Trade, wages, and productivity

Kristian Behrens* Giordano Mion† Yasusada Murata‡ Jens Südekum§

October 4, 2010

Abstract

We develop a new general equilibrium model of trade in which wages, productivity, and markups are endogenous and need not be equalized across asymmetric regions. Using Canada-US regional data, we structurally estimate a gravity equation for bilateral trade subject to general equilibrium conditions. The estimated model is then used to simulate the impacts of removing all trade barriers generated by the Canada-US border. The counterfactual analysis reveals that ignoring endogenous wage and productivity responses to trade integration can lead to substantial biases in predicted trade flows, and that these biases systematically depend on the origin and the destination of the flows. In particular, we find that Canada-US flows are dampened by 34% due to relative wage increases in Canada, whereas US-Canada flows are reduced by 56% due to productivity increases driven by firm selection in Canada. We also find that the impacts on regional economic aggregates differ both between and within countries. For example, our results suggest that trade integration reduces markups in Canadian provinces by 4.65% to 14.8% against 0.07% to 2.56% in US states.

Keywords: firm heterogeneity; gravity equation; endogenous markups; monopolistic competition; general equilibrium

JEL Classification: F12; F15; F17

*Canada Research Chair, Department of Economics, Université du Québec à Montréal (UQAM), Canada; CIRPÉE, Canada; and CEPR, London. E-mail: behrens.kristian@uqam.ca.

†Department of Geography and Environment, London School of Economics (LSE), UK; National Bank of Belgium (NBB), Belgium; CEP, London; and CEPR, London. E-mail: g.mion@lse.ac.uk

‡Advanced Research Institute for the Sciences and Humanities (ARISH), Nihon University, Tokyo, Japan; and Nihon University Population Research Institute (NUPRI), Japan. E-mail: murata.yasusada@nihon-u.ac.jp

§Mercator School of Management, Universität Duisburg-Essen, Germany; Ruhr Graduate School of Economics, Germany; and IZA, Germany. E-mail: jens.suedekum@uni-due.de

1 Introduction

Recent theories of international trade predict that countries with larger domestic markets have higher wages (Krugman, 1980) and higher aggregate productivity (Melitz, 2003). Both wages and productivity respond to trade integration as it affects firms' profit opportunities. Other things being equal, larger markets also have lower price-cost margins, and more exposure to trade reduces markups due to pro-competitive effects (Krugman, 1979). All of these features, as well as changes in consumption diversity, are crucial for assessing the gains from trade (Feenstra, 2010). Various studies have confronted these issues separately, but there has so far been no comprehensive analysis that incorporates all of these aspects into a single framework.

We develop and quantify a new general equilibrium model of trade in which both wages and productivity are endogenous and need not be equalized across asymmetric regions. To account for firms' entry decisions and endogenous varieties, as well as markups that can vary depending on firms, regions, and trade costs, we consider a monopolistic competition model with firm heterogeneity and variable demand elasticity. The model delivers a *gravity equation system* — a gravity equation for bilateral trade *and* general equilibrium conditions that determine wages and productivity. Using Canada-US regional data, we structurally estimate the model's key parameters, namely trade frictions and region-specific technological possibilities. The estimated model is then used to simulate the impacts of removing all trade barriers generated by the Canada-US border. This counterfactual analysis relates to the vast literature on border effects (McCallum, 1995; Anderson and van Wincoop, 2003) and allows us to quantify the impacts of trade integration on various regional economic aggregates.

Our contribution is threefold. First, we develop a framework that accommodates the key qualitative features of the recent workhorse trade models. In particular, we show that, other things being equal, regions with larger population size or better technological possibilities have higher wages and higher aggregate productivity. Consumers in those regions are less exposed to market power, as measured by the (expenditure share) weighted average of firm-level markups. They also enjoy greater consumption diversity and higher welfare. Moreover, we show that trade integration induces regional convergence in wages, productivity, weighted averages of markups, consumption diversity, and welfare, thereby building intuition for the counterfactual experiment.

Second, structural estimation of our gravity equation system allows us to obtain not only trade friction parameters but also region-specific technological possibilities. The latter are related to regional productivity distributions and are not observable from the data. However, they appear in both the gravity equation and the general equilibrium conditions that encompass labor market clearing, zero profit, and trade balance. Estimating the gravity equation system then reveals their values that are consistent with equilibrium wages and productivity. Assessing the fit of our estimated model, we find that it behaves well and can replicate several empirical facts, both at the regional and at the firm level.

Last, we run a counterfactual experiment. In the estimated model, we remove the Canada-US border and consider a world where trade costs depend only on distance. This allows us to compute a series of *bilateral border effects* that convey how trade flows between any two regions would be affected by such a trade integration. These effects can be decomposed into a 'pure' border effect, relative and absolute wage effects, and a selection (or productivity) effect, thus highlighting each channel that affects trade flows. We find that disregarding endogenous wage and productivity responses can lead to substantial biases in predicted trade flows, and that these biases systematically depend on the origin and the destination of the flows. In particular, endogenous wage responses would dampen the predicted expansion of Canadian exports to the US by 34% on average when compared to a fixed-wage case. The reason is that the border removal would increase Canadian relative wages, thereby raising production costs and reducing exports. In contrast, endogenous productivity responses would reduce the predicted expansion of US exports to Canada by 56% when compared to a fixed-productivity case because tougher selection of low productivity firms in Canada, driven by the border removal, would make market penetration more difficult.

Our counterfactual experiment further reveals that the impacts of trade integration on regional economic aggregates differ both between and within countries. We find that the border removal would increase Canadian relative wages by 2.45% to 8.78% and enhance Canadian average labor productivity by 6.78%, whereas US average labor productivity would rise by just 0.32%.[1] Canadian weighted averages of markups would fall by 4.65% to 14.8%, whereas the corresponding fall in the US would be 0.07% to 2.56%, thereby suggesting the magnitude of the pro-competitive effects of trade integration. Given these figures, welfare gains would naturally be larger in Canada than in the US, ranging from 4.9% to 17.4%, against 0.08% to 2.26%. Investigating what drives the regional variations, we find that *geography and size matter*: less populous regions closer to the border would be affected more by the border removal.

This paper extends the two strands of literature on firm heterogeneity and gravity equations. Melitz (2003) introduces productivity differences across firms into the monopolistic competition model by Krugman (1980). His model explains how trade liberalization generates aggregate productivity gains by intensifying firm selection and by reallocating market shares from less to more efficient producers.[2] Despite its merits, that framework has two restrictive features: factor price equalization (FPE) and constant elasticity of substitution (CES). Though analytically convenient,

[1] Our predicted Canadian productivity gain is roughly similar to the 7.4% increase estimated by Trefler (2004, pp.880-881) for the Canada-US Free Trade Agreement. It is worth emphasizing that he attributes the sources of these productivity gains to "market share shifts favoring high-productivity plants. Such share shifting would come about from the growth of high-productivity plants and the demise and/or exit of low-productivity plants [...]". These are precisely the key channels highlighted by our model.

[2] The fact that trade integration leads to selection and productivity-enhancing market share reallocations is well documented empirically. See, e.g., Aw *et al.* (2000), Pavcnik (2002), Trefler (2004) or Bernard *et al.* (2007a).

the first feature neglects differential wage responses following trade liberalization across asymmetric economies. The second feature does not accord with abundant empirical evidence that markups differ across firms and markets, and that trade integration reduces price-cost margins due to pro-competitive effects (e.g., Tybout, 2003; Syverson, 2004, 2007; Badinger, 2007; Foster *et al.*, 2008; Feenstra and Weinstein, 2010).

To deal with these limitations, several alternatives have been put forward in the literature. Bernard *et al.* (2007b) embed Melitz's model into a Heckscher-Ohlin framework, thereby allowing for factor price differences between two asymmetric countries. Arkolakis *et al.* (2008) extend Melitz's model to the case of multiple asymmetric countries, thereby endogenizing wages and relaxing FPE. However, these two models rely on the CES specification so that markups (both at the firm level and in the aggregate) are invariant: they are thus silent on the pro-competitive effects of trade integration that we quantify in the counterfactual analysis. Taking an alternative route, Melitz and Ottaviano (2008) propose a monopolistic competition model with multiple asymmetric countries, in which markups depend on trade costs and can vary across firms and regions. However, due to the quasi-linear specification and the associated assumptions, their trade equilibrium displays FPE and precludes differential wage responses. Finally, Bernard *et al.* (2003) extend Eaton and Kortum's (2002) multi-country Ricardian model to allow for imperfect competition. Although they relax FPE, wages do not respond to trade integration due to the assumption of exogenous wage differences across countries.[3] Also, their framework does not allow for pro-competitive effects as markups are independent of the number of competing firms and identically distributed across countries.

Turning to the gravity equation literature, the conventional estimation method relies on importer and exporter fixed effects. While this yields consistent estimates of the trade friction parameters (Feenstra, 2004), labor market clearing, zero profit, and trade balance conditions are ignored. Hence, as argued by Anderson and van Wincoop (2003), fixed-effects estimation has severe limitations when it comes to conducting counterfactual analysis. The two major reasons why we adopt a structural estimation approach are that it allows us to obtain trade friction parameters and region-specific technological possibilities that are consistent with general equilibrium, and that we can compute the wages, productivity, weighted averages of markups as well as other variables that we would observe in a 'borderless' world. Our departure from Anderson and van Wincoop (2003), who use a CES specification without firm heterogeneity, is to consistently accommodate endogenous wages, productivity, and markups within the gravity equation system.[4] Our frame-

[3]Alvarez and Lucas (2007) extend Eaton and Kortum's (2002) model with perfect competition to include endogenously determined wages. To the best of our knowledge, such an extension is missing to date for the model by Bernard *et al.* (2003) with imperfect competition.

[4]Anderson and van Wincoop (2003) abstract from the direct impact of endogenous wages on the predicted trade flows, although they take into account how wage changes are related to the predicted price indices.

work also extends recent gravity equations derived from heterogeneous firms models (e.g., Chaney, 2008; Helpman *et al.*, 2008; Melitz and Ottaviano, 2008), in which either FPE holds or wages are assumed to differ exogenously across regions.

The remainder of the paper is organized as follows. Section 2 sets up the basic closed economy model. Section 3 extends it to the open economy case and conducts comparative statics analysis. Section 4 derives the gravity equation system, describes the data, and presents the estimation procedure and results. Section 5 carries out the counterfactual experiment. Section 6 concludes.

2 Closed economy

Consider a closed economy with a final consumption good, provided as a continuum of horizontally differentiated varieties. We denote by Ω the endogenously determined set of available varieties, with measure N. There are L consumers, each of whom supplies inelastically one unit of labor, which is the only factor of production.

2.1 Preferences and demands

All consumers have identical preferences that display 'love of variety' and give rise to demands with variable elasticity. Following Behrens and Murata (2007), the utility maximization problem of a representative consumer is given by:

$$\max_{q(j),\, j\in\Omega} \; U \equiv \int_\Omega \left[1 - \mathrm{e}^{-\alpha q(j)}\right] \mathrm{d}j \qquad \text{s.t.} \quad \int_\Omega p(j) q(j) \mathrm{d}j = E, \tag{1}$$

where E denotes expenditure; $p(j) > 0$ and $q(j) \geq 0$ stand for the price and the per capita consumption of variety j; and $\alpha > 0$ is a parameter. As shown in Appendix A.1, solving (1) yields the following demand functions:

$$q(i) = \frac{E}{N\overline{p}} - \frac{1}{\alpha}\left\{\ln\left[\frac{p(i)}{N\overline{p}}\right] + h\right\}, \quad \forall i \in \Omega, \tag{2}$$

where

$$\overline{p} \equiv \frac{1}{N}\int_\Omega p(j)\mathrm{d}j \quad \text{and} \quad h \equiv -\int_\Omega \ln\left[\frac{p(j)}{N\overline{p}}\right]\frac{p(j)}{N\overline{p}}\mathrm{d}j$$

denote the average price and the differential entropy of the price distribution, respectively.[5] Since marginal utility at zero consumption is bounded, the demand for a variety need not be positive.

[5] As shown in Reza (1994, pp.278-279), the differential entropy h takes its maximum value when there is no price dispersion, i.e., $p(i) = \overline{p}$ for all $i \in \Omega$. In that case, we would observe $h = -\ln(1/N)$ and thus $q(i) = E/(N\overline{p})$ by (2). Behrens and Murata (2007) entirely focus on such a symmetric case. On the contrary, this paper considers firm heterogeneity, so that not only the average price $\overline{p}$ but also the entire price distribution matter for the demand $q(i)$. The differential entropy h in (2) does capture the latter price dispersion.

Indeed, as can be seen from (2), the demand for variety i is positive if and only if its price is lower than the reservation price p^d. Formally,

$$q(i) > 0 \quad \iff \quad p(i) < p^d \equiv N\overline{p}\,\mathrm{e}^{\frac{\alpha E}{N\overline{p}} - h}. \tag{3}$$

Note that the reservation price p^d is a function of the price aggregates $\overline{p}$ and h. Combining expressions (2) and (3) allows us to express the demand for variety i concisely as follows:

$$q(i) = \frac{1}{\alpha}\ln\left[\frac{p^d}{p(i)}\right]. \tag{4}$$

Observe that the price elasticity of demand for variety i is given by $1/[\alpha q(i)]$. Thus, if individuals consume more of this variety (which is, e.g., the case when their expenditure increases), they become less price sensitive. Since $\mathrm{e}^{-\alpha q(i)} = p(i)/p^d$, the indirect utility is given by

$$U = N - \int_\Omega \frac{p(i)}{p^d}\mathrm{d}i = N\left(1 - \frac{\overline{p}}{p^d}\right), \tag{5}$$

which we use to compute the equilibrium utility in the subsequent analysis.

2.2 Technology and market structure

The labor market is assumed to be perfectly competitive so that all firms take the wage rate w as given. Prior to production, each firm enters the market by engaging in research and development, which requires a fixed amount F of labor paid at the market wage. Each firm discovers its marginal labor requirement $m(i) \geq 0$ only after making this irreversible entry decision. We assume that $m(i)$ is drawn from a common and known, continuously differentiable distribution G. Since research and development costs are sunk, an entrant will survive (i.e., operate) in the market provided it can charge a price $p(i)$ above marginal cost $m(i)w$.

Each surviving firm sets its price to maximize operating profit

$$\pi(i) = L\big[p(i) - m(i)w\big]q(i), \tag{6}$$

where $q(i)$ is given by (4). Since there is a continuum of firms, no individual firm has any impact on p^d so that the first-order conditions for (operating) profit maximization are given by:

$$\ln\left[\frac{p^d}{p(i)}\right] = \frac{p(i) - m(i)w}{p(i)}, \quad \forall i \in \Omega. \tag{7}$$

A price distribution satisfying (7) is called a *price equilibrium.* Equations (4) and (7) imply that $q(i) = (1/\alpha)[1 - m(i)w/p(i)]$, which allows us to derive the upper and lower bounds for the marginal labor requirement. The maximum output is given by $q(i) = 1/\alpha$ at $m(i) = 0$. The minimum output is given by $q(i) = 0$ at $p(i) = m(i)w$, which by (7) implies that $p(i) = p^d$. Therefore, the cutoff marginal labor requirement is defined as $m^d \equiv p^d/w$. All firms that draw $m \geq m^d$ choose not to produce, whereas all firms with a draw $m < m^d$ will operate in equilibrium. Hence, given a mass of entrants N^E, only a fraction $G(m^d)$ of them will have positive output. The mass of surviving firms is then given by $N = N^E G(m^d)$.

Since firms differ only by their marginal labor requirement, we can express all firm-level variables in terms of m. Solving (7) by using the Lambert W function, defined as $\varphi = W(\varphi)\mathrm{e}^{W(\varphi)}$, the profit-maximizing prices and quantities, as well as operating profits, can be expressed as follows (see Appendix A.2 for the derivations):

$$p(m) = \frac{mw}{W}, \quad q(m) = \frac{1}{\alpha}(1 - W), \quad \pi(m) = \frac{Lmw}{\alpha}\left(W^{-1} + W - 2\right), \tag{8}$$

where we suppress the argument $\mathrm{e}m/m^d$ of W to alleviate notation. As shown in Appendix A.2, $W' > 0$ for all non-negative arguments. Further, $W(0) = 0$ and $W(\mathrm{e}) = 1$. Hence, $0 \leq W \leq 1$ if $0 \leq m \leq m^d$. The expressions in (8) show that a firm with a draw m^d charges a price equal to marginal cost, faces zero demand, and earns zero operating profit. Since $W' > 0$, we readily obtain $\partial p(m)/\partial m > 0$, $\partial q(m)/\partial m < 0$, and $\partial \pi(m)/\partial m < 0$. In words, firms with better draws charge lower prices, sell larger quantities, and earn higher operating profits. Furthermore, markups defined as

$$\Lambda(m) \equiv \frac{p(m)}{mw} = \frac{1}{W} \tag{9}$$

are higher for more productive firms, because $\partial\Lambda(m)/\partial m < 0$.

2.3 Equilibrium

The equilibrium conditions for the closed economy consist of zero expected profits and labor market clearing. These two conditions can be solved for the cutoff m^d and the mass of entrants N^E. Using (6), the zero expected profit condition for each firm is given by:

$$L\int_0^{m^d} [p(m) - mw]\, q(m)\mathrm{d}G(m) = Fw, \tag{10}$$

which, combined with (8), can be rewritten as a function of m^d only:

$$\frac{L}{\alpha}\int_0^{m^d} m\left(W^{-1} + W - 2\right)\mathrm{d}G(m) = F. \tag{11}$$

As the left-hand side of (11) is strictly increasing in m^d from 0 to ∞, there always exists a unique equilibrium cutoff (see Appendix A.3 for the proof). Furthermore, the labor market clearing condition is given by:[6]

$$N^E\left[L\int_0^{m^d} mq(m)\mathrm{d}G(m) + F\right] = L, \tag{12}$$

[6] Note that by using (10) and the budget constraint $N^E\int_0^{m^d} p(m)q(m)\mathrm{d}G(m) = E$, we obtain $EL/(wN^E) = L\int_0^{m^d} mq(m)\mathrm{d}G(m) + F$ which, together with (12), yields $E = w$ in equilibrium.

which, combined with (8), can be rewritten as a function of m^d and N^E:

$$N^E \left[\frac{L}{\alpha} \int_0^{m^d} m\,(1-W)\,\mathrm{d}G(m) + F \right] = L. \tag{13}$$

Given the equilibrium cutoff m^d from (11), equation (13) can be uniquely solved for N^E.

How does population size affect entry and firms' survival probabilities? Using the equilibrium conditions (11) and (13), we can show that a larger L leads to more entrants N^E and a smaller cutoff m^d, respectively (see Appendix A.4 for the proofs). Hence, the survival probability $G(m^d)$ of entrants is lower in larger markets. The effect of population size on the mass of surviving firms N cannot be signed for a general distribution G. However, under the commonly made assumption that firms' productivity draws $1/m$ follow a Pareto distribution

$$G(m) = \left(\frac{m}{m^{\max}}\right)^k,$$

with upper bound $m^{\max} > 0$ and shape parameter $k \geq 1$, we can show that N is increasing in L.[7] Using this distributional assumption, we readily obtain the following closed-form solutions for the equilibrium cutoff and mass of entrants:

$$m^d = \left(\frac{\mu^{\max}}{L}\right)^{\frac{1}{k+1}} \quad \text{and} \quad N^E = \frac{\kappa_2}{\kappa_1+\kappa_2}\frac{L}{F}, \tag{14}$$

where κ_1 and κ_2 are positive constants that solely depend on k (see Appendices B.1 and B.2); and $\mu^{\max} \equiv \left[\alpha F (m^{\max})^k\right]/\kappa_2$.[8] The term $\mu^{\max}$ can be interpreted as a measure of 'technological possibilities': the lower is the fixed labor requirement for entry F or the lower is the upper bound $m^{\max}$, the lower is $\mu^{\max}$ and hence the equilibrium cutoff m^d. Selection is therefore tougher in markets with better technological possibilities. Since $\overline{m} = [k/(k+1)]m^d$ holds when productivity follows a Pareto distribution, a larger population or better technological possibilities also map into higher average productivity $1/\overline{m}$. The mass of surviving firms is given by

$$N = \frac{1}{\kappa_1+\kappa_2}\frac{\alpha}{m^d} = \frac{\alpha}{\kappa_1+\kappa_2}\left(\frac{L}{\mu^{\max}}\right)^{\frac{1}{k+1}}, \tag{15}$$

which is also higher in larger markets or markets with better technological possibilities.

We next turn to the issue of markups that consumers face. As firms are heterogeneous and have different markups and market shares, the simple (unweighted) average of markups is not an adequate measure of consumers' exposure to market power. Using (8) and (9), we hence define the (expenditure share) weighted average of firm-level markups as follows:

$$\overline{\Lambda} \equiv \frac{1}{G(m^d)} \int_0^{m^d} \frac{p(m)q(m)}{E} \Lambda(m)\mathrm{d}G(m) = \frac{\kappa_3 m^d}{\alpha}, \tag{16}$$

[7] The Pareto distribution has been extensively used in the previous literature on heterogeneous firms (e.g., Bernard *et al.*, 2007; Helpman *et al.*, 2008; Melitz and Ottaviano, 2008).

[8] For this solution to be consistent, we must ensure that $m^d \leq m^{\max}$, i.e., $m^{\max} \geq \alpha F/(\kappa_2 L)$.

where κ_3 is a positive constant that solely depends on k (see Appendix B.3).[9] Note that the weighted average of markups is proportional to the cutoff. It thus follows from (15) and (16) that our model displays pro-competitive effects, since $\overline{\Lambda}$ decreases with the mass N of competing firms:

$$\overline{\Lambda} = \frac{\kappa_3}{\kappa_1 + \kappa_2} \frac{1}{N}.$$

Note further that expression (16), together with (14), shows that $\overline{\Lambda}$ is smaller in larger markets or markets with better technological possibilities, as more firms compete in these markets.[10]

Finally, we show in Appendix A.5 that the indirect utility is given by

$$U = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\alpha}{m^d} = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\kappa_3}{\overline{\Lambda}}, \tag{17}$$

where the term in square brackets is, by construction of the utility function, positive for all $k \geq 1$. Alternatively, indirect utility can be written as $U = [1/(k+1) - (\kappa_1 + \kappa_2)]N$. Hence, as can be seen from (14)–(17), larger markets or markets with better technological possibilities allow for higher utility because of tougher selection, tougher competition, and greater consumption diversity.

3 Open economy

We now turn to the open economy case. As dealing with two regions only marginally alleviates the notational burden, we first derive the equilibrium conditions for the general case with K asymmetric regions that we use when taking our model to the data. We then present some clear-cut analytical results for the special case of two asymmetric regions in order to guide the intuition for the general case. We finally compare our results with the existing literature.

3.1 Preferences and demands

Preferences in the open economy case are analogous to the ones described in the previous section. Let $p_{sr}(i)$ and $q_{sr}(i)$ denote the price and the per capita consumption of variety i when it is produced in region s and consumed in region r. The utility maximization problem of a consumer in region r is given by:

$$\max_{q_{sr}(j),\, j \in \Omega_{sr}} \; U_r \equiv \sum_s \int_{\Omega_{sr}} \left[1 - \mathrm{e}^{-\alpha q_{sr}(j)}\right] \mathrm{d}j \qquad \text{s.t.} \quad \sum_s \int_{\Omega_{sr}} p_{sr}(j) q_{sr}(j) \mathrm{d}j = E_r, \tag{18}$$

[9] Recent empirical work by Feenstra and Weinstein (2010) uses a similar (expenditure share) weighted average of markups in a translog framework.

[10] A similar result can be obtained using an alternative definition of market power like the weighted average of firm-level Lerner indices $[p(m) - mw]/p(m)$. In that case, $\overline{\Lambda}$ is given by $(\kappa_2 m^d)/\alpha$, which differs from the weighted average of markups by a constant multiplicative term only. An *unweighted* average of firm-level markups would be a constant in our model. The latter result also occurs in other models with heterogeneous firms (Bernard *et al.*, 2003; Melitz and Ottaviano, 2008).

where Ω_{sr} denotes the set of varieties produced in region s and consumed in region r. It is readily verified that the demand functions are given as follows:

$$q_{sr}(i) = \frac{E_r}{N_r^c \overline{p}_r} - \frac{1}{\alpha}\left\{\ln\left[\frac{p_{sr}(i)}{N_r^c \overline{p}_r}\right] + h_r\right\}, \quad \forall i \in \Omega_{sr},$$

where N_r^c is the mass of varieties consumed in region r, and

$$\overline{p}_r \equiv \frac{1}{N_r^c}\sum_s \int_{\Omega_{sr}} p_{sr}(j)\mathrm{d}j \quad \text{and} \quad h_r \equiv -\sum_s \int_{\Omega_{sr}} \ln\left[\frac{p_{sr}(j)}{N_r^c \overline{p}_r}\right]\frac{p_{sr}(j)}{N_r^c \overline{p}_r}\mathrm{d}j$$

denote the average price and the differential entropy of the price distribution of all varieties consumed in region r. As in the closed economy case, the demand for domestic variety i (resp., foreign variety j) is positive if and only if the price of variety i (resp., variety j) is lower than the reservation price p_r^d. Formally,

$$q_{rr}(i) > 0 \iff p_{rr}(i) < p_r^d \quad \text{and} \quad q_{sr}(j) > 0 \iff p_{sr}(j) < p_r^d,$$

where $p_r^d \equiv N_r^c \overline{p}_r \mathrm{e}^{\alpha E_r/(N_r^c \overline{p}_r) - h_r}$ is a function of the price aggregates $\overline{p}_r$ and h_r. The demands for domestic and foreign varieties can then be concisely expressed as follows:

$$q_{rr}(i) = \frac{1}{\alpha}\ln\left[\frac{p_r^d}{p_{rr}(i)}\right] \quad \text{and} \quad q_{sr}(j) = \frac{1}{\alpha}\ln\left[\frac{p_r^d}{p_{sr}(j)}\right]. \tag{19}$$

Since $\mathrm{e}^{-\alpha q_{sr}(j)} = p_{sr}(j)/p_r^d$, the indirect utility is given by

$$U_r = N_r^c - \sum_s \int_{\Omega_{sr}} \frac{p_{sr}(j)}{p_r^d}\mathrm{d}j = N_r^c\left(1 - \frac{\overline{p}_r}{p_r^d}\right), \tag{20}$$

which we use to compute the equilibrium utility in the subsequent analysis.

3.2 Technology and market structure

Technology and the entry process are identical to the ones described in Section 2. We assume that markets are segmented, where costs of resale or third-party arbitrage are sufficiently high, and that firms are free to price discriminate.

Firms in region r independently draw their value of m from a region-specific distribution G_r. Assuming that shipments from r to s are subject to trade costs $\tau_{rs} > 1$ for all r and s, which firms incur in terms of labor, the operating profit of firm i in r is given by:

$$\pi_r(i) = \sum_s \pi_{rs}(i) = \sum_s L_s q_{rs}(i)\left[p_{rs}(i) - \tau_{rs} m_r(i) w_r\right]. \tag{21}$$

Each firm maximizes (21) with respect to its prices $p_{rs}(i)$ separately. Since it has no impact on the price aggregates and on the wages, the first-order conditions are given by:

$$\ln\left[\frac{p_s^d}{p_{rs}(i)}\right] = \frac{p_{rs}(i) - \tau_{rs} m_r(i) w_r}{p_{rs}(i)}, \quad \forall i \in \Omega_{rs}. \tag{22}$$

Equations (19) and (22) imply that $q_{rs}(i) = (1/\alpha)[1 - \tau_{rs} m_r(i) w_r / p_{rs}(i)]$, which shows that $q_{rs}(i) = 0$ at $p_{rs}(i) = \tau_{rs} m_r(i) w_r$. It then follows from (22) that $p_{rs}(i) = p_s^d$. Hence, a firm located in r with draw $m_{rs}^x \equiv p_s^d/(\tau_{rs} w_r)$ is just indifferent between selling and not selling in region s. All firms in r with draws below m_{rs}^x are productive enough to sell to region s. In what follows, we refer to $m_{ss}^x \equiv m_s^d$ as the *domestic cutoff* in region s, whereas m_{rs}^x with $r \neq s$ is the *export cutoff*. Export and domestic cutoffs are linked as follows:

$$m_{rs}^x = \frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d. \tag{23}$$

Expression (23) reveals the key relationship between trade costs, wages and productivity. In particular, it shows how trade costs and wage differences affect firms' ability to break into market s. When wages are equalized ($w_r = w_s$) and internal trade is costless ($\tau_{ss} = 1$), all export cutoffs must fall short of the domestic cutoffs since $\tau_{rs} > 1$. Breaking into market s is then always harder for firms in $r \neq s$ than for firms in s, which is the standard case in the literature (e.g., Melitz, 2003; Melitz and Ottaviano, 2008). However, consider a case where $w_s > \tau_{rs} w_r$ and $\tau_{ss} = 1$. In that case, firms from the low wage region r have a cost advantage in selling to the high wage market s compared to their competitors located in region s. Surviving in market s is then easier for exporters from r than for domestic firms in s, i.e., $m_{rs}^x > m_s^d$. More generally, in the presence of wage differences and internal trade costs, the domestic cutoff in market s need not be larger than the export cutoff from market r to market s in equilibrium. The usual ranking, namely $m_{rs}^x < m_s^d$, prevails only when $\tau_{ss} w_s < \tau_{rs} w_r$.[11] We will show in Section 4.4 that the case $m_{rs}^x > m_s^d$ is not negligible as it shows up for more than 6.5% of all possible pairs of export and domestic cutoffs in our estimated model.

Given a mass of entrants N_r^E and export cutoffs m_{rs}^x as in (23), only $N_r^p = N_r^E \, G_r\left(\max_s \{m_{rs}^x\}\right)$ firms survive in region r, namely those which are productive enough to sell at least in one market (which need not be the local market). The mass of varieties consumed in region r is

$$N_r^c = \sum_s N_s^E G_s(m_{sr}^x), \tag{24}$$

which is the sum of all firms that are productive enough to serve market r.

Finally, the first-order conditions (22) can be solved as in the closed economy case by using the Lambert W function. Switching to notation in terms of m, the profit-maximizing prices and quantities, as well as operating profits, are given by:

$$p_{rs}(m) = \frac{\tau_{rs} m w_r}{W}, \quad q_{rs}(m) = \frac{1}{\alpha}\left(1 - W\right), \quad \pi_{rs} = \frac{L_s \tau_{rs} m w_r}{\alpha}(W^{-1} + W - 2), \tag{25}$$

[11] As recently emphasized by Foster *et al.* (2008), firm selection and export performance depend more generally on profitability. In addition to physical productivity, locational heterogeneity matters in our model because firms face different wages and have different market access depending on which region they are located in.

where W denotes the Lambert W function with argument $\mathrm{e}\tau_{rs} m w_r / p_s^d$, which we suppress to alleviate notation. It is readily verified that more productive firms again charge lower prices, sell larger quantities, and earn higher operating profits. Furthermore, markups defined as

$$\Lambda_{rs}(m) \equiv \frac{p_{rs}(m)}{\tau_{rs} m w_r} = \frac{1}{W} \tag{26}$$

are higher for more productive firms.

3.3 Equilibrium

The zero expected profit condition for each firm in region r is given by

$$\sum_s L_s \int_0^{m_{rs}^x} \left[p_{rs}(m) - \tau_{rs} m w_r\right] q_{rs}(m) \mathrm{d}G_r(m) = F_r w_r, \tag{27}$$

where F_r is the region-specific fixed labor requirement. Furthermore, each labor market clears in equilibrium, which requires that

$$N_r^E \left[\sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m q_{rs}(m) \mathrm{d}G_r(m) + F_r\right] = L_r. \tag{28}$$

Last, in equilibrium trade must be balanced for each region:

$$N_r^E \sum_{s \neq r} L_s \int_0^{m_{rs}^x} p_{rs}(m) q_{rs}(m) \mathrm{d}G_r(m) = L_r \sum_{s \neq r} N_s^E \int_0^{m_{sr}^x} p_{sr}(m) q_{sr}(m) \mathrm{d}G_s(m).$$

As in the foregoing section, we can restate the equilibrium conditions using the Lambert W function (see Appendix C for details).

In what follows, we assume that productivity draws $1/m$ follow Pareto distributions with identical shape parameters $k \geq 1$. However, we allow the upper bounds to vary across regions, i.e., $G_r(m) = (m/m_r^{\max})^k$. Under the Pareto parametrization, the equilibrium conditions can be greatly simplified. First, using the expressions in Appendices B.1 and C.1, labor market clearing requires that

$$N_r^E \left[\frac{\kappa_1}{\alpha \left(m_r^{\max}\right)^k} \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d\right)^{k+1} + F_r\right] = L_r. \tag{29}$$

Second, using the expressions in Appendices B.2 and C.2, zero expected profits imply that

$$\mu_r^{\max} = \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d\right)^{k+1}, \tag{30}$$

where, analogously to the closed economy case, $\mu_r^{\max} \equiv \left[\alpha F_r \left(m_r^{\max}\right)^k\right] / \kappa_2$ denotes region r's technological possibilities. Last, using the expressions in Appendices B.4 and C.3, balanced trade requires that

$$\frac{N_r^E w_r}{\left(m_r^{\max}\right)^k} \sum_{s \neq r} L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d\right)^{k+1} = L_r \sum_{s \neq r} \tau_{sr} \frac{N_s^E w_s}{\left(m_s^{\max}\right)^k} \left(\frac{\tau_{rr}}{\tau_{sr}} \frac{w_r}{w_s} m_r^d\right)^{k+1}. \tag{31}$$

The $3K$ conditions (29)–(31) depend on $3K$ unknowns: the wages w_r, the masses of entrants N_r^E, and the domestic cutoffs m_r^d. The export cutoffs m_{rs}^x can then be computed using (23). Combining (29) and (30) immediately shows that

$$N_r^E = \frac{\kappa_2}{\kappa_1 + \kappa_2} \frac{L_r}{F_r}. \tag{32}$$

Thus, in the open economy case the mass of entrants in region r still positively depends on that region's size L_r and negatively on its fixed labor requirement F_r.

Adding the term in r that is missing on both sides of (31), and using (30) and (32), we obtain the following equilibrium relationship:

$$\frac{1}{(m_r^d)^{k+1}} = \sum_s L_s \tau_{rr} \left(\frac{\tau_{rr}}{\tau_{sr}} \frac{w_r}{w_s} \right)^k \frac{1}{\mu_s^{\max}}. \tag{33}$$

Expressions (30) and (33) summarize how wages, technological possibilities, cutoffs, trade costs and population sizes are related in general equilibrium.

Using the foregoing expressions, we can show that the mass of varieties consumed in a region is inversely proportional to the domestic cutoff in that region (see Appendix A.6 for the derivation):

$$N_r^c = \frac{1}{\kappa_1 + \kappa_2} \frac{\alpha}{\tau_{rr} m_r^d}. \tag{34}$$

Furthermore, the (expenditure share) weighted average of markups that consumers face in region r is given by (see Appendix A.7 for the derivation):

$$\overline{\Lambda}_r \equiv \frac{\sum_s N_s^E \int_0^{m_{sr}^x} \frac{p_{sr}(m) q_{sr}(m)}{E_r} \Lambda_{sr}(m) \mathrm{d}G_s(m)}{\sum_s N_s^E G_s(m_{sr}^x)} = \frac{\kappa_3 \tau_{rr} m_r^d}{\alpha}. \tag{35}$$

Hence, the weighted average of markups is proportional to the cutoff. It thus follows from (34) and (35) that there are pro-competitive effects, since $\overline{\Lambda}_r$ decreases with the mass N_r^c of competing firms in region r:

$$\overline{\Lambda}_r = \frac{\kappa_3}{\kappa_1 + \kappa_2} \frac{1}{N_r^c}.$$

Finally, we show in Appendix A.8 that the indirect utility in region r can be expressed as follows:

$$U_r = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\alpha}{\tau_{rr} m_r^d} = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\kappa_3}{\overline{\Lambda}_r}, \tag{36}$$

which implies that tougher selection and a lower weighted average of markups in region r translate into higher welfare. Alternatively, the indirect utility can be written as $U_r = [1/(k+1) - (\kappa_1 + \kappa_2)] N_r^c$, i.e., it is proportional to the mass of varieties consumed. Note that the welfare gains come from foreign varieties (Broda and Weinstein, 2006), as the mass of domestic varieties

$N_r^E G_r(m_r^d)$ decreases when trade integration reduces m_r^d (which is indeed the case in our counterfactual analysis below). This is in accord with Feenstra and Weinstein (2010), who show that new import varieties have contributed to US welfare gains even when taking into account the displaced domestic varieties.

3.4 Some comparative statics

In order to build intuition for the multi-region case, we illustrate some comparative statics results for two asymmetric regions. Using (30)–(32), an equilibrium is characterized by a system of three equations with three unknowns (the two cutoffs m_1^d and m_2^d, and the relative wage $\omega \equiv w_1/w_2$):

$$\mu_1^{\max} = L_1\tau_{11}\left(m_1^d\right)^{k+1} + L_2\tau_{12}\left(\frac{\tau_{22}}{\tau_{12}}\,\frac{1}{\omega}\,m_2^d\right)^{k+1} \tag{37}$$

$$\mu_2^{\max} = L_2\tau_{22}\left(m_2^d\right)^{k+1} + L_1\tau_{21}\left(\frac{\tau_{11}}{\tau_{21}}\,\omega\,m_1^d\right)^{k+1} \tag{38}$$

$$\omega^{2k+1} = \rho\left(\frac{\tau_{21}}{\tau_{12}}\right)^{k}\left(\frac{\tau_{22}}{\tau_{11}}\right)^{k+1}\left(\frac{m_2^d}{m_1^d}\right)^{k+1}, \tag{39}$$

where $\rho \equiv \mu_2^{\max}/\mu_1^{\max}$. When $\rho > 1$, region 1 has better technological possibilities than region 2. Equations (37) and (38) can readily be solved for the cutoffs as a function of the relative wage:

$$(m_1^d)^{k+1} = \frac{\mu_1^{\max}}{L_1\tau_{11}}\frac{1-\rho\left(\frac{\tau_{22}}{\tau_{12}}\right)^k\omega^{-(k+1)}}{1-\left(\frac{\tau_{11}\tau_{22}}{\tau_{12}\tau_{21}}\right)^k} \quad \text{and} \quad (m_2^d)^{k+1} = \frac{\mu_2^{\max}}{L_2\tau_{22}}\frac{1-\rho^{-1}\left(\frac{\tau_{11}}{\tau_{21}}\right)^k\omega^{k+1}}{1-\left(\frac{\tau_{22}\tau_{11}}{\tau_{21}\tau_{12}}\right)^k}. \tag{40}$$

Substituting the cutoffs (40) into (39) yields after some simplification

$$\text{LHS} \equiv \omega^k = \rho\frac{L_1}{L_2}\left(\frac{\tau_{21}}{\tau_{12}}\right)^k\frac{\rho\tau_{11}^{-k} - \tau_{21}^{-k}\omega^{k+1}}{\tau_{22}^{-k}\omega^{k+1} - \rho\tau_{12}^{-k}} \equiv \text{RHS}. \tag{41}$$

Assume that intraregional trade is less costly than interregional trade, i.e., $\tau_{11} < \tau_{21}$ and $\tau_{22} < \tau_{12}$. Then, the RHS of (41) is decreasing in ω on its relevant domain, whereas the LHS is increasing in ω. Hence, there exists a unique equilibrium such that the equilibrium relative wage ω^* is bounded by relative trade costs τ_{22}/τ_{12} and τ_{21}/τ_{11}, relative technological possibilities ρ, and the shape parameter k (see Appendix A.9). Since the RHS of (41) is decreasing in ω, the following comparative statics results are straightforward to derive (see Appendix A.10 for the derivations).

Suppose that the two regions are identical except that region 1 is (i) larger or has (ii) better technological possibilities than region 2. Then, region 1 has the higher wage and the lower cutoff, i.e., the higher average productivity. The reason is that the larger market size or the better technological possibilities in region 1 are, ceteris paribus, associated with higher profitability of entry. To offset this advantage in region 1 requires, in equilibrium, a higher wage and a lower cutoff in that region. Expression (34) shows that a lower cutoff maps into greater consumption diversity, which intensifies competition and lowers the weighted average of markups by (35). Consequently, welfare increases as can be seen by (36).

(iii) Turning to the impacts of trade liberalization, suppose that the two regions have symmetric (both internal and bilateral) trade costs, but differ in size or technological possibilities. Then, as bilateral trade costs fall, wages, average productivity, weighted averages of markups, consumption diversity and welfare converge between the two regions. This suggests that bilateral trade liberalization tends to attenuate regional economic differences, rather than to exacerbate them.[12]

3.5 Comparison with other models

To our knowledge, there exists so far no trade model with heterogeneous firms that can capture all the foregoing endogenous economic differences between regions. First, CES models that extend the standard Melitz (2003) framework to the case of asymmetric countries, such as Arkolakis *et al.* (2008), may come to similar conclusions about the impacts of size or technological differences on wages and cutoffs.[13] However, firm-level markups are fixed in CES models because the price elasticity of demand is constant by construction. This result contradicts abundant empirical evidence which shows that firm-level markups depend on productivity and local market size (e.g., Tybout, 2003; Syverson, 2004, 2007; Foster *et al.*, 2008). Furthermore, industry-level analysis suggests that markups react to trade liberalization (Badinger, 2007; Feenstra and Weinstein, 2010).

Second, Melitz and Ottaviano (2008) develop a model with quasi-linear preferences without income effects for the differentiated good. Their trade equilibrium displays FPE and, therefore, precludes differential wage responses across asymmetric regions. Last, Bernard *et al.* (2003) consider exogenous wage differences across countries, thus ruling out endogenous wage responses triggered by different trade liberalization scenarios. Contrary to these studies, however, recent empirical evidence suggests that wages do respond to trade integration. For example, Trefler (2004) finds that the Canada-US Free Trade Agreement has led to small, but significantly positive impacts on Canadian wages. As will be shown in Section 5.1, ignoring such small wage changes generates large biases in predicted cross-border trade flows.

[12] We can also prove that (iv) if the two regions are identical except that region 1 has better access to region 2 than vice versa ($\tau_{12} < \tau_{21}$), then region 1 has the higher wage, the higher average productivity, the lower weighted average of markups, greater consumption diversity and higher welfare than region 2. An advantage in accessing the other market, thus, works like a local size advantage or better technological possibilities.

[13] Chaney (2008) extends the Melitz model to the case of multiple asymmetric countries. In that model, however, there exists a homogeneous and freely tradable numeraire good, which exogenously pins down wages. Similarly, in Demidova (2008) the existence of such a good leads to FPE in any equilibrium. Hence, both models cannot cope with endogenous wage responses driven by trade integration.

4 Estimation

We now take the model with K asymmetric regions to the data. To this end, we first derive a gravity equation with general equilibrium conditions. Using the well-known Canada-US interregional trade data by Anderson and van Wincoop (2003), we then structurally estimate trade friction parameters as well as other variables of the model. Afterwards, we assess the performance of our estimated model by comparing its predictions to several empirical facts, both at the regional and at the firm level. The next section turns to a counterfactual analysis where we consider the impacts of removing all trade barriers generated by the Canada-US border.

4.1 Gravity equation system

The value of exports from region r to region s is given by

$$X_{rs} = N_r^E L_s \int_0^{m_{rs}^x} p_{rs}(m) q_{rs}(m) \mathrm{d}G_r(m).$$

Using equations (25), (32), and the Pareto distribution for $G_r(m)$, we obtain the following gravity equation:[14]

$$\frac{X_{rs}}{L_r L_s} = \tau_{rs}^{-k} \tau_{ss}^{k+1} (w_s/w_r)^{k+1} w_r \left(m_s^d\right)^{k+1} \left(\mu_r^{\max}\right)^{-1}. \tag{42}$$

As can be seen from (42), exports depend on bilateral trade costs τ_{rs}, internal trade costs in the destination τ_{ss}, origin and destination wages w_r and w_s, the destination cutoff m_s^d, and origin technological possibilities $\mu_r^{\max}$. A higher relative wage w_s/w_r raises the value of exports as firms in r face relatively lower production costs, whereas a higher absolute wage w_r raises the value of exports by increasing export prices p_{rs}. Furthermore, a larger m_s^d raises the value of exports since firms located in the destination are on average less productive. Last, a lower $\mu_r^{\max}$ implies that firms in region r have higher expected productivity, which raises the value of their exports. Expressions (30) and (33) give us the following general equilibrium conditions:

$$\mu_r^{\max} = \sum_v L_v \tau_{rv}^{-k} \tau_{vv}^{k+1} \left(\frac{w_v}{w_r}\right)^{k+1} \left(m_v^d\right)^{k+1} \qquad r = 1, 2. \dots K \tag{43}$$

$$\frac{1}{\left(m_s^d\right)^{k+1}} = \sum_v L_v \tau_{vs}^{-k} \tau_{ss}^{k+1} \left(\frac{w_s}{w_v}\right)^{k} \left(\mu_v^{\max}\right)^{-1} \qquad s = 1, 2, \dots K \tag{44}$$

The *gravity equation system* consists of the gravity equation (42) and the $2K$ general equilibrium conditions (43) and (44) that summarize the interactions between the $2K$ endogenous variables, namely the wages and cutoffs.

[14] Contrary to standard practice in the gravity literature, we do not move the GDPs but instead move the population sizes to the left-hand side. Applying the former approach to our model would amount to assuming that wages are exogenous in the gravity estimation, which is not the case in general equilibrium (see Bergstrand, 1985, for an early contribution on this issue).

Expressions (43) and (44) are reminiscent of the conditions in Anderson and van Wincoop (2003), who argue that general equilibrium interdependencies need to be taken into account when conducting a counterfactual analysis based on the gravity equation.[15] One of our contributions is to go a step further by extending their approach to cope with endogenous wages and productivity. Note that expression (42) is similar to gravity equations that have been derived in previous models with heterogeneous firms. These models rely, however, either on exogenous wages (Chaney, 2008) or on FPE (Melitz and Ottaviano, 2008) and also often disregard general equilibrium conditions when being estimated (Helpman *et al.*, 2008).[16]

4.2 Data and estimation procedure

To estimate the gravity equation system (42)–(44), we rely on aggregate bilateral trade flows X_{rs} and internal absorption X_{rr} for 10 Canadian provinces and 30 US states in 1993.[17] We further have geographical coordinates of the capitals, regional surface measures, and regional population sizes L_r in 1993 for these 40 regions. The latter are obtained from Statistics Canada and the US Census Bureau. For the specification of trade costs τ_{rs} we stick to standard practice by assuming that $\tau_{rs} \equiv d_{rs}^{\gamma} \mathrm{e}^{\theta b_{rs}}$, where d_{rs} stands for distance between r and s and is computed using the great circle formula. The internal distances are measured as $d_{rr} = (2/3)\sqrt{\text{surface}_r/\pi}$ like in Redding and Venables (2004).[18] The term b_{rs} is a border dummy valued 1 if r and s are not in the same country and 0 otherwise. With this specification of trade costs we relate our analysis to the vast literature on border effects (McCallum, 1995), which has shown that regional trade flows are not only affected by physical distance, but also by the presence of the Canada-US border. The trade friction parameters γ and θ are to be estimated.

The estimation of the gravity equation system poses four main difficulties which we need to deal with. First, although we require the value of the shape parameter k, it cannot be structurally identified from the estimated parameters of the model. We thus proceed as follows. We choose an arbitrary initial value for k to estimate the gravity equation system as described below. Using the estimates thus obtained, as well as the chosen value of k, we then compute the productivity advan-

[15]It is possible to treat w_r, w_s, m_s^d and $\mu_r^{\max}$ as fixed effects in equation (42) and obtain consistent estimates of trade friction parameters while ignoring the general equilibrium conditions (43) and (44). However, this approach cannot be used for a counterfactual analysis since the effect of the hypothetical decrease in trade frictions on the estimated fixed effects is not known. In our approach, the endogenous responses of wages and cutoffs are crucial when evaluating the counterfactual trade liberalization scenario below.

[16]One exception is Balistreri and Hillberry (2007), who allow regional incomes to respond to trade liberalization. However, they do not consider endogenous productivity as their model abstracts from firm heterogeneity.

[17]This publicly available dataset has been widely used in the literature (see Anderson and van Wincoop, 2003; Feenstra 2004), which makes it easy to compare our results to existing ones.

[18]As a robustness check we also consider the alternative measure $d_{rr} = (1/4)\min_{s\neq r}\{d_{rs}\}$ as in Anderson and van Wincoop (2003). Results are little sensitive to that choice.

tage of exporters from a random sample of firms drawn from the fitted productivity distributions of our model (see Appendix D for more details on the procedure). We repeat this procedure for different values of k until our sample allows us to match the 33% productivity advantage of US exporters that is reported by Bernard *et al.* (2003) from 1992 Census of Manufactures data. This calibration yields $k = 7.5$, which we henceforth take as our benchmark. As additional robustness checks we also consider $k = 3.6$, which is the value that has been used by Bernard *et al.* (2003), as well as a symmetrically larger value $k = 11.4$.

Second, there exists no data for $\mu_r^{\max}$ as it depends on the unobservables α, F_r and $m_r^{\max}$. To address this issue, we use the general equilibrium conditions (43) and (44). Ideally, we would plug data for $\mu_r^{\max}$ into these $2K$ conditions to solve for the $2K$ endogenous variables w_r and m_r^d. However, as $\mu_r^{\max}$ is unobservable, we rely instead on data for the K cutoffs m_r^d. This allows us to solve the $2K$ conditions (43) and (44) for theoretically consistent values of the $2K$ variables w_r and $\mu_r^{\max}$. Recall that, under the Pareto distribution, the domestic cutoff in each region is proportional to the inverse of the average productivity, i.e., $m_r^d = [(k+1)/k]\overline{m}_r$. We measure $\overline{m}_r$ by using the GDP per employee in Canadian dollars for each province and state in 1993, which is obtained from Statistics Canada and the US Census Bureau. Once we have computed the theory-consistent values of w_r from the general equilibrium conditions, we evaluate the model fit in Section 4.4 by comparing computed with observed wages.

Third, the estimates of the trade friction parameters γ and θ depend on w_r and $\mu_r^{\max}$, which depend themselves on the estimates of γ and θ. Put differently, the conditions (43) and (44) include the trade friction parameters, but to estimate the parameters of the gravity equation we need the solution to these conditions. We tackle this problem by estimating the gravity equation system iteratively as follows:

1. Given our specification of τ_{rs}, the gravity equation (42) can be rewritten in log-linear stochastic form as follows:

$$\ln\left(\frac{X_{rs}}{L_r L_s}\right) = -k\gamma \ln d_{rs} - k\theta b_{rs} + \zeta_r^1 + \zeta_s^2 + \varepsilon_{rs}, \tag{45}$$

   where all terms specific to the origin and the destination are collapsed into exporter and importer fixed effects ζ_r^1 and ζ_s^2; and where ε_{rs} is an error term with the usual properties. From (45), we obtain initial *unconstrained* estimates of the parameters $(\widehat{\gamma}', \widehat{\theta}')$.[19]

2. Using the initial estimates $(\widehat{\gamma}', \widehat{\theta}')$ and the observed cutoffs m_s^d in (43) and (44), we solve simultaneously for the equilibrium wages and the technological possibilities $(\widehat{w}_r', \widehat{\mu}_r^{\max\prime})$.

[19] Although we could choose the initial values for $(\widehat{\gamma}', \widehat{\theta}')$ arbitrarily, the fixed effects estimates provide a reasonable 'guess' for the starting values and allow for faster convergence of the iterative procedure. We experimented with different starting values and obtained the same estimates.

3. We use the computed values $(\widehat{w}_r', \widehat{\mu}_r^{\max\prime})$ to estimate the gravity equation (42) as follows:

$$\ln\left(\frac{X_{rs}}{L_r L_s}\right) + k\ln\widehat{w}_r' - (k+1)\ln\widehat{w}_s' - \ln m_s^d + \ln \widehat{\mu}_r^{\max\prime}$$
$$= -\gamma k \ln d_{rs} + \gamma(k+1)\ln d_{rr} - k\theta b_{rs} + \varepsilon_{rs},$$

which yields *constrained* estimates $(\widehat{\gamma}'', \widehat{\theta}'')$.

4. We iterate through steps 2 to 3 until convergence to obtain $(\widehat{\gamma}, \widehat{\theta})$ and $(\widehat{w}_r, \widehat{\mu}_r^{\max})$.

Last, we have to deal with the fact that bilateral trade flows among the 40 regions for which we have sufficient data are also affected by other out-of-sample regions and countries. This concern is particularly relevant in the context of a counterfactual analysis, since the trade creation and diversion effects of a hypothetical trade integration also feature general equilibrium repercussions with other trading partners. Fortunately, our gravity equation system allows us to take this issue into account, because we can include further regions and countries into the equilibrium conditions (43)–(44), even if we lack bilateral trade flow data for them.

Specifically, our full dataset includes $K = 83$ areas, namely the 10 Canadian provinces, all 50 US states plus the District of Columbia, the 21 OECD members in 1993, and Mexico.[20] The distance, population and cutoff data for the 43 areas out of the gravity sample are defined in an analogous way as those for the areas in the gravity sample. For the rest of the world (ROW), we use OECD data (including Mexico). We construct average productivities by converting 1993 hourly labor productivity in national currency into yearly figures (using hours worked) expressed in Canadian dollars.

In the iterative procedure, once we obtain initial unconstrained estimates for the structural parameters $(\widehat{\gamma}', \widehat{\theta}')$ using only the 40 regions from the 'in gravity sample', we can solve (43) and (44) for wages and technological possibilities $(\widehat{w}_r', \widehat{\mu}_r^{\max\prime})$ for the full sample of areas. The solutions for the 40 'in gravity' regions, which are affected by the general equilibrium interactions among *all* trading partners, are plugged back into (42) and the estimation proceeds as described above. The resulting final estimates of trade frictions $(\widehat{\gamma}, \widehat{\theta})$ and of wages and technological possibilities $(\widehat{w}_r, \widehat{\mu}_r^{\max})$ for all 83 areas are consistent with theory as they take into account all the equilibrium information of the model. Using this information we can then retrieve the fitted values of bilateral trade flows $\widehat{X}_{rs}$ for *all* pairs of areas, even for those not in the gravity sample.

[20] See Table 7 for the list of the 40 Canadian and US regions used in the gravity equation ('in gravity sample') and for the 21 regions used only in the general equilibrium conditions ('out of gravity sample'). Because of their extremely small population sizes we have excluded Yukon, Northwest Territories and Nunavut from the analysis. The rest of the world consists of Australia, Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, UK and Mexico, which together with Canada account for the lion's share of US trade in 1993 (66.5% of total US exports and 64.7% of total US imports).

4.3 Estimation results

Our estimation results for the gravity equation system are summarized in Table 1, where we report bootstrapped standard errors in parentheses.[21] Column 1 presents the benchmark case with $K = 83$ areas and $k = 7.5$, whereas Columns 2-6 contain alternative specifications used as robustness checks.

Insert Table 1 about here.

As can be seen from Column 1 in Table 1, all coefficients have the correct sign and are precisely estimated. In our benchmark case, the estimated distance elasticity is -1.4457, which implies that $\widehat{\gamma} = 0.1928$. The border coefficient estimate is -1.5657, which implies that $\widehat{\theta} = 0.2088$. Note that our estimated border coefficient is not statistically different from the -1.59 obtained by Anderson and van Wincoop (2003). However, as shown later, the impacts of a border removal on trade flows differ substantially once endogenous wages and firm selection are taken into account.

Columns 2-3 report results for different values of k. Column 4 presents results obtained when we exclude the 49 zero trade flows in the gravity sample.[22] In column 5, we present results using an alternative measure for internal distance as proposed by Anderson and van Wincoop (2003). In that case, we exclude the ROW from the general equilibrium conditions as this distance measure is not really appropriate in an international context. The coefficient of the border dummy varies only little across the different specifications. Column 6 finally shows the results of fixed effects estimation (step 1 of our estimation procedure). Observe that the two estimates of the border coefficient in columns 1 and 6 are not statistically different from each other as the corresponding 95% confidence intervals overlap. However, as explained above, the fixed effects approach is of little help in performing a counterfactual analysis, although it is certainly a consistent method for estimating trade frictions in our model.

4.4 Model fit and behavior

Before turning to the counterfactual experiment, we assess the performance of our estimated model by comparing its predictions to several empirical facts, both at the regional and at the firm level. We first compare the wages $\widehat{w}_r$, obtained from the general equilibrium conditions, with observed

[21] To this end, we proceed as follows. First, we randomly permute $\widehat{\varepsilon}_{rs}$ obtained after estimation step 1 to get $\widehat{\varepsilon}^b_{rs}$. We then compute the $\widehat{X}^b_{rs}$ that are consistent with the permutation to obtain new initial values $(\widehat{\theta}^b, \widehat{\gamma}^b)$. Finally, we apply steps 2-4 until convergence. By repeating this procedure, we end up with a distribution for $(\widehat{\theta}^b, \widehat{\gamma}^b)$ from which we compute standard errors.

[22] Even when focusing on just 40 regions, we still have to deal with 49 zero trade flows out of 1600 observations. Since there is no generally agreed-upon methodology to deal with this problem (see, e.g., Anderson and van Wincoop, 2004; Disdier and Head, 2008), we include a dummy variable for zero flows in the regressions. Note that our zeros are unlikely to be 'true zeros', as this would imply no aggregate manufacturing trade between several US states (see Helpman *et al.* 2008 on the treatment of 'true zeros').

wages. We construct observed wages across provinces and states using hourly wage data from Statistics Canada and the Bureau of Labor Statistics. In order to match the unit of measurement of trade and GDP data, we compute average yearly wages in million Canadian dollars based on an average of 1930 hours worked yearly in Canada, and 2080 hours worked yearly in the US in 1993. For the ROW, we use OECD data (including Mexico) on hourly wages and hours worked to construct observed wages, which are then converted into Canadian dollars. In our benchmark case, the correlation between computed and observed wages is 0.63 when including the ROW, and 0.75 when focusing only on Canadian provinces and US states. Thus, the predicted wages match observed ones fairly well.

Although our main focus is on regional aggregates such as wages, average productivity, and average markup, we can also assess our estimated model by using well-established firm-level facts, namely, the share of exporters and the distribution of export intensities. With respect to the former, Bernard *et al.* (2009) document that only 2.6% of all US firms reported exporting anything at all in 1993. Our model delivers an exporter share of 3.54% across all US firms (see Appendix D for details), which is fairly close to the observed number. The corresponding share of Canadian exporters is given by 10.37%. Turning to the export intensity, defined as the share of export sales in total sales of a firm, the first column in Table 2 reports the observed distribution across all US exporters. It shows that the large bulk of exporters sells little to nothing in the export markets, whereas some firms have much higher export intensity. Bernard *et al.* (2003) nicely replicate this feature of the data, in particular the lower tail of the distribution. As can be seen from the second and third column in Table 2, however, our model can explain the empirical distribution of export intensities at least as well. For the sake of interest we also report the computed distribution of export intensity across Canadian firms in column 4.

Insert Table 2 about here.

We can further illustrate more particular behaviors of our model that have been emphasized in Section 3, namely the ranking of the export and domestic cutoffs, and the distribution of average markups across regions. We have shown in (23) that m^x_{rs} need not be smaller than m^d_s in our model, as exporters from low wage regions may find it easier to break even in high wage markets than domestic entrants in those regions, despite the existence of trade costs. Using the data for m^d_s, our estimation procedure yields $\widehat{w}_r$, $\widehat{\gamma}$ and $\widehat{\theta}$, which allows us to recover m^x_{rs} for all r and $s \neq r$. For 451 out of 6,889 cases (i.e., about 6.5%) we find that $m^x_{rs} > m^d_s$.

Insert Figure 1 about here.

The bottom-left part of Figure 1 illustrates the case where export cutoffs m^x_{rs} are larger than domestic cutoffs m^d_s. As can be seen from (23), the intuition underlying this result is that wage differences dominate trade cost differences, thereby making it easier for firms in region r to sell to region s. Two examples where this occurs are exports from Nevada to California, and exports from Québec to New York. In the former case, exporters from Nevada can break even in California more easily than Californian entrants, because they can take advantage of the relatively low wages that prevail in Nevada. The exporters from Nevada, of course, face trade costs, but these costs are relatively small as the two states are adjacent to each other. In the latter case, trade costs between Québec and New York are larger. Still, exporters from Québec can break even in New York more easily, because the wage difference is even larger than in the case of California and Nevada. Observe from the top-left panel that export cutoffs need not exceed domestic cutoffs as trade cost differences can dominate wage differences. In that case, we get the 'standard' cutoff ranking. Note also that Melitz and Ottaviano (2008), where wages are equalized, cannot generate such diverse patterns as in Figure 1 as they restrict their attention to the vertical axis.

5 Counterfactual analysis

Having estimated the gravity equation system and having assessed the estimated model, we now conduct a counterfactual experiment. In particular, we consider a hypothetical trade integration scenario where all trade barriers generated by the Canada-US border are completely eliminated (Anderson and van Wincoop, 2003). This allows us to measure the border effects that convey how trade flows would be affected by such a trade integration. These effects can be decomposed into a pure border effect, relative and absolute wage effects, and a selection effect. This decomposition is useful as it highlights each channel that affects trade flows. We finally quantify the impacts of trade integration on regional aggregates (Eaton and Kortum, 2002; Bernard *et al.*, 2003; Del Gatto *et al.*, 2006). Unlike existing studies, we consider changes in wages, average productivity, weighted averages of markups, consumption diversity, and welfare within a single framework.

5.1 The impacts on regional trade flows

To quantify changes in regional trade flows we define *bilateral border effects* as the ratio of trade flows from r to s in a borderless world to those in a world with borders:

$$B_{rs} \equiv \frac{\widetilde{X}_{rs}}{\widehat{X}_{rs}} = \mathrm{e}^{k\widehat{\theta}b_{rs}} \left(\frac{\widetilde{w}_s/\widetilde{w}_r}{\widehat{w}_s/\widehat{w}_r}\right)^{k+1} \left(\frac{\widetilde{w}_r}{\widehat{w}_r}\right) \left(\frac{\widetilde{m}_s^d}{m_s^d}\right)^{k+1}, \tag{46}$$

where variables with a tilde refer to values in a borderless world. The value of B_{rs} can be computed as follows. First, we use the estimated wages $\widehat{w}_r$ and the observed cutoffs m_s^d in the presence of the border to obtain the relevant information for the initial fitted trade flows $\widehat{X}_{rs}$ in (46). Second, holding the shape parameter k, the estimated technological possibilities $\widehat{\mu}_r^{\max}$, and trade frictions $(\widehat{\gamma}, \widehat{\theta})$ constant, we solve (43) and (44) by setting $b_{rs} = 0$ for all r and s. This yields the wages $\widetilde{w}_r$ and the cutoffs $\widetilde{m}_s^d$ that would prevail in a borderless world. Plugging these values into (46), we obtain $61 \times 61 = 3721$ bilateral border effects, each of which gives the change in the trade flow from r to s after the border removal.[23]

Insert Table 3 about here.

The bilateral border effects B_{rs} are typically greater than one when regions r and s are in different countries. The reason is that exports from region r to region s partly substitute for domestic sales as international trade frictions are reduced. For analogous reasons, the values of B_{rs} are typically less than one when r and s are in the same country. Table 3 provides some descriptive statistics on the series of computed bilateral border effects for the various specifications given in Table 1. One can see that the different specifications yield almost identically distributed and strongly correlated bilateral border effects.

5.1.1 Decomposing bilateral border effects

What drives bilateral border effects? As can be seen from expression (46), B_{rs} can be decomposed into the following four components: (i) a *pure border effect* $\mathrm{e}^{k\widehat{\theta} b_{rs}}$; (ii) a *relative wage effect* $\Delta(w_s/w_r) \equiv [(\widetilde{w}_s/\widetilde{w}_r)/(\widehat{w}_s/\widehat{w}_r)]^{k+1}$; (iii) an *absolute wage effect* $\Delta w_r \equiv \widetilde{w}_r/\widehat{w}_r$; and (iv) a *selection effect* $\Delta m_s^d \equiv \left(\widetilde{m}_s^d/m_s^d\right)^{k+1}$. Tables 4 and 5 provide two examples of this decomposition. Depending on the origin and the destination, we can classify all possible cases into four categories which we discuss in turn: (a) Canada-US bilateral trade; (b) Canada-Canada bilateral trade; (c) US-Canada bilateral trade; and (d) US-US bilateral trade.

Insert Tables 4 and 5 about here.

(a) Canada-US bilateral trade. Table 4 lists the components of B_{rs} for exports from British Columbia (BC) to all Canadian provinces and US states. Consider, for example, the bilateral border effect with Washington (WA). The pure border effect corresponds to the predicted change in bilateral trade flows that would prevail if endogenous changes in wages and cutoffs were not taken into account. In this example, it states that the value of exports from BC to WA would rise by a factor of 4.7862. Yet, the wage in BC rises relative to that in WA after the border removal, and BC firms thus become less competitive in WA due to relatively higher production costs. This change is captured by the relative wage effect which decreases the export value by some 45% in this case. The absolute wage effect, on the contrary, raises the value of BC exports by about 8.78% as the higher wage is reflected in the higher prices. When taken together, these two wage effects reduce the pure bilateral border effect from BC to WA by about 40.9% (as $0.5429 \times 1.0878 = 0.5906$). Put differently, neglecting the endogenous reaction of wages to the border removal leads to overstating

[23] We could compute 83×83 bilateral border effects, but in the remainder of this paper we concentrate on the effects of the hypothetical border removal for the 61 regions in Canada and in the US only.

the bilateral border effect by almost 41%. Finally, there is a selection effect. The border removal reduces the cutoff marginal labor requirement that firms need to match to survive in WA. In other words, trade integration induces tougher selection and makes it harder for BC firms to sell in WA. This selection effect decreases the export value by a factor of 0.8065, i.e., it further reduces the bilateral border effect by 19.4%. Putting together the different components, the bilateral border effect is then given by $4.7862 \times 0.5429 \times 1.0878 \times 0.8065 = 2.2799$, which is more than half the size of the pure border effect without endogenous wage and productivity responses.

The top-left panel of Figure 2 depicts the distributions of wage and selection effects for 510 bilateral trade flows from Canadian provinces to US states. The solid line is the product of the relative and absolute wage effects, whereas the dashed line corresponds to the selection effect. Values are reported on the same log scale across the different panels of Figure 2 to allow for a direct comparison of magnitude among the four categories. Whenever the effects are larger (smaller) than one they imply higher (lower) bilateral border effects, and the further away they are from one the larger is their impact on trade flows. In accordance with the BC-WA example, the top-left panel of Figure 2 shows that wage effects strongly dampen export values because Canadian provinces experience significant wage increases relative to US states. Neglecting endogenous wage changes would therefore lead to strong upward biases when assessing changes in trade flows from Canada to the US. Neglecting endogenous productivity changes would also generate upward biases in bilateral border effects, but these biases are somewhat smaller because little selection is induced in the already competitive US markets by the border removal.

Insert Figure 2 about here.

(b) Canada-Canada bilateral trade. Trade flows between regions within the same country would also be affected by the border removal. Consider, for example, exports from BC to Ontario (ON) in Table 4. There is, of course, no pure border effect for this intranational trade flow, but due to the endogenous changes in wages and cutoffs we find a bilateral border effect equal to $1 \times 0.9461 \times 1.0878 \times 0.2835 = 0.2918$. The border removal thus reduces the value of exports from BC to ON by 70.8%. Note that wages in BC rise relative to those in ON, which provides BC firms with a cost disadvantage and per se decreases exports to ON by around 5.4%. The main effect at work in this case, however, is the tougher selection in ON due to the increased presence of more productive US firms. This makes it much harder for BC firms to sell in ON and reduces the export value by more than 70%.[24]

The top-right panel of Figure 2 shows the distributions of wage and selection effects across all 100 bilateral trade flows within Canada. It conveys a message that is similar to the specific

[24]The induced selection effects are also visible in the bilateral border effect of BC with itself. The value of local sales by BC firms drops by 72.2%, which is caused by the now much tougher selection in that market.

BC-ON example. As the hypothetical border removal would induce strong selection effects in the Canadian markets, it is crucial to take these endogenous productivity changes into account since otherwise there would be strong upward biases in bilateral border effects. Neglecting endogenous wage changes would also affect bilateral border effects, but for intranational trade flows their impacts are somewhat smaller and can go in either direction.

(c) US-Canada bilateral trade. Table 5 provides the B_{rs} for exports from New York (NY) to all Canadian provinces and US states. Consider, for example, exports from NY to Québec (QC), which would rise by a factor of $4.7862 \times 1.6444 \times 1.0033 \times 0.3658 = 2.8882$. In this example, wages in QC rise relative to those in NY, which gives NY firms a relative cost advantage and per se boosts export values, whereas the tougher selection in QC makes market penetration by NY firms more difficult, which per se reduces export values. The bottom-left panel of Figure 2, which shows the distributions of wage and selection effects across all 510 trade flows from US states to Canadian provinces, confirms this pattern. Put differently, neglecting endogenous wages leads to downward biases, whereas there are strong upward biases from neglecting endogenous productivity responses.

(d) US-US bilateral trade. Finally, within the US there are only small effects on trade flows. For example, exports from NY to California (CA) in Table 5 change little after the border removal ($1 \times 0.9719 \times 1.0033 \times 0.9842 = 0.9597$). The explanation is that CA is large and far away from the border, so that little additional selection is induced there, while the wage in NY rises only slightly when compared to that in CA. Similarly, there is also only a slight reduction of local sales of NY firms (by 6.4%) since the wage and the cutoff in NY are virtually unaffected by the relatively small Canadian economy.[25] The bottom-right panel of Figure 2 confirms that both wage and selection effects are quite small for the 2601 flows within the US. Whereas selection reduces trade flows overall, the impacts of endogenous wage responses can go either way.

Insert Table 6 about here.

Table 6 provides a summary of these four types of trade flows. As can be seen, both relative wage and selection effects are crucial for assessing how trade flows would change after the border removal. In particular, we find that Canada-US flows are dampened by 34% due to relative wage increases in Canada, whereas US-Canada flows are reduced by 56% due to productivity increases driven by firm selection in Canada. Hence, ignoring endogenous wage and productivity responses can lead to substantial biases in predicted trade flows and these biases depend systematically on the origin and the destination of the flows.

[25]For similar reasons there is also only a slight reduction of local sales of NY firms in their home market (by about 3.4%), which is due to the fact that the relatively small size of the Canadian economy does not induce much firm selection in NY.

5.1.2 Regional and national border effects

In his seminal work on border effects, McCallum (1995) finds that, conditional on regional GDP and distance, trade between Canadian provinces exceeds by roughly 22 times trade between Canadian provinces and US states. Anderson and van Wincoop (2003) argue that this estimate is substantially upward biased due to the omission of general equilibrium conditions. They find that, on average, the border increases trade between Canadian provinces 'only' by a factor of 10.7 when compared to trade with US states. The corresponding number for the US is 2.24.

What does our approach, which adds endogenous wages and firm selection to the analysis, predict for the impacts of the border removal on *overall* Canadian and US trade flows? To evaluate this, we need to aggregate bilateral border effects first at the regional and then at the national level. We define the *regional border effect* for Canadian province r as follows:

$$B_r \equiv \frac{\sum_{s\in\mathrm{US}} \widetilde{X}_{rs} / \sum_{s\in\mathrm{US}} \widehat{X}_{rs}}{\sum_{s\in\mathrm{CAN}} \widetilde{X}_{rs} / \sum_{s\in\mathrm{CAN}} \widehat{X}_{rs}} = \frac{\sum_{s\in\mathrm{US}} \lambda_{rs}^{\mathrm{US}} B_{rs}}{\sum_{s\in\mathrm{CAN}} \lambda_{rs}^{\mathrm{CAN}} B_{rs}},$$

where $\lambda_{rs}^{\mathrm{US}} = \widehat{X}_{rs} / \sum_{s\in\mathrm{US}} \widehat{X}_{rs}$ and $\lambda_{rs}^{\mathrm{CAN}} = \widehat{X}_{rs} / \sum_{s\in\mathrm{CAN}} \widehat{X}_{rs}$ are the fitted trade shares. The numerator is the trade weighted average of international bilateral border effects, whereas the denominator is the trade weighted average of the intranational B_{rs}. The regional border effects B_r thus summarize by how much cross-border trade would rise as compared to domestic trade for each Canadian province and US state.

Using these regional border effects we can then compute the *national border effect* for Canada as follows:

$$B_{\mathrm{CAN}} \equiv \frac{\sum_{r\in\mathrm{CAN}} \sum_{s\in\mathrm{US}} \widetilde{X}_{rs} / \sum_{r\in\mathrm{CAN}} \sum_{s\in\mathrm{US}} \widehat{X}_{rs}}{\sum_{r\in\mathrm{CAN}} \sum_{s\in\mathrm{CAN}} \widetilde{X}_{rs} / \sum_{r\in\mathrm{CAN}} \sum_{s\in\mathrm{CAN}} \widehat{X}_{rs}} = \frac{1}{K_{\mathrm{CAN}}} \sum_{r\in\mathrm{CAN}} B_r,$$

where $K_{\mathrm{CAN}} = 10$ is the number of Canadian provinces. An analogous definition applies to the US. Using this approach we obtain $B_{\mathrm{CAN}} = 7.0831$ and $B_{\mathrm{US}} = 2.9535$.[26]

Endogenous wages and productivity are crucial for explaining the difference between our results and those of Anderson and van Wincoop (2003). The border removal breaks the zero expected profit conditions in all regions. To recover zero expected profits, wages and productivity in Canada must rise relatively to those in the US, as will be shown in more detail in the next subsection. In Anderson and van Wincoop (2003), who abstract from the direct impact of endogenous wages on

[26] Strictly speaking, our definition of the national border effect differs slightly from that of Anderson and van Wincoop (2003). When using their definition in terms of geometric means (see Feenstra, 2004), we obtain 6.88 for Canada and 3.33 for the US. The advantage of our definition is that it precisely measures the (multiplicative) impacts of the border removal on trade flows. Let us emphasize here that our primary objective is not to 'downsize the border effect' but rather to understand how it is affected by endogenous wages and cutoffs. There are many competing explanations for why measured border effects may be too large (see, e.g., Yi, 2010, for the impacts of trade in intermediate goods).

the predicted trade flows, the measured Canadian border effect is therefore overstated, because the export dampening effects of the higher relative wage are not taken into account. The measured US border effect is understated for analogous reasons. This may explain why Anderson and van Wincoop (2003) find more dissimilar national border effects (10.7 for Canada and 2.24 for the US). The change in productivity also affects national border effects. The border removal intensifies competition in all markets and, therefore, bilateral border effects for all pairs of regions are reduced. A model that does not take such effects into account will therefore incorrectly assess the impacts of the border removal on changes in trade flows.

5.2 The impacts on key economic aggregates

We finally investigate how trade integration would affect other key economic aggregates at the regional level. More specifically, we describe the impacts of the hypothetical border removal on wages, average productivity, weighted averages of markups, and welfare. Table 7 reports the regional changes that would occur after eliminating all trade barriers generated by the border.

Insert Table 7 about here.

Which regions would be affected the most? To explore the patterns of these hypothetical changes, we can use a simple OLS approach, where we regress these changes on two crucial regional characteristics: *geography and size*. The former dimension is captured by (the log of) the distance of region r to the closest foreign region, and the latter by (the log of) population size L_r. We further include a US dummy to pick up overall pattern differences between Canada and the US. Observe that a multivariate regression analysis allows us to address in a simple way questions like whether the border removal would mainly affect regions closer to the border or smaller regions.

Table 8 reports the results. Starting with the wage changes, the first specification confirms the aforementioned wage convergence between Canada and the US, because the dummy variable is significantly negative and US wages prior to the border removal are higher than those in Canada. For Canadian provinces the wage changes range from 2.45% for Prince Edward Island to 8.78% for British Columbia, whereas those in the US are much smaller.[27] Apart from this wage convergence between the two countries, we find that regions further away from the border tend to experience smaller wage increases, and that there is a positive and significant relationship with population size. These findings are likely to be driven by the fact that the 10 Canadian provinces display little variation in their distance to the border (at least when compared to US states). In the second specification we therefore consider interaction terms of our proxies for *geography* and *size* with the US dummy in order to capture parameter heterogeneity. This specification indeed shows that wage increases are stronger in *smaller* US states, thus favoring wage convergence across them, whereas

[27] This result is in line with result (iii) proved in the two-region setting in Section 3.4. Note that all wages are expressed relative to that in Alabama, which we set to one by choice of numeraire.

the opposite holds for Canadian provinces. The elasticity of wage change with respect to distance to the border is negative in both countries, but more so in Canada. The intuition for this result is that for the US, being a much larger market, proximity to the new market opportunities matters less than for Canada. Finally, we report at the bottom of Table 8 beta coefficients obtained from the first specification when restricting estimations to US states.[28] These beta coefficients reveal that distance to the border is the more important determinant of the regional variation in wage changes. In fact, this measure of *geography* is more than twice as important as that of *size*.

Insert Table 8 about here.

Looking at other aggregate changes reveals a similar pattern. Predicted cutoff changes are negative for all Canadian provinces and US states, which shows that removing the border induces tougher selection and increases average productivity in all regions.[29] The national productivity gain in Canada, with weights given by regions' share of surviving firms, would be 6.78%, whereas in the US it is much smaller and amounts to only 0.32%. Clearly, since Canada is the smaller economy with less selection prior to the border removal, trade integration has more substantial consequences there. Still, across US states we find stronger productivity gains in smaller regions and in regions closer to the border, with *geography* adding more to our understanding of the regional variation than *size*. The weighted averages of markups fall in all regions, but the reductions in Canada (where they fall between 4.65% and 14.8%) are more substantial than in the US (0.07% to 2.56%). These pro-competitive effects arise because the border removal increases the share of firms engaged in cross-border transactions. More firms compete in each market, especially in the Canadian one, after the border removal which puts downward pressure on markups. Lastly, turning to the welfare impacts of the border removal, we know from our model that a decrease in the cutoff, a lower weighted average of markups, and greater consumption diversity in a region translate into regional welfare gains. As can be seen from Table 7, welfare gains are larger in Canada than in the US, and they range from about 4.87% to about 17.42%.[30] The corresponding range for the US is from 0.07% to 2.62%. The welfare effects are thus again more pronounced in Canada than in the US, and welfare gains in the US are larger in smaller regions close to the border.

[28] The value of -0.0382 for the beta coefficient on size means that a one standard deviation increase in regional size lowers the regional wage change by 3.82% standard deviation points.

[29] Quite naturally, the hypothetical border removal between Canada and the US hurts the ROW countries, who see their cutoffs marginally increase. Results are available upon request.

[30] Table 7 reports cardinal percentage changes in welfare. Therefore, they are sensitive to a monotonically increasing transformation of the utility function. However, their ranking reported in the last column of the table is invariant to such a transformation.

6 Conclusions

We have developed a new general equilibrium model of trade that accommodates the key qualitative features of the recent workhorse trade models. In particular, larger regions have higher wages as in Krugman (1980), higher aggregate productivity as in Melitz (2003), and lower markups as in Krugman (1979). All these variables, as well as product diversity, do respond to changes in trade costs, thus making our framework well suited to simulating the impacts of trade integration. To this end, we have structurally estimated a gravity equation subject to general equilibrium conditions. Although our iterative estimation procedure requires some customized programming, it has the advantage of yielding estimates that take into account all general equilibrium effects. Contrary to the conventional fixed effects approach, our framework can thus capture responses of all endogenous variables at the regional level.

We have used the estimated model to simulate the impacts of removing all trade barriers generated by the Canada-US border. The counterfactual analysis reveals that disregarding endogenous wage and productivity responses can lead to substantial biases in predicted trade flows, and that these biases systematically depend on the origin and the destination of the flows. The reasons are that the border removal would raise relative wages in Canadian provinces, and that productivity would rise more in Canadian provinces than in US states. Furthermore, we have shown that the border removal would lead to lower weighted averages of markups, greater consumption diversity, and higher welfare in all provinces and states, but particularly so in Canada. Hence, in accord with our comparative statics results, trade integration favors regional convergence among Canadian provinces and US states.

Although this paper has focused on the border removal, our model can be applied to various other issues. We could, for example, investigate a scenario where trade costs decrease only for a single pair of regions, say, as the result of an infrastructure project. Policymakers at the federal level would certainly be interested in how such a project affects other regions, and our framework can shed light on such questions. As our structural estimation reveals region-specific technological possibilities, another striking counterfactual exercise would be to examine how their changes would spread across space. We could also quantify the effects of narrowing the technology gap between Canada and the US that still exists according to our estimation.

Our framework can be further extended in many directions. An obvious extension would be to incorporate differential factor proportions in order to cope with a broader international setting including North-South trade. Another possible extension would be to endogenize regional populations by allowing for interregional and international migration based on utility maximization. Taking this road could give rise to a new generation of spatial economics in which theory, structural estimation, and counterfactual experiments are tightly linked.

Acknowledgements. We thank Thomas Chaney, Robert C. Johnson, Alain Gabler, Marc Melitz, Gianmarco Ottaviano, Esteban Rossi-Hansberg, as well as seminar and conference participants at Princeton, GRIPS, Vienna University of Economics and Business, TU Dresden, Universität Bayreuth, University of Saint-Etienne, University of Tokyo, Kyoto University, Nagoya City University, Tohoku University, Yokohama National University, UQAM, Nihon University, Keio University, the 2009 International Trade Workshop in Göttingen, the 2009 Applied Regional Science Conference in Yamagata, the 2008 CEPR ERWIT Meetings in Appenzell, the 2008 Trade and Geography Workshop in Nagoya, the 2008 Annual Meeting of the German Economics Association in Graz, and the 2008 New Economic Geography Workshop in Passau for helpful comments and suggestions. Behrens is holder of the *Canada Research Chair in Regional Impacts of Globalization*. Financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada is gratefully acknowledged. Behrens also acknowledges financial support from FQRSC Québec (Grant NP-127178). Murata gratefully acknowledges financial support from the Japan Society for the Promotion of Science (17730165) and MEXT.ACADEMIC FRONTIER (2006-2010). The usual disclaimer applies.

References

[1] Alvarez, F.E. and R.E. Lucas (2007) General equilibrium analysis of the Eaton-Kortum model of international trade, *Journal of Monetary Economics* 54, 1726-1768.

[2] Anderson, J.E. and E. van Wincoop (2003) Gravity with gravitas: A solution to the border puzzle, *American Economic Review* 93, 170-192.

[3] Anderson, J.E. and E. van Wincoop (2004) Trade costs, *Journal of Economic Literature* 42, 691-751.

[4] Arkolakis, C., S. Demidova, P.J. Klenow and A. Rodríguez-Clare (2008) Endogenous variety and the gains from trade, *American Economic Review* 98, 444-450.

[5] Aw, B., S. Chung and M. Roberts (2000) Productivity and turnover in the export market: Micro-level evidence from the Republic of Korea and Taiwan (China), *World Bank Economic Review* 14, 65-90.

[6] Badinger, H. (2007) Has the EU's Single Market Programme fostered competition? Testing for a decrease in mark-up ratios in EU industries, *Oxford Bulletin of Economics and Statistics* 69, 497-519.

[7] Balistreri, E.J. and R. Hillberry (2007) Structural estimation and the border puzzle, *Journal of International Economics* 72, 451-463.

[8] Behrens, K. and Y. Murata (2007) General equilibrium models of monopolistic competition: A new approach, *Journal of Economic Theory* 136, 776-787.

[9] Bergstrand, J.H. (1985) The gravity equation in international trade: Some microeconomic foundations and empirical evidence, *Review of Economics and Statistics* 67, 474-481.

[10] Bernard, A.B., J. Eaton, J.B. Jensen and S. Kortum (2003) Plants and productivity in international trade, *American Economic Review* 93, 1268-1290.

[11] Bernard, A.B., J.B. Jensen, S.J. Redding and P.K. Schott (2007a) Firms in international trade, *Journal of Economic Perspectives* 21, 105-130.

[12] Bernard, A.B., J.B. Jensen and P.K. Schott (2009) Importers, exporters and multinationals: A portrait of firms in the US that trade goods. In T. Dunne, J.B. Jensen and M.J. Roberts (eds.), *Producer Dynamics: New Evidence from Micro Data*, NBER Book Series Studies in Income and Wealth. University of Chicago Press.

[13] Bernard, A.B., S.J. Redding and P.K. Schott (2007b) Comparative advantage and heterogeneous firms, *Review of Economic Studies* 73, 31-66.

[14] Broda, C. and D. Weinstein (2006) Globalization and the gains from variety, *Quarterly Journal of Economics* 121, 541-585.

[15] Chaney, T. (2008) Distorted gravity: The intensive and extensive margins of international trade, *American Economic Review* 98, 1707-1721.

[16] Corless, R.M., G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey and D.E. Knuth (1996) On the Lambert W function, *Advances in Computational Mathematics* 5, 329-359.

[17] Del Gatto, M., G. Mion and G.I.P. Ottaviano (2006) Trade integration, firm selection and the cost of non-Europe, *CEPR Discussion Paper* #5730.

[18] Demidova, S. (2008) Productivity improvements and falling trade costs: Boon or bane?, *International Economic Review* 49, 1437-1462.

[19] Disdier, A.-C. and K. Head (2008) The puzzling persistence of the distance effect on bilateral trade, *Review of Economics and Statistics* 90, 37-48.

[20] Eaton, J. and S. Kortum (2002) Technology, geography and trade, *Econometrica* 70, 1741-1779.

[21] Feenstra, R.E. (2004) *Advanced International Trade*. Princeton, NJ: Princeton Univ. Press.

[22] Feenstra, R.E. and D.E. Weinstein (2010) Globalization, markups, and the U.S. price level, *NBER Working Paper* #15749.

[23] Foster, L., J.C. Haltiwanger and C. Syverson (2008) Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?, *American Economic Review* 98, 394-425.

[24] Helpman, E., M.J. Melitz and Y. Rubinstein (2008) Estimating trade flows: Trading partners and trading volumes, *Quarterly Journal of Economics* 123, 441-487.

[25] Krugman, P.R. (1979) Increasing returns, monopolistic competition, and international trade, *Journal of International Economics* 9, 469-479.

[26] Krugman, P.R. (1980) Scale economies, product differentiation and the pattern of trade, *American Economic Review* 70, 950-959.

[27] McCallum, J.M. (1995) National borders matter: Canada-US regional trade patterns, *American Economic Review* 85, 615-623.

[28] Melitz, M.J. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* 71, 1695-1725.

[29] Melitz, M.J. and G.I.P. Ottaviano (2008) Market size, trade, and productivity, *Review of Economic Studies* 75, 295-316.

[30] Pavcnik, N. (2002) Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants, *Review of Economic Studies* 69, 245-276.

[31] Redding, S.J. and A.J. Venables (2004) Economic geography and international inequality, *Journal of International Economics* 62, 53-82.

[32] Reza, F.M. (1994) *An Introduction to Information Theory*. Mineola, NY: Dover Publications, Inc.

[33] Syverson, C. (2004) Market structure and productivity: A concrete example, *Journal of Political Economy* 112, 1181-1222.

[34] Syverson, C. (2007) Prices, spatial competition and heterogeneous producers: An empirical test, *Journal of Industrial Economics* 55, 197-222.

[35] Trefler, D. (2004) The long and short of the Canada-US free trade agreement, *American Economic Review* 94, 870-895.

[36] Tybout, J. (2003) Plant and firm-level evidence on new trade theories. In: Choi, E.K. and J. Harrigan (eds.) *Handbook of International Trade*. Oxford: Basil-Blackwell, pp. 388-415.

[37] Yi, K.-M. (2010) Can multistage production explain the home bias in trade, *American Economic Review* 100, 364-393.

Appendix A: Proofs and computations

A.1. Derivation of the demand functions (2). Letting λ stand for the Lagrange multiplier, the first-order condition for an interior solution to the maximization problem (1) satisfies:

$$\alpha \mathrm{e}^{-\alpha q(i)} = \lambda p(i), \quad \forall i \in \Omega \tag{47}$$

and the budget constraint $\int_\Omega p(k)q(k)\mathrm{d}k = E$. Taking the ratio of (47) with respect to i and j yields

$$q(i) = q(j) + \frac{1}{\alpha}\ln\left[\frac{p(j)}{p(i)}\right] \quad \forall i, j \in \Omega.$$

Multiplying this expression by $p(j)$ and integrating with respect to $j \in \Omega$, we obtain

$$q(i)\int_\Omega p(j)\mathrm{d}j = \underbrace{\int_\Omega p(j)q(j)\mathrm{d}j}_{\equiv E} + \frac{1}{\alpha}\int_\Omega \ln\left[\frac{p(j)}{p(i)}\right]p(j)\mathrm{d}j \tag{48}$$

which, letting $\int_\Omega p(j)\mathrm{d}j \equiv N\bar{p}$, can be rewritten as follows:

$$\begin{aligned} q(i) &= \frac{E}{N\bar{p}} - \frac{1}{\alpha}\ln p(i) + \frac{1}{\alpha N\bar{p}}\int_\Omega \ln\left[p(j)\right]p(j)\mathrm{d}j \\ &= \frac{E}{N\bar{p}} - \frac{1}{\alpha}\ln\left[\frac{p(i)}{N\bar{p}}\right] + \frac{1}{\alpha}\int_\Omega \ln\left[\frac{p(j)}{N\bar{p}}\right]\frac{p(j)}{N\bar{p}}\mathrm{d}j. \end{aligned}$$

Factorizing α and using the definition of h then yields the demand function (2).

A.2. Derivation of (8) and properties of W. Using $p^d = m^d w$, the first-order conditions (7) can be rewritten as

$$\ln\left[\frac{m^d w}{p(m)}\right] = 1 - \frac{mw}{p(m)}.$$

Taking the exponential of both sides and rearranging terms, we have

$$\mathrm{e}\frac{m}{m^d} = \frac{mw}{p(m)}\mathrm{e}^{\frac{mw}{p(m)}}.$$

Noting that the Lambert W function is defined as $\varphi = W(\varphi)\mathrm{e}^{W(\varphi)}$ and setting $\varphi = \mathrm{e}m/m^d$, we obtain $W(\mathrm{e}m/m^d) = mw/p(m)$, which implies $p(m)$ as given in (8). The expressions for quantities $q(m) = (1/\alpha)\left[1 - mw/p(m)\right]$ and operating profits $\pi(m) = \left[p(m) - mw\right]q(m)$ are then straightforward to compute.

Turning to the properties of the Lambert W function, $\varphi = W(\varphi)\mathrm{e}^{W(\varphi)}$ implies that $W(\varphi) \geq 0$ for all $\varphi \geq 0$. Taking logarithms on both sides and differentiating yield

$$W'(\varphi) = \frac{W(\varphi)}{\varphi[W(\varphi)+1]} > 0$$

for all $\varphi > 0$. Finally, we have $0 = W(0)\mathrm{e}^{W(0)}$, which implies $W(0) = 0$; and $\mathrm{e} = W(\mathrm{e})\mathrm{e}^{W(\mathrm{e})}$, which implies $W(\mathrm{e}) = 1$.

A.3. Existence and uniqueness of the equilibrium cutoff m^d. To see that there exists a unique equilibrium cutoff m^d, we apply the Leibniz integral rule to the left-hand side of (11) and use $W(\mathrm{e}) = 1$ to obtain

$$\frac{\mathrm{e}L}{\alpha(m^d)^2}\int_0^{m^d} m^2\left(W^{-2}-1\right)W'\mathrm{d}G(m) > 0,$$

where the sign comes from $W' > 0$ and $W^{-2} \geq 1$ for $0 \leq m \leq m^d$. Hence, the left-hand side of (11) is strictly increasing. This uniquely determines the equilibrium cutoff m^d, because

$$\lim_{m^d\to 0}\int_0^{m^d} m\left(W^{-1}+W-2\right)\mathrm{d}G(m) = 0 \quad \text{and} \quad \lim_{m^d\to \infty}\int_0^{m^d} m\left(W^{-1}+W-2\right)\mathrm{d}G(m) = \infty.$$

A.4. Market size, the equilibrium cutoff, and the mass of entrants. Differentiating (11) and using the Leibniz integral rule, we readily obtain

$$\frac{\partial m^d}{\partial L} = -\frac{\alpha F\left(m^d\right)^2}{\mathrm{e}L^2}\left[\int_0^{m^d} m^2\left(W^{-2}-1\right)W'\mathrm{d}G(m)\right]^{-1} < 0,$$

because $W' > 0$ and $W^{-2} \geq 1$ for $0 \leq m \leq m^d$. Differentiating (13) with respect to L yields

$$\frac{\partial N^E}{\partial L} = \frac{F(N^E)^2}{L^2}\left\{1 - \frac{\mathrm{e}L^3}{\alpha F\left(m^d\right)^2}\left[\int_0^{m^d} m^2 W'\mathrm{d}G(m)\right]\frac{\partial m^d}{\partial L}\right\} > 0,$$

where the sign comes from $\partial m^d/\partial L < 0$ as established in the foregoing.

A.5. Indirect utility in the closed economy. To derive the indirect utility, we first compute the (unweighted) average price across all varieties. Multiplying both sides of (7) by $p(i)$, integrating over Ω, and using (4), we obtain

$$\overline{p} = \overline{m}w + \frac{\alpha E}{N},$$

where $\overline{m} \equiv (1/N)\int_\Omega m(j)\mathrm{d}j$ denotes the average marginal labor requirement of the surviving firms. Using $\overline{p}$, expression (5) can be rewritten as $U = N - (\alpha + N\overline{m})/m^d$. When combined with $\overline{m} = [k/(k+1)]m^d$, and with (15) and (16), we obtain the expression for U as given in (17).

A.6. The mass of varieties consumed in the open economy. Using N_r^c as defined in (24), and the export cutoff and the mass of entrants as given by (23) and (32), and making use of the Pareto distribution, we obtain:

$$N_r^c = \frac{\kappa_2}{\kappa_1+\kappa_2}(m_r^d)^k \sum_s \frac{L_s}{F_s(m_s^{\max})^k}\left(\frac{\tau_{rr}}{\tau_{sr}}\frac{w_r}{w_s}\right)^k = \frac{\alpha}{\kappa_1+\kappa_2}\frac{(m_r^d)^k}{\tau_{rr}}\sum_s L_s\tau_{rr}\left(\frac{\tau_{rr}}{\tau_{sr}}\frac{w_r}{w_s}\right)^k \frac{\kappa_2}{\alpha F_s(m_s^{\max})^k}.$$

Using the definition of $\mu_s^{\max}$, and noting that the summation in the foregoing expression appears in the equilibrium relationship (33), we can then express the mass of varieties consumed in region r as given in (34).

A.7. The weighted average of markups in the open economy. Plugging (25) and (26) into the definition (35), the weighted average of markups in the open economy can be rewritten as

$$\overline{\Lambda}_r = \frac{1}{\alpha E_r \sum_s N_s^E G_s(m_{sr}^x)}\sum_s N_s^E \tau_{sr} w_s \int_0^{m_{sr}^x} m\left(W^{-2}-W^{-1}\right)\mathrm{d}G_s(m),$$

where the argument $\mathrm{e}m/m_{sr}^x$ of the Lambert W function is suppressed to alleviate notation. As shown in Appendix B.3, the integral term is given by $\kappa_3(m_s^{\max})^{-k}(m_{sr}^x)^{k+1} = \kappa_3 G_s(m_{sr}^x)m_{sr}^x$. Using this together with (23) and $E_r = w_r$ yields the expression in (35).

A.8. Indirect utility in the open economy. To derive the indirect utility, we first compute the (unweighted) average price across all varieties sold in each market. Multiplying both sides of (22) by $p_{rs}(i)$, integrating over Ω_{rs}, and summing the resulting expressions across r, we obtain:

$$\overline{p}_s \equiv \frac{1}{N_s^c}\sum_r \int_{\Omega_{rs}} p_{rs}(j)\mathrm{d}j = \frac{1}{N_s^c}\sum_r \tau_{rs}w_r\int_{\Omega_{rs}} m_r(j)\mathrm{d}j + \frac{\alpha E_s}{N_s^c},$$

where the first term is the average of marginal delivered costs. Under the Pareto distribution, $\int_{\Omega_{sr}} m_s(j)\mathrm{d}j = N_s^E\int_0^{m_{sr}^x} m\mathrm{d}G_s(m) = [k/(k+1)]m_{sr}^x N_s^E G_s(m_{sr}^x)$. Hence, the (unweighted) average price can be rewritten for region r as follows

$$\overline{p}_r = \frac{1}{N_r^c}\sum_s \tau_{sr}w_s\left(\frac{k}{k+1}\right)m_{sr}^x N_s^E G_s(m_{sr}^x) + \frac{\alpha E_r}{N_r^c} = \left(\frac{k}{k+1}\right)p_r^d + \frac{\alpha E_r}{N_r^c}, \tag{49}$$

where we have used (24) and $p_r^d = \tau_{sr}w_s m_{sr}^x$. Plugging (49) into (20) and using (23), the indirect utility is then given by

$$U_r = \frac{N_r^c}{k+1} - \frac{\alpha}{\tau_{rr}m_r^d}, \tag{50}$$

which together with (34) and (35) yields (36).

A.9. Existence and uniqueness of equilibrium in the two-region case. Under our assumptions on trade costs, the RHS of (41) is non-negative if and only if $\underline{\omega} < \omega < \overline{\omega}$, where $\underline{\omega} \equiv \rho^{1/(k+1)} (\tau_{22}/\tau_{12})^{k/(k+1)}$ and $\overline{\omega} \equiv \rho^{1/(k+1)} (\tau_{21}/\tau_{11})^{k/(k+1)}$. Furthermore, the RHS is strictly decreasing in $\omega \in (\underline{\omega}, \overline{\omega})$ with $\lim_{\omega \to \underline{\omega}+} \text{RHS} = \infty$ and $\lim_{\omega \to \overline{\omega}-} \text{RHS} = 0$. The LHS of (41) is, on the contrary, strictly increasing in $\omega \in (0, \infty)$. Hence, there exists a unique equilibrium relative wage $\omega^* \in (\underline{\omega}, \overline{\omega})$.

A.10. Comparative statics results. In this appendix, we prove the different comparative statics results of Section 3.4.

(i) Assume that $\rho = 1$, $\tau_{12} = \tau_{21} = \tau$, and $\tau_{11} = \tau_{22} = t$. The equilibrium relative wage ω^* is increasing in L_1/L_2 as an increase in L_1/L_2 raises the RHS of (41) without affecting the LHS. This implies that if the two regions have equal technological possibilities and face symmetric trade costs, the larger region has the higher relative wage. Using (39), one can verify that $\omega^{2k+1} = \left(m_2^d/m_1^d\right)^{k+1}$ holds in that case. As $L_1 > L_2$ implies $\omega > 1$, it directly follows that $m_1^d < m_2^d$. Finally, we show in (34)–(36) that a lower cutoff maps into greater consumption diversity, lower weighted average of markups and higher welfare.

(ii) Assume next that $L_1 = L_2$, $\tau_{12} = \tau_{21} = \tau$, and $\tau_{11} = \tau_{22} = t$. Since $t < \tau$ holds, the RHS of (41) shifts up as ρ increases, which then also increases ω^*. This implies that if the two regions are of equal size and face symmetric trade costs, the region with the better technological possibilities has the higher wage. Furthermore, evaluate (41) at $\omega = \rho^{1/(k+1)}$. The LHS is equal to $\rho^{k/(k+1)}$, which falls short of the RHS given by ρ (since $\rho > 1$ and $k \geq 1$). Since the LHS is increasing and the RHS is decreasing, it must be that $\omega^* > \rho^{1/(k+1)}$. It is then straightforward to see that $m_1^d < m_2^d$, because we can rewrite (39) as $\omega^{2k+1}/\rho = \left(m_2^d/m_1^d\right)^{k+1}$ and the LHS of this expression must be larger than one since $(\omega^*)^{2k+1} > (\omega^*)^{k+1} > \rho$. It then follows from (34)–(36) that $m_1^d < m_2^d$ implies $N_1^c > N_2^c$, $\overline{\Lambda}_1 < \overline{\Lambda}_2$, and $U_1 > U_2$.

(iii) Assume that $\tau_{12} = \tau_{21} = \tau$ and that $\tau_{11} = \tau_{22} = t$. One can verify that

$$\frac{\partial(\text{RHS})}{\partial\tau} = -\frac{k\rho t^k}{\tau^{k+1}} \frac{L_1}{L_2} \frac{\rho^2 - \omega^{2(k+1)}}{[\omega^{k+1} - \rho(t/\tau)^k]^2} \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} 0 \quad \text{for} \quad \left\{ \begin{array}{c} \underline{\omega} < \rho^{\frac{1}{k+1}} < \omega^* < \overline{\omega} \\ \underline{\omega} < \omega^* = \rho^{\frac{1}{k+1}} < \overline{\omega} \\ \underline{\omega} < \omega^* < \rho^{\frac{1}{k+1}} < \overline{\omega} \end{array} \right\}. \tag{51}$$

When regions are of equal size, but have different technological possibilities ($\rho > 1$), the first case of (51) applies since $\omega^* > \rho^{1/(k+1)}$ as shown in (ii) before. Hence, lower trade costs reduce the relative wage of the more productive region. Furthermore, when regions have the same technological possibilities but different sizes ($L_1 > L_2$), we obtain $\omega^* > \rho^{k/(k+1)} = 1$, so that the first case of (51) applies again. In other words, when regions differ in size or technological possibilities, wages converge as bilateral trade barriers fall. Since $\omega^{2k+1} = \rho \left(m_2^d/m_1^d\right)^{k+1}$ always holds, this wage convergence directly implies (conditional) convergence of the regional cutoff productivities, and thus (conditional) convergence of consumption diversity, weighted averages of markups, and welfare between the two regions.

(iv) Assume finally that $\rho = 1$, $L_1 = L_2$ and $\tau_{11} = \tau_{22} = t$. Better access to the foreign market raises the domestic relative wage, whereas better access from the foreign to the domestic market reduces the domestic relative wage, because (41) implies

$$\frac{\partial(\text{RHS})}{\partial \tau_{12}} < 0 \quad \text{iff} \quad \omega^* < \overline{\omega} \quad \text{and} \quad \frac{\partial(\text{RHS})}{\partial \tau_{21}} > 0 \quad \text{iff} \quad \omega^* > \underline{\omega}.$$

Since $\omega^{2k+1} = \left(m_2^d/m_1^d\right)^{k+1}$ holds, it follows that the region that ends up with the higher wage also ends up with the lower cutoff and, thus, with greater consumption diversity, lower weighted average of markups, and higher welfare.

Appendix B: Integrals involving the Lambert W function

To derive closed-form solutions for various expressions throughout the paper we need to compute integrals involving the Lambert W function. This can be done by using the change in variables suggested by Corless *et al.* (1996, p.341). Let

$$z \equiv W\left(\mathrm{e}\,\frac{m}{I}\right), \quad \text{so that} \quad \mathrm{e}\,\frac{m}{I} = z\mathrm{e}^z, \quad \text{where} \quad I = m_r^d, m_{rs}^x,$$

where subscript r can be dropped in the closed economy. The change in variables then yields $\mathrm{d}m = (1+z)\mathrm{e}^{z-1}I\mathrm{d}z$, with the new integration bounds given by 0 and 1. Under our assumption of a Pareto distribution for productivity draws, the change in variables allows to rewrite integrals in simplified form.

B.1. First, consider the following expression, which appears when integrating firms' outputs:

$$\int_0^I m\left[1 - W\left(\mathrm{e}\frac{m}{I}\right)\right]\mathrm{d}G_r(m) = \kappa_1\left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_1 \equiv k\mathrm{e}^{-(k+1)}\int_0^1 (1-z^2)\left(z\mathrm{e}^z\right)^k \mathrm{e}^z\mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k.

B.2. Second, the following expression appears when integrating firms' operating profits:

$$\int_0^I m\left[W\left(\mathrm{e}\frac{m}{I}\right)^{-1} + W\left(\mathrm{e}\frac{m}{I}\right) - 2\right]\mathrm{d}G_r(m) = \kappa_2\left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_2 \equiv k\mathrm{e}^{-(k+1)}\int_0^1 (1+z)\left(z^{-1}+z-2\right)\left(z\mathrm{e}^z\right)^k \mathrm{e}^z\mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k.

B.3. Third, the following expression appears when deriving the weighted average of firm-level markups:

$$\int_0^I m \left[W\left(\mathrm{e}\frac{m}{I}\right)^{-2} - W\left(\mathrm{e}\frac{m}{I}\right)^{-1} \right] \mathrm{d}G_r(m) = \kappa_3 \left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_3 \equiv k\mathrm{e}^{-(k+1)} \int_0^1 (z^{-2} - z^{-1})(1+z)(z\mathrm{e}^z)^k \mathrm{e}^z \mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k.

B.4. Finally, the following expression appears when integrating firms' revenues:

$$\int_0^I m \left[W\left(\mathrm{e}\frac{m}{I}\right)^{-1} - 1 \right] \mathrm{d}G_r(m) = \kappa_4 \left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_4 \equiv k\mathrm{e}^{-(k+1)} \int_0^1 (z^{-1} - z) \left(z\mathrm{e}^z\right)^k \mathrm{e}^z \mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k. Using the expressions for κ_1 and κ_2, one can verify that $\kappa_4 = \kappa_1 + \kappa_2$.

Appendix C: Open economy equilibrium conditions using the Lambert W function

In this appendix we restate the open economy equilibrium conditions of Section 3 using the Lambert W function.

C.1. Using (25), the labor market clearing condition can be rewritten as follows:

$$N_r^E \left\{ \frac{1}{\alpha} \sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[1 - W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right) \right] \mathrm{d}G_r(m) + F_r \right\} = L_r. \tag{52}$$

C.2. Plugging (25) into (27), zero expected profits require that

$$\frac{1}{\alpha} \sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right)^{-1} + W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right) - 2 \right] \mathrm{d}G_r(m) = F_r. \tag{53}$$

As in the closed economy case, the zero expected profit condition depends solely on the cutoffs m_{rs}^x and is independent of the mass of entrants.

C.3. Finally, the trade balance condition is given by

$$\begin{aligned} N_r^E w_r \sum_{s \neq r} L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right)^{-1} - 1 \right] \mathrm{d}G_r(m) \\ = L_r \sum_{s \neq r} N_s^E \tau_{sr} w_s \int_0^{m_{sr}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{sr}^x}\right)^{-1} - 1 \right] \mathrm{d}G_s(m). \end{aligned} \tag{54}$$

Applying the region-specific Pareto distributions $G_r(m) = (m/m_r^{\max})^k$ to (52)–(54) yields, after some algebra and using the results of Appendix B, expressions (29)–(31) given in the main text.

Appendix D: Calibrating the value of k and generating firm-level variables

For a given value of k we can simulate our model at the firm level by using the estimates from the gravity equation system: the wages ($\widehat{w}_r$), the technological possibilities ($\widehat{\mu}_r^{\max}$), and the trade friction parameters ($\widehat{\gamma}$, $\widehat{\theta}$). These estimates, together with k and data on the domestic cutoffs (m_s^d), provide all the information required to construct the export cutoffs ($\widehat{m}_{rs}^x$). We can then compute the following variables.

Share of exporters. We define the share of exporters in a US state as the share of firms selling to at least one Canadian province or to one country in the ROW. Formally, it is given by $G_r(\max_{s\in\text{CAN,ROW}}\{m_{rs}^x\})/G_r(\max_s\{m_{rs}^x\})$. The share of US exporters is then computed as the weighted average of the states' exporter shares, where the weights are proportional to the mass of surviving firms in each state (see below). The corresponding share of Canadian exporters is defined in an analogous way.

Export intensity. Let $\widehat{\chi}_{rs} = 1$ if $m < \widehat{m}_{rs}^x$ and $\widehat{\chi}_{rs} = 0$ otherwise. The export intensity of a firm in country $I = \text{CAN}, \text{US}$ is defined as

$$\text{expint}_r(m) = \frac{\text{expsls}_r(m)}{\text{domsls}_r(m) + \text{expsls}_r(m)},$$

where domestic and export sales are given by

$$\begin{aligned}\text{domsls}_r(m) &= \sum_{s\in I} \widehat{\chi}_{rs} L_s p_{rs}(m) q_{rs}(m) \\ &= \frac{\widehat{w}_r m}{\alpha} \sum_{s\in I} \widehat{\chi}_{rs} L_s d_{rs}^{\widehat{\gamma}} [W(\text{e}m/\widehat{m}_{rs}^x)^{-1} - 1]\end{aligned}$$

$$\begin{aligned}\text{expsls}_r(m) &= \sum_{s\notin I} \widehat{\chi}_{rs} L_s p_{rs}(m) q_{rs}(m) \\ &= \frac{\widehat{w}_r m}{\alpha} \sum_{s\notin I} \widehat{\chi}_{rs} L_s d_{rs}^{\widehat{\gamma}} \text{e}^{\widehat{\theta} b_{rs}} [W(\text{e}m/\widehat{m}_{rs}^x)^{-1} - 1].\end{aligned}$$

Note that information on α is not required to obtain export intensity, although domestic and export sales depend on α.

Revenue-based productivity. The revenue-based productivity, excluding the labor used for shipping goods, is given by:

$$\begin{aligned}\mathrm{rbprod}_r(m) &= \frac{\mathrm{domsls}_r(m) + \mathrm{expsls}_r(m)}{m \sum_s \widehat{\chi}_{rs} L_s q_{rs}(m)} \\ &= \frac{\mathrm{domsls}_r(m) + \mathrm{expsls}_r(m)}{(m/\alpha) \sum_s \widehat{\chi}_{rs} L_s (1 - W(\mathrm{e}m/\widehat{m}_{rs}^x))},\end{aligned}$$

which is again independent of α.

We can now compute the productivity advantage of exporters. To make the sample representative, we draw firms in all regions in proportion to that region's share of surviving firms in the national number of surviving firms. We know that

$$N_r^p = N_r^E G_r \left(\max_s m_{rs}^x \right) = \frac{\alpha}{\kappa_1 + \kappa_2} L_r \left(\mu_r^{\max} \right)^{-1} \left(\max_s m_{rs}^x \right)^k$$

so that each region's share of surviving firms in country $I = \mathrm{CAN}, \mathrm{US}$ is given by

$$\widehat{\theta}_r = \frac{\widehat{N}_r^p}{\sum_{s \in I} \widehat{N}_s^p} = \frac{L_r \left(\widehat{\mu}_r^{\max} \right)^{-1} \left(\max_j \widehat{m}_{rj}^x \right)^k}{\sum_{s \in I} L_s \left(\widehat{\mu}_s^{\max} \right)^{-1} \left(\max_j \widehat{m}_{sj}^x \right)^k}, \quad r \in I.$$

Note that the foregoing expression is again independent of the unobservable parameter α. For a sample size $N_{CAN} = 10,000$ and $N_{US} = 100,000$, we randomly draw $\mathrm{int}(\widehat{\theta}_s N_{CAN})$ firms for each Canadian province and $\mathrm{int}(\widehat{\theta}_r N_{US})$ firms for each US state from the region-specific productivity distribution, where $\mathrm{int}(\cdot)$ stands for the integer part. This yields a representative sample for each country, while the overall sample respects country's relative sizes in 1993. To calibrate k, we search over the parameter space in order to match the US productivity advantage of exporters generated by our model with the 33% figure reported by Bernard *et al.* (2003).

Table 1: Estimation of the gravity equation system

	Benchmark(1)	Robustness(2)	Robustness(3)	Robustness(4)	Robustness(5)	FixedEffects(6)
Regions (in)	83 (40)	83 (40)	83 (40)	83 (40)	61 (40)	83 (40)
Trade flows	1560	1560	1560	1511	1560	1560
k	7.5	3.6	11.4	7.5	7.5	7.5
Internal dist.	Surface	Surface	Surface	Surface	AvW	Surface
Procedure	Iterative	Iterative	Iterative	Iterative	Iterative	OLS
constant	−4.4584***	−4.4047***	−4.4635***	−4.4112***	−4.4515***	−16.255***
	(0.0384)	(0.0386)	(0.0391)	(0.0364)	(0.0560)	(0.3720)
$\ln d_{rs}$	−1.4457***	−1.5605***	−1.4165***	−1.4659***	−1.1993***	−1.2411***
	(0.0431)	(0.0347)	(0.0452)	(0.0428)	(0.0349)	(0.0417)
$\ln d_{rr}$	1.6384***	1.9939***	1.5408***	1.6614***	1.3591***	—
	(0.0488)	(0.0443)	(0.0492)	(0.0486)	(0.0395)	—
b_{rs}	−1.5657***	−1.5677***	−1.5576***	−1.6627***	−1.6656***	−1.4508***
	(0.0552)	(0.0613)	(0.0557)	(0.0409)	(0.0531)	(0.0654)
$0-$ dummy	−17.392***	−17.449***	−17.375***	—	−17.735***	−16.976***
	(0.1301)	(0.1508)	(0.1394)	—	(0.1354)	(0.1581)
Adjusted R^2	0.9131	0.9110	0.9134	0.7036	0.8923	0.9214

Notes: Bootstrapped standard errors for the parameters obtained using the iterative procedure are given in parenthesis (with 100 replications for Benchmark(1) and for FixedEffects(6), and 40 replications for Robustness(2)–Robustness(5)). All specifications, except Robustness(4), use 1560 trade flows, excluding intraregional flows X_{rr} as in Anderson and van Wincoop (2003). Robustness(4) drops the 49 reported zero trade flows from the sample instead of using a dummy variable for them. 'Surface' refers to the surface-based measure of internal distance of Redding and Venables' (2004), whereas AvW refers to Anderson and van Wincoop's (2003) measure. The convergence criterion for the iterative procedure is based on the difference of norms of the vector of regression coefficients between two successive iterations, with threshold 10^{-12}. Starting points for the iterative solver are obtained using OLS with importer-exporter fixed effects. We choose $w_{\text{Alabama}} \equiv 1$ as numeraire. Results are invariant to that choice. Coefficients significant at 10% level (*), 5% level (**), and 1% level (***).

Table 2: Export intensity distributions

Export intensity (percent)	US Observed % of exporters (1992 Census, BEJK)	US Predicted % of exporters (BEJK model)	US Predicted % of exporters (our model, with border)	Canada Predicted % of exporters (our model, with border)
0-10	66	76	64.83	35.78
10-20	16	19	23.40	9.35
20-30	7.7	4.2	7.56	12.44
30-40	4.4	0.0	1.19	21.31
40-50	2.4	0.0	1.47	10.80
50-60	1.5	0.0	1.02	3.57
60-70	1.0	0.0	0.54	2.12
70-80	0.6	0.0	0.00	2.22
80-90	0.5	0.0	0.00	1.16
90-100	0.7	0.0	0.00	1.25

Notes: Export intensity is defined as in Appendix D as the firm's share of export revenue in total revenue, conditional upon exporting something. Figures in column 2 and 3 are provided by Bernard *et al.* (2003), or BEJK for short. Column 2 reports the observed distribution using 1992 Census of Manufactures data, whereas column 3 provides the simulation results obtained by Bernard *et al.* (2003). Columns 4 and 5 provide our own simulation results with $k = 7.5$ for the US and for Canada.

Table 3: Descriptive statistics for bilateral border effects

Descriptive statistics for bilateral border effect series:					
	Benchmark(1)	Robustness(2)	Robustness(3)	Robustness(4)	Robustness(5)
Minimum	0.2779	0.2890	0.2727	0.2651	0.3796
Maximum	3.9438	4.2580	3.8198	4.3661	4.3155
Geo. Mean	1.1170	1.1320	1.1115	1.1301	1.1363
Std. dev.	0.8325	0.8937	0.8057	0.9453	0.9682
Median	0.9784	0.9810	0.9779	0.9777	0.9708
Skewness	2.1108	2.1527	2.0981	2.1302	2.0268
Kurtosis	5.9050	6.0856	5.8617	5.9626	5.3473
Correlation matrix for bilateral border effect series:					
	(1)	(2)	(3)	(4)	(5)
(1)	1	0.9997	1.0000	1.0000	0.9501
(2)		1	0.9995	0.9998	0.9447
(3)			1	0.9999	0.9510
(4)				1	0.9491
(5)					1

Table 4: Decomposition of bilateral border effects with British Columbia as exporter

	Pure border $e^{\hat{\theta} b_{rs}}$	Rel. wage $\Delta(w_s/w_r)$	Abs. wage Δw_r	Selection Δm_s^d	Bil. border B_{rs}
Importer:	**In Gravity sample**				
Alberta	1.0000	0.6673	1.0878	0.5469	0.3971
British Columbia	1.0000	1.0000	1.0878	0.2555	0.2779
Manitoba	1.0000	0.8503	1.0878	0.3466	0.3206
New Brunswick	1.0000	0.6581	1.0878	0.5615	0.4020
Newfoundland	1.0000	0.6570	1.0878	0.5633	0.4026
Nova Scotia	1.0000	0.6292	1.0878	0.6110	0.4182
Ontario	1.0000	0.9461	1.0878	0.2835	0.2918
Prince Edward Island	1.0000	0.6004	1.0878	0.6673	0.4359
Quebec	1.0000	0.8264	1.0878	0.3658	0.3288
Saskatchewan	1.0000	0.7798	1.0878	0.4080	0.3461
Alabama	4.7862	0.4889	1.0878	0.9826	2.5010
Arizona	4.7862	0.4914	1.0878	0.9732	2.4898
California	4.7862	0.4884	1.0878	0.9842	2.5029
Florida	4.7862	0.4883	1.0878	0.9846	2.5034
Georgia	4.7862	0.4896	1.0878	0.9799	2.4978
Idaho	4.7862	0.5024	1.0878	0.9332	2.4413
Illinois	4.7862	0.4902	1.0878	0.9774	2.4948
Indiana	4.7862	0.4923	1.0878	0.9697	2.4856
Kentucky	4.7862	0.4922	1.0878	0.9702	2.4862
Louisiana	4.7862	0.4888	1.0878	0.9829	2.5013
Maine	4.7862	0.5444	1.0878	0.8023	2.2744
Maryland	4.7862	0.4912	1.0878	0.9738	2.4905
Massachusetts	4.7862	0.4958	1.0878	0.9569	2.4702
Michigan	4.7862	0.5079	1.0878	0.9144	2.4181
Minnesota	4.7862	0.4998	1.0878	0.9424	2.4526
Missouri	4.7862	0.4906	1.0878	0.9759	2.4930
Montana	4.7862	0.5214	1.0878	0.8702	2.3626
New Hampshire	4.7862	0.5037	1.0878	0.9287	2.4357
New Jersey	4.7862	0.4913	1.0878	0.9736	2.4903
New York	4.7862	0.5025	1.0878	0.9329	2.4409
North Carolina	4.7862	0.4924	1.0878	0.9694	2.4852
North Dakota	4.7862	0.5199	1.0878	0.8750	2.3687
Ohio	4.7862	0.4968	1.0878	0.9534	2.4659
Pennsylvania	4.7862	0.4995	1.0878	0.9437	2.4541
Tennessee	4.7862	0.4905	1.0878	0.9765	2.4937
Texas	4.7862	0.4894	1.0878	0.9807	2.4987
Vermont	4.7862	0.5246	1.0878	0.8603	2.3499
Virginia	4.7862	0.4933	1.0878	0.9660	2.4811
Washington	4.7862	0.5429	1.0878	0.8065	2.2799
Wisconsin	4.7862	0.4963	1.0878	0.9550	2.4678
Importer:	**Out of Gravity sample**				
Alaska	4.7862	0.5162	1.0878	0.8868	2.3836
Arkansas	4.7862	0.4905	1.0878	0.9763	2.4935
Colorado	4.7862	0.4942	1.0878	0.9627	2.4772
Connecticut	4.7862	0.4914	1.0878	0.9731	2.4897
Delaware	4.7862	0.4912	1.0878	0.9739	2.4907
Hawaii	4.7862	0.4861	1.0878	0.9932	2.5136
Iowa	4.7862	0.4934	1.0878	0.9657	2.4808
Kansas	4.7862	0.4922	1.0878	0.9701	2.4861
Mississippi	4.7862	0.4894	1.0878	0.9806	2.4986
Nebraska	4.7862	0.4943	1.0878	0.9624	2.4768
Nevada	4.7862	0.4859	1.0878	0.9939	2.5145
New Mexico	4.7862	0.4933	1.0878	0.9662	2.4814
Oklahoma	4.7862	0.4907	1.0878	0.9755	2.4925
Oregon	4.7862	0.5009	1.0878	0.9385	2.4478
Rhode Island	4.7862	0.4917	1.0878	0.9718	2.4882
South Carolina	4.7862	0.4910	1.0878	0.9745	2.4913
South Dakota	4.7862	0.5055	1.0878	0.9227	2.4283
Utah	4.7862	0.4970	1.0878	0.9524	2.4647
West Virginia	4.7862	0.4957	1.0878	0.9573	2.4707
Wyoming	4.7862	0.4945	1.0878	0.9616	2.4758
District of Columbia	4.7862	0.4863	1.0878	0.9923	2.5126

Notes: Border effects are decomposed as indicated by (46).

Table 5: Decomposition of bilateral border effects with New York as exporter

	Pure border $e^{\hat{\theta} b_{rs}}$	Rel. wage $\Delta(w_s/w_r)$	Abs. wage Δw_r	Selection Δm_s^d	Bil. border B_{rs}
Importer:	**In Gravity sample**				
Alberta	4.7862	1.3280	1.0033	0.5469	3.4876
British Columbia	4.7862	1.9899	1.0033	0.2555	2.4409
Manitoba	4.7862	1.6920	1.0033	0.3466	2.8164
New Brunswick	4.7862	1.3095	1.0033	0.5615	3.5309
Newfoundland	4.7862	1.3074	1.0033	0.5633	3.5361
Nova Scotia	4.7862	1.2521	1.0033	0.6110	3.6734
Ontario	4.7862	1.8827	1.0033	0.2835	2.5631
Prince Edward Island	4.7862	1.1948	1.0033	0.6673	3.8284
Quebec	4.7862	1.6444	1.0033	0.3658	2.8882
Saskatchewan	4.7862	1.5518	1.0033	0.4080	3.0398
Alabama	1.0000	0.9728	1.0033	0.9826	0.9590
Arizona	1.0000	0.9778	1.0033	0.9732	0.9546
California	1.0000	0.9719	1.0033	0.9842	0.9597
Florida	1.0000	0.9717	1.0033	0.9846	0.9599
Georgia	1.0000	0.9742	1.0033	0.9799	0.9577
Idaho	1.0000	0.9998	1.0033	0.9332	0.9361
Illinois	1.0000	0.9755	1.0033	0.9774	0.9566
Indiana	1.0000	0.9796	1.0033	0.9697	0.9531
Kentucky	1.0000	0.9794	1.0033	0.9702	0.9533
Louisiana	1.0000	0.9727	1.0033	0.9829	0.9591
Maine	1.0000	1.0834	1.0033	0.8023	0.8721
Maryland	1.0000	0.9774	1.0033	0.9738	0.9549
Massachusetts	1.0000	0.9866	1.0033	0.9569	0.9471
Michigan	1.0000	1.0107	1.0033	0.9144	0.9272
Minnesota	1.0000	0.9946	1.0033	0.9424	0.9404
Missouri	1.0000	0.9763	1.0033	0.9759	0.9559
Montana	1.0000	1.0376	1.0033	0.8702	0.9059
New Hampshire	1.0000	1.0024	1.0033	0.9287	0.9339
New Jersey	1.0000	0.9776	1.0033	0.9736	0.9548
New York	1.0000	1.0000	1.0033	0.9329	0.9359
North Carolina	1.0000	0.9798	1.0033	0.9694	0.9529
North Dakota	1.0000	1.0346	1.0033	0.8750	0.9082
Ohio	1.0000	0.9885	1.0033	0.9534	0.9455
Pennsylvania	1.0000	0.9939	1.0033	0.9437	0.9410
Tennessee	1.0000	0.9760	1.0033	0.9765	0.9562
Texas	1.0000	0.9738	1.0033	0.9807	0.9581
Vermont	1.0000	1.0440	1.0033	0.8603	0.9010
Virginia	1.0000	0.9816	1.0033	0.9660	0.9513
Washington	1.0000	1.0804	1.0033	0.8065	0.8742
Wisconsin	1.0000	0.9876	1.0033	0.9550	0.9462
Importer:	**Out of Gravity sample**				
Alaska	1.0000	1.0273	1.0033	0.8868	0.9139
Arkansas	1.0000	0.9761	1.0033	0.9763	0.9561
Colorado	1.0000	0.9834	1.0033	0.9627	0.9498
Connecticut	1.0000	0.9778	1.0033	0.9731	0.9546
Delaware	1.0000	0.9774	1.0033	0.9739	0.9550
Hawaii	1.0000	0.9673	1.0033	0.9932	0.9638
Iowa	1.0000	0.9818	1.0033	0.9657	0.9512
Kansas	1.0000	0.9794	1.0033	0.9701	0.9532
Mississippi	1.0000	0.9739	1.0033	0.9806	0.9580
Nebraska	1.0000	0.9836	1.0033	0.9624	0.9497
Nevada	1.0000	0.9669	1.0033	0.9939	0.9641
New Mexico	1.0000	0.9815	1.0033	0.9662	0.9514
Oklahoma	1.0000	0.9765	1.0033	0.9755	0.9557
Oregon	1.0000	0.9968	1.0033	0.9385	0.9385
Rhode Island	1.0000	0.9785	1.0033	0.9718	0.9540
South Carolina	1.0000	0.9771	1.0033	0.9745	0.9553
South Dakota	1.0000	1.0059	1.0033	0.9227	0.9311
Utah	1.0000	0.9890	1.0033	0.9524	0.9451
West Virginia	1.0000	0.9863	1.0033	0.9573	0.9473
Wyoming	1.0000	0.9840	1.0033	0.9616	0.9493
District of Columbia	1.0000	0.9677	1.0033	0.9923	0.9634

Notes: Border effects are decomposed as indicated by (46).

Table 6: Decomposition of average bilateral border effects

	Canada-US	Canada-Canada	US-Canada	US-US
Pure border effect	4.7862	1.0000	4.7862	1.0000
Relative wage effect	0.6630	1.0000	1.5082	1.0000
Absolute wage effect	1.0517	1.0517	1.0021	1.0021
Selection effect	0.9509	0.4387	0.4387	0.9509

Notes: Bilateral border effects are given by expression (46). The geometric mean of those bilateral border effects for, say, Canada-US flows is given by

$$\overline{B} = \left[\prod_{r,s} \mathrm{e}^{k\widehat{\theta} b_{rs}}\right]^{\frac{1}{N}} \times \left[\prod_{r,s} \frac{\widetilde{w}_r}{\widehat{w}_r}\right]^{\frac{1}{N}} \times \left[\prod_{r,s} \left(\frac{\widetilde{w}_s/\widetilde{w}_r}{\widehat{w}_s/\widehat{w}_r}\right)^{k+1}\right]^{\frac{1}{N}} \times \left[\prod_{r,s} \left(\frac{\widetilde{m}_s^d}{m_s^d}\right)^{k+1}\right]^{\frac{1}{N}},$$

where $r \in$ CAN and $s \in$ US, and N denotes the number of Canada-US flows. The 'percentage' change with respect to the 'pure' border effect is then given by for example $(1 - 0.6630) \times 100\% \approx 34\%$ for the relative wage effect for Canada-US flows. Put differently, disregarding this effect overstates changes in Canada-US flows by about 34%.

Table 7: Impacts of removing all trade barriers generated by the Canada-US border

	Wages $\Delta w_r\%$	Cutoffs and Markups $\Delta m_r^d\%$ and $\Delta\overline{\Lambda}_r\%$	Varieties and Welfare $\Delta N_r^c\%$ and $\Delta U_r^*\%$	Rank of $\Delta U_r^*\%$
States/Provinces	**In Gravity sample**			
Alberta	3.7294	-6.8527	7.3569	6
British Columbia	8.7844	-14.8329	17.4162	1
Manitoba	6.7285	-11.7186	13.2741	3
New Brunswick	3.559	-6.5639	7.025	7
Newfoundland	3.5389	-6.5299	6.986	8
Nova Scotia	3.0144	-5.6319	5.968	9
Ontario	8.0778	-13.7817	15.9847	2
Prince Edward Island	2.4483	-4.648	4.8745	10
Quebec	6.3708	-11.1589	12.5606	4
Saskatchewan	5.6477	-10.0108	11.1245	5
Alabama	0	-0.2062	0.2066	55
Arizona	0.0603	-0.3195	0.3205	42
California	-0.0101	-0.1872	0.1876	57
Florida	-0.0127	-0.1823	0.1826	58
Georgia	0.0174	-0.2389	0.2395	52
Idaho	0.323	-0.8102	0.8168	21
Illinois	0.0333	-0.2688	0.2695	51
Indiana	0.0825	-0.361	0.3623	37
Kentucky	0.0796	-0.3557	0.3569	39
Louisiana	-0.0016	-0.2032	0.2036	56
Maine	1.2748	-2.5576	2.6247	11
Maryland	0.0561	-0.3115	0.3125	44
Massachusetts	0.1659	-0.517	0.5197	28
Michigan	0.4509	-1.0477	1.0588	17
Minnesota	0.2612	-0.6951	0.6999	23
Missouri	0.0428	-0.2867	0.2875	48
Montana	0.762	-1.622	1.6487	14
New Hampshire	0.3534	-0.8667	0.8742	19
New Jersey	0.0575	-0.3141	0.3151	43
New York	0.3251	-0.8141	0.8208	20
North Carolina	0.0847	-0.3652	0.3665	36
North Dakota	0.7278	-1.5591	1.5838	15
Ohio	0.1887	-0.5597	0.5628	26
Pennsylvania	0.2529	-0.6795	0.6841	24
Tennessee	0.0392	-0.2799	0.2807	50
Texas	0.0124	-0.2295	0.2301	54
Vermont	0.8344	-1.7549	1.7863	13
Virginia	0.1068	-0.4065	0.4081	34
Washington	1.2419	-2.4981	2.5621	12
Wisconsin	0.1784	-0.5405	0.5435	27
States/Provinces	**Out of Gravity sample**			
Alaska	0.6433	-1.4035	1.4235	16
Arkansas	0.0405	-0.2823	0.2831	49
Colorado	0.1277	-0.4457	0.4476	32
Connecticut	0.0607	-0.3202	0.3212	41
Delaware	0.0554	-0.3101	0.311	45
Hawaii	-0.0668	-0.0806	0.0806	60
Iowa	0.1085	-0.4097	0.4114	33
Kansas	0.0799	-0.3562	0.3575	38
Mississippi	0.0129	-0.2305	0.2311	53
Nebraska	0.13	-0.4499	0.452	31
Nevada	-0.0714	-0.0719	0.0719	61
New Mexico	0.1053	-0.4037	0.4053	35
Oklahoma	0.0454	-0.2915	0.2924	47
Oregon	0.2875	-0.7442	0.7497	22
Rhode Island	0.069	-0.3356	0.3367	40
South Carolina	0.0519	-0.3036	0.3045	46
South Dakota	0.3943	-0.9426	0.9516	18
Utah	0.1951	-0.5717	0.575	25
West Virginia	0.163	-0.5117	0.5144	29
Wyoming	0.1353	-0.46	0.4621	30
District of Columbia	-0.0615	-0.0905	0.0906	59

Notes: See Section 5 for details on computations.

Table 8: Determinants of changes in regional aggregates

	Dependent variable		
	Wages $\Delta w_r\%$	Cutoffs and Markups $\Delta m_r^d\%$ and $\Delta\overline{\Lambda}_r\%$	Varieties and Welfare $\Delta N_r^c\%$ and $\Delta U_r^*\%$
Regressor	**Estimated coefficients specification 1**, (All regions, $N = 61$)		
US-dummy	-4.7784***	8.2041***	-9.2707***
DISTANCE TO BORDER (log)	-0.6134***	1.0506***	-1.1917***
SIZE (log)	0.2152**	-0.3322**	0.4304**
Constant	5.8895***	-10.8866***	11.4409***
Adjusted R^2	0.8428	0.8576	0.8380
Regressor	**Estimated coefficients specification 2**, (All regions, $N = 61$)		
DISTANCE TO BORDER (log)	-1.4251***	2.2409***	-2.8390***
DISTANCE TO BORDER (log) × US-dummy	1.1080***	-1.6535***	2.2384***
SIZE (log)	1.0407***	-1.6957***	2.0511***
SIZE (log) × US-dummy	-1.1124***	1.8286***	-2.1868***
US-dummy	4.2658	-7.7008	8.1381
Constant	-0.8601	1.1810	-1.4803
Adjusted R^2	0.9280	0.9332	0.9262
Regressor	**Beta coefficients specification 1** (Only US States, $N = 51$)		
DISTANCE TO BORDER (log)	-0.0960***	0.1043***	-0.0935***
SIZE (log)	-0.0382**	0.0416**	-0.0372**

Notes: See Section 5 for additional details on computations. Coefficients significant at 10% level (*), 5% level (**), and 1% level (***).

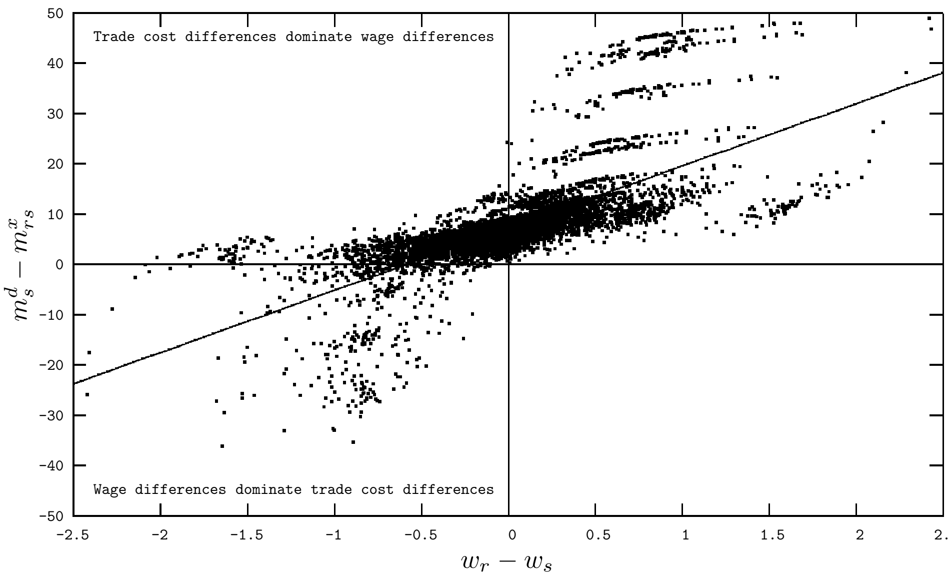

Figure 1. Cutoff rankings in the presence of wage differences

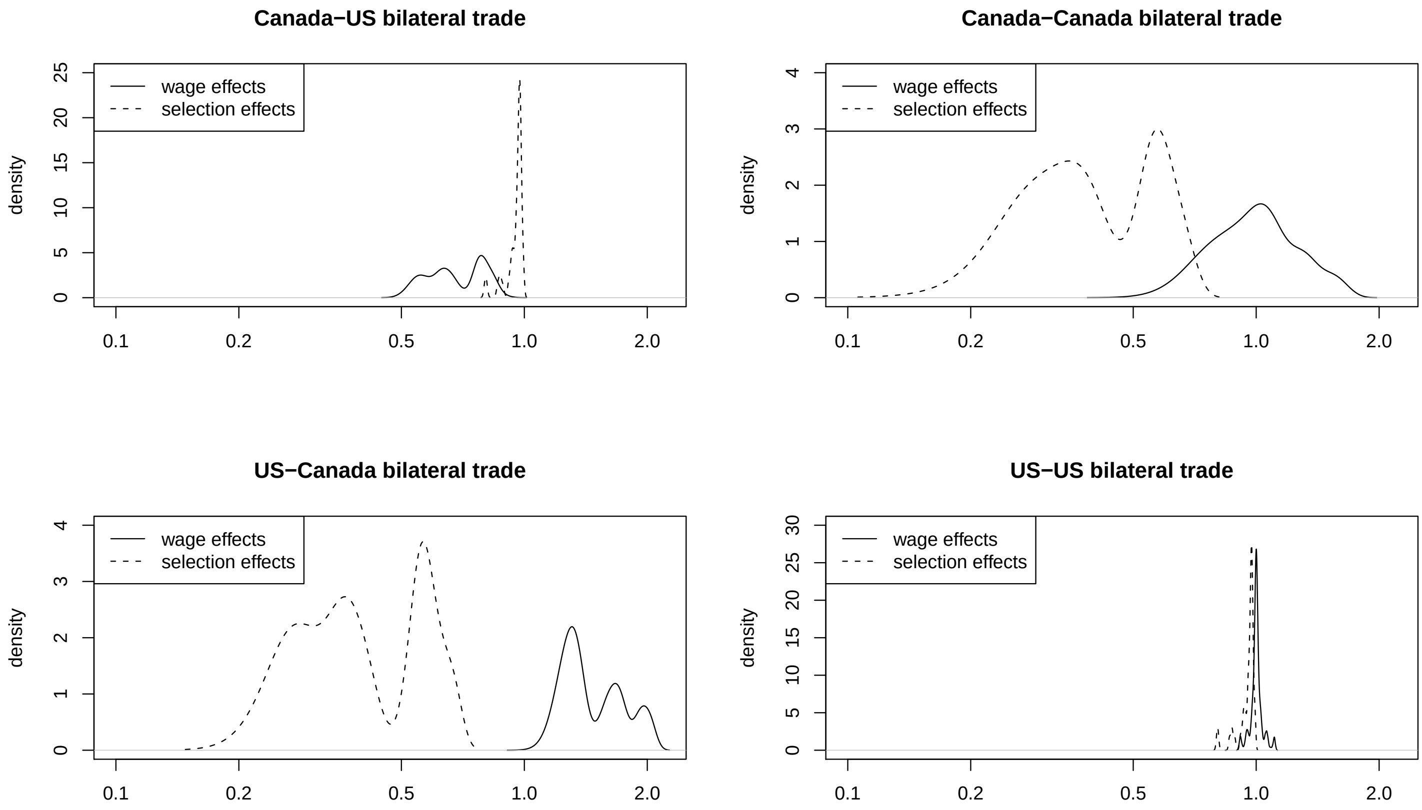

Figure 2. Distributions of wage and selection effects