billion. The CFS data are therefore scaled down by  $3,025/5,846$ . Of the total 870 trade pairs among the 30 states in the sample, data are available for 832 pairs.

There are several reasons to believe that the adjusted U.S. trade data are not so bad. First, for both the two-country model and the multicountry model the estimated model coefficients are similar when estimating the model without the use of interstate data (an experiment considered in sensitivity analysis), and the difference is not statistically significant. Second, the average

squared error term is smaller for the interstate data than for the interprovincial data, respectively 0.48 and 1.40 in the multicountry model. This is not the result of the dominance of interstate trade data. When estimating the multicountry model without interstate trade data, the average squared error term of interprovincial data remains 1.44. Consistent with that, Table 1 also reports a higher  $\bar{R}^2$  when estimating McCallum's equation for the United States (0.86) than for Canada (0.77).

We do not pretend to have solved all measurement problems with the adjustment factor applied to the U.S. commodity flow data. As discussed above, the data used in the original McCallum study are not without measurement problems either, with even much larger adjustment factors applied to the original state-province data. These data nonetheless remain by far the best currently available to study the impact of borders on trade. Moreover, as reported in the sensitivity analysis, the key findings of this paper do not rely on the U.S. trade data set.

#### APPENDIX B: SOLUTION TO THE BORDERLESS TRADE EQUILIBRIUM

To solve for the borderless trade equilibrium of the model we set  $b_{ij} = 1 \forall i, j$ . When solving for the new equilibrium prices  $p_i$ , or alternatively for the price indices  $P_i$  from (12), we need to take into account that the income shares  $\theta_i$  change. Let a 1 superscript denote the "no borders" equilibrium with a 0 superscript denote the estimated model with borders present. Since quantities produced are assumed fixed,  $y_i^1 = (p_i^1/p_i^0)y_i^0$ . We observe  $y_i^0$  and have solved for  $p_i^0$ . The new income shares  $\theta_i^1$  then become functions of the new prices  $p_i^1$  that are being solved.

While equilibrium trade flows with border barriers can be computed using only the estimated trade cost parameters  $(1 - \sigma)\ln b_{ij}$  and  $(1 - \sigma)\rho$ , we need to know the elasticity  $\sigma$  in order to compute equilibrium trade flows under borderless trade. In the equilibrium with border barriers we can solve for  $p_i^{1-\sigma} \forall i$  as a function of the estimated trade cost parameters. This determines the equilibrium  $P_i^{1-\sigma} \forall i$ , which determines equilibrium trade flows. But in the borderless trade equilibrium the  $p_i^{1-\sigma}$  also depend on the income shares  $\theta_i$ , which are func-

tions of the prices  $p_i$ . We therefore need to know  $\sigma$  in order to solve for  $p_i^{1-\sigma}$ .

## REFERENCES

- Anderson, James E.** "A Theoretical Foundation for the Gravity Equation." *American Economic Review*, March 1979, 69(1), pp. 106–16.
- Anderson, James E. and Marcouiller, Douglas.** "Insecurity and the Pattern of Trade: An Empirical Investigation." *Review of Economics and Statistics*, May 2002, 84(2), pp. 345–52.
- Anderson, Michael A. and Smith, Stephen L. S.** "Canadian Provinces in World Trade: Engagement and Detachment." *Canadian Journal of Economics*, February 1999a, 32(1), pp. 23–37.
- . "Do National Borders Really Matter? Canada–U.S. Regional Trade Reconsidered." *Review of International Economics*, May 1999b, 7(2), pp. 219–27.
- Bergstrand, Jeffrey H.** "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence." *Review of Economics and Statistics*, August 1985, 67(3), pp. 474–81.
- . "The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International Trade." *Review of Economics and Statistics*, February 1989, 71(1), pp. 143–53.
- . "The Heckscher-Ohlin-Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade." *Economic Journal*, December 1990, 100(403), pp. 1216–29.
- Bernard, Andrew and Wagner, Joachim.** "Export Entry and Exit by German Firms." *Weltwirtschaftliches Archiv*, 2001, 137(1), pp. 105–23.
- Brenton, Paul; Di Mauro, Francesca and Lucke, Matthias.** "Economic Integration and FDI: An Empirical Analysis of Foreign Investment in the EU and in Central and Eastern Europe." *Empirica*, 1999, 26(2), pp. 95–121.
- Chen, Natalie A.** "Intra-national Versus International Trade in the European Union: Why Do National Borders Matter?" Mimeo, ECARES, Université Libre de Bruxelles, April 2000.
- Deardorff, Alan V.** "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in J. A. Frankel, ed., *The regionalization of the world economy*. Chicago: University of Chicago Press, 1998, pp. 7–22.
- Eichengreen, Barry and Irwin, Douglas A.** "The Role of History in Bilateral Trade Flows," in J. A. Frankel, ed., *The regionalization of the world economy*. Chicago: University of Chicago Press, 1998, pp. 33–57.
- Evans, Carolyn L.** "National Border Effects and Heterogeneous Fixed Costs of International Trade." Mimeo, Federal Reserve Bank of New York, 2000a.
- . "The Economic Significance of National Border Effects." Mimeo, Federal Reserve Bank of New York, 2000b.
- Feenstra, Robert C.** "New Product Varieties and the Measurement of International Prices." *American Economic Review*, March 1994, 84(1), pp. 157–77.
- . "U.S. Exports, 1972–1994: With State Exports and Other U.S. Data." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5990, April 1997.
- Frankel, Jeffrey; Stein, Ernesto and Wei, Shang-Jin.** "Continental Trading Blocs: Are They Natural or Supernatural?" in J. A. Frankel, ed., *The regionalization of the world economy*. Chicago: University of Chicago Press, 1998, pp. 91–113.
- Grossman, Gene.** "Comment on Alan V. Deardorff, Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" in J. A. Frankel, ed., *The regionalization of the world economy*. Chicago: University of Chicago Press, 1998, pp. 29–31.
- Haveman, Jon and Hummels, David.** "Alternative Hypotheses and the Volume of Trade: Evidence on the Extent of Specialization." Mimeo, Krannert School of Management, Purdue University, 1999.
- Head, Keith and Ries, John.** "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.–Canada Trade." *American Economic Review*, September 2001, 91(4), pp. 858–76.
- Helliwell, John F.** "Do National Borders Matter for Quebec's Trade?" *Canadian Journal of Economics*, August 1996, 29(3), pp. 507–22.
- . "National Borders, Trade and Migration." *Pacific Economic Review*, October 1997, 2(3), pp. 165–85.
- . *How much do national borders matter?* Washington, DC: Brookings Institution, 1998.

- Helliwell, John F. and McCallum, John.** "National Borders Still Matter for Trade." *Policy Options*, July/August 1995, 16, pp. 44–48.
- Helliwell, John F. and Verdier, Genevieve.** "Measuring Internal Trade Distances: A New Method Applied to Estimate Provincial Border Effects in Canada." *Canadian Journal of Economics*, November 2001, 34(4), pp. 1024–41.
- Hillberry, Russell.** "Regional Trade and the Medicine Line: The National Border Effect in U.S. Commodity Flow Data." *Journal of Borderlands Studies*, Fall 1998, 8(2), pp. 1–17.
- . "Explaining the Border Effect: What Can We Learn from Disaggregated Commodity Flow Data?" Mimeo, International Trade Commission, 1999.
- . "Aggregation Bias, Compositional Change and the Border Effect." Mimeo, Federal Trade Commission, 2001.
- Hillberry, Russell and Hummels, David.** "Explaining Home Bias in Consumption: The Role of Intermediate Input Trade." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 9020, 2002.
- Hummels, David.** "Toward a Geography of Trade Costs." Mimeo, Krannert School of Management, Purdue University, 1999.
- McCallum, John.** "National Borders Matter: Canada–U.S. Regional Trade Patterns." *American Economic Review*, June 1995, 85(3), pp. 615–23.
- Messinger, Hans.** "Interprovincial Trade Flows of Goods and Services." *Canadian Economic Observer*, 1993, pp. 3.8–3.14.
- Obstfeld, Maurice and Rogoff, Kenneth.** "The Six Major Puzzles in International Macroeconomics. Is There a Common Cause?" in Ben S. Bernanke and Kenneth Rogoff, eds., *NBER macroeconomics annual 2000*. Cambridge, MA: MIT Press, 2001, pp. 339–90.
- Portes, Richard and Rey, Helene.** "The Euro and International Equity Flows." *Journal of the Japanese and International Economies*, December 1998, 12(4), pp. 406–23.
- Ravenstein, Ernst G.** "The Laws of Migration." *Journal of the Royal Statistical Society*, 1885, 48(2), pp. 167–235.
- Roberts, Mark J. and Tybout, James R.** "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs." *American Economic Review*, September 1997, 87(4), pp. 545–64.
- Rose, Andrew K.** "One Money, One Market: The Effect of Common Currencies on Trade." *Economic Policy*, April 2000, 15(30), pp. 7–33.
- Rose, Andrew K. and van Wincoop, Eric.** "National Money as a Barrier to Trade: The Real Case for Currency Union." *American Economic Review*, May 2001 (*Papers and Proceedings*), 91(2), pp. 386–90.
- Trefler, Daniel.** "The Case of the Missing Trade and Other Mysteries." *American Economic Review*, December 1995, 85(5), pp. 1029–46.
- Wei, Shang-Jin.** "Intra-national versus International Trade: How Stubborn Are Nations in Global Integration?" National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5531, 1996.
- Wolf, Holger.** "(Why) Do Borders Matter For Trade?" in G. D. Hess and E. van Wincoop, eds., *Intranational macroeconomics*. Cambridge: Cambridge University Press, 2000, pp. 112–28.
- Yi, Kei-Mu.** "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy*, 2003 (forthcoming).