CESifo Working Papers
www.cesifo.org/wp

Trade, Wages, and Productivity

Kristian Behrens
Giordano Mion
Yasusada Murata
Jens Südekum

CESIFO WORKING PAPER NO. 4011
CATEGORY 8: TRADE POLICY
NOVEMBER 2012

CESifo *Working Paper No. 4011*

Trade, Wages, and Productivity

Abstract

We develop a new general equilibrium monopolistic competition model with variable demand elasticity, heterogeneous firms, and multiple asymmetric regions. Wages, productivity, consumption diversity, and markups across firms and markets are all endogenously determined and respond to trade integration in a way that is consistent with empirical evidence. Using Canada-US regional data, we structurally estimate the model and simulate the impacts of removing all trade barriers generated by the Canada-US border. We find that Canadian average labor productivity increases by 8.03%, whereas US average labor productivity rises by just 1.02%. Consumers' exposure to market power falls sizably by up to 12.11% in the Canadian provinces, and by up to 2.82% in the US states. At the firm level, however, markup changes are ambiguous and depend on the firm's productivity and location. Our results suggest that markups on the firms' side provide a very different piece of information than markups on the consumers' side, which are central to any welfare statement.

JEL-Code: F120, F150, F170.

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

Kristian Behrens
Canada Research Chair
Department of Economics
University of Québec at Montréal
(UQAM) / Canada
behrens.kristian@uqam.ca

Giordano Mion
Department of Geography and Environment
London School of Economics
London WC2A 2AE / United Kingdom
g.mion@lse.ac.uk

Yasusada Murata
Advanced Research Institute for the
Sciences and Humanities (ARISH)
Nihon University
Tokyo / Japan
murata.yasusada@nihon-u.ac.jp

Jens Südekum
Mercator School of Management
University Duisburg-Essen
Duisburg / Germany
jens.suedekum@uni-due.de

November 18, 2012

1 Introduction

Many general equilibrium models of international trade yield equivalent results about the aggregate impacts of trade liberalization for welfare and trade flows as captured by the gravity equation (Arkolakis *et al.*, 2012). However, models differ in their specific predictions along which margins an economy adjusts to freer trade. Recent workhorse frameworks have focused on combinations of wages, productivity, and consumption diversity as adjustment mechanisms, triggered by firm selection and market share reallocations.[1] Yet, those models do not come to grips with the fact that trade integration also changes firms' price-cost margins.

There has been vastly growing empirical interest in markups recently, and important contributions by Feenstra and Weinstein (2010), De Loecker (2011), Simonovska (2011), De Loecker *et al.* (2012), and others, have established some basic facts: (i) markups differ substantially across firms even within industries, and firms with lower marginal costs tend to charge higher markups; (ii) firms apply different markups across different markets; and (iii) trade integration affects price-cost margins.[2] Models based on constant elasticity of substitution (CES) preferences, such as Krugman (1980) and Melitz (2003), cannot accommodate these facts, as markups are constant both within industries and across markets. The model by Bernard *et al.* (2003) also cannot deal with the evidence on the pro-competitive effects of trade, as markups are independent of the number of competing firms and identically distributed across countries.

The contribution of this paper is twofold. First, we develop a new general equilibrium model of trade under monopolistic competition with variable demand elasticity, heterogeneous firms, and multiple asymmetric regions. Wages, productivity, and consumption diversity are all endogenously determined, and in line with the facts (i)–(iii), markups differ across firms and across markets, and respond to trade integration. Second, using Canada-US regional data, we structurally estimate the model's parameters and then simulate the impacts of removing all trade barriers generated by the Canada-US border. This counterfactual analysis allows us to quantitatively explore the features of our model, both at the firm and at the regional level, and to relate it to the large literature on border effects (McCallum, 1995; Anderson and van Wincoop, 2003) and the growing literature on firm-level markup responses to trade integration.

Our framework accommodates all key features of the recent workhorse models, and allows for a richer portrait of the various effects of trade liberalization. Despite its richness, our *gravity equation system* – a gravity equation for bilateral trade and general equilibrium conditions involv-

[1]See the influential models by Eaton and Kortum (2002), Melitz (2003) and Bernard *et al.* (2003), as well as the extensions by Bernard *et al.* (2007b), Arkolakis *et al.* (2008), and Chaney (2008), among others. Empirical studies show that these are indeed relevant channels in trade liberalization episodes, e.g. Aw *et al.* (2000), Pavcnik (2002), Trefler (2004), and Bernard *et al.* (2007a).

[2]Previous studies by Tybout (2003), Syverson (2004, 2007) and Foster *et al.* (2008) are also consistent with this evidence, in particular with fact (i), even though these studies use more restrictive approaches to measure markups.

ing multilateral resistance terms – turns out to be the same as that in Anderson and van Wincoop (2003), except that their elasticity of substitution between varieties is replaced by the shape parameter of our productivity distribution. Interestingly, however, this does not mean that welfare gains from trade are the same. Rather, our prediction for welfare gains is shown to differ in a systematic way from that in the class of models studied by Arkolakis *et al.* (2012).[3]

Since the seminal work by Krugman (1979), there have been several attempts to incorporate endogenous markups into models of international trade. In particular, Melitz and Ottaviano (2008) propose a monopolistic competition model in which markups depend on trade costs and vary across firms and across markets. However, due to their quasi-linear specification, there are no income effects of demand for varieties, and their model displays factor price equalization (FPE) and precludes differential wage responses. Other notable contributions include the translog model by Feenstra (2003, 2010a), the multi-sector model with Cournot competition by Atkeson and Burstein (2008), the model with non-homothetic preferences by Simonovska (2011), and the model by Holmes *et al.* (2012). Our model is complementary to theirs and allows for multiple regions that are asymmetric along various dimensions. Due to this flexible and tractable general equilibrium setup, our approach lends itself quite naturally to structural estimation and counterfactual analysis, where multiple margins of adjustment to trade liberalization can be taken into account.

What are the main quantitative insights of our counterfactual experiment? We find that the hypothetical border removal would increase Canadian labor income shares by 2.18% to 6.56% and enhance Canadian average labor productivity by 8.03%, whereas US average labor productivity would rise by just 1.02%.[4] Gains in consumption diversity and welfare are also larger in Canada, where they range from 5.09% to 13.77%. In the US they only vary up to 2.90%. Investigating what drives these regional variations, we find that geography and size matter: more populous regions and regions closer to the border are on average affected more by the border removal.

Furthermore, in our framework, the markup associated with each transaction depends on the origin region where a firm locates and on the destination region where a consumer locates. Such micro-level origin-destination markups can be aggregated in two different ways.

First, for each consumer in a particular destination, we aggregate markups across all origins,

[3]Arkolakis *et al.* (2012) show that in a class of models, including Krugman (1980), Eaton and Kortum (2002), and Melitz (2003), welfare gains from trade can be expressed by a simple statistic that depends on the import penetration rate and the trade elasticity. In our case a similar – though not identical – formula holds. More specifically, welfare gains are *smaller*, although our model features pro-competitive effects as an additional channel for welfare gains.

[4]Our predicted Canadian productivity gain is not far from 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 channels that are at work at the disaggregate level in our model.

using expenditure shares as weights. This yields a measure of consumers' exposure to market power in that destination. We show that changes in the expenditure share weighted average of markups are a sufficient statistic to assess welfare changes. In our trade liberalization scenario, consumers' exposure to market power falls sizably by up to 12.11% in the Canadian provinces, and by up to 2.82% in the US states. Trade integration – by reducing markups that consumers face – thus translates into substantial aggregate welfare gains.

Second, for each firm in a particular origin, we aggregate markups across all destinations, using sales shares as weights. This yields firm-level markups that measure firms' market power. Contrary to pro-competitive effects on the consumers' side, we show that changes in firm-level markups are ambiguous. Interestingly, the correlation between changes in both types of markups is virtually zero. Our results thus suggest that markups on the firms' side, which are the focus of a number of recent microeconometric studies (e.g., De Loecker *et al.*, 2012), provide a very different piece of information than markups on the consumers' side, which are central to any welfare statement.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 derives the gravity equation system, describes the data, and presents the estimation procedure and results. Section 4 carries out the counterfactual experiment. Section 5 concludes.

2 Model

We consider an economy that consists of K regions. There are L_r consumers in region $r = 1, ..., K$, each of whom inelastically supplies one unit of labor, which is the only factor of production. Workers are immobile across regions. We first set up the model in Sections 2.1 and 2.2, and then analyze the equilibrium in Section 2.3. We finally illustrate some comparative statics in Section 2.4. Several proofs and derivations can be found in Appendices A-E.

2.1 Preferences and demands

There is a final consumption good, provided as a continuum of horizontally differentiated varieties. Consumers have identical preferences that display 'love of variety' and give rise to demands with variable elasticity. 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. Following Behrens and Murata (2007, 2012a), the utility maximization problem of a representative consumer in region r is given by:

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

where Ω_{sr} denotes the endogenously determined set of varieties produced in s and consumed in r. Solving (1) yields the following demand functions:

$$q_{sr}(i) = \frac{E_r}{N_r^c \overline{p}_r} - \frac{1}{\alpha}\left\{\ln\left[\frac{p_{sr}(i)}{N_r^c \overline{p}_r}\right] + h_r\right\}, \quad \forall i \in \Omega_{sr}, \tag{2}$$

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, respectively.[5] Since marginal utility at zero consumption is bounded, the demand for a variety need not be positive. Indeed, as can be seen from (2), 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 $p_r^d \equiv N_r^c \overline{p}_r \mathrm{e}^{\alpha E_r/(N_r^c \overline{p}_r) - h_r}$ depends on the price aggregates $\overline{p}_r$ and h_r. The definition of the reservation price allows us to express the demands for domestic and foreign varieties concisely as follows:

$$q_{rr}(i) = \frac{1}{\alpha}\ln\left[\frac{p_r^d}{p_{rr}(i)}\right] \quad \text{and} \quad q_{sr}(j) = \frac{1}{\alpha}\ln\left[\frac{p_r^d}{p_{sr}(j)}\right]. \tag{3}$$

Observe that the price elasticity of demand for domestic variety i (resp., foreign variety j) is given by $1/[\alpha q_{rr}(i)]$ (resp., $1/[\alpha q_{sr}(j)]$). Thus, if individuals consume more of those varieties, which is for instance the case when their expenditure increases, they become less price sensitive. Last, since $\mathrm{e}^{-\alpha q_{sr}(j)} = p_{sr}(j)/p_r^d$, the indirect utility in region r is given by

$$U_r = N_r^c - \sum_s \int_{\Omega_{sr}} \frac{p_{sr}(j)}{p_r^d}\mathrm{d}j = N_r^c\left(1 - \frac{\overline{p}_r}{p_r^d}\right), \tag{4}$$

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

2.2 Technology and market structure

Prior to production, firms decide in which region they enter and engage in research and development. The labor market in each region is perfectly competitive, so that all firms take the wage rate as given. Entry in region r requires a fixed amount F_r of labor paid at the market wage w_r.

[5] As shown in Reza (1994, pp.278-279), the differential entropy takes its maximum value when there is no dispersion, i.e., $p_{sr}(i) = \overline{p}_r$ for all $i \in \Omega_{sr}$ for all s. In that case, we would observe $h_r = -\ln(1/N_r^c)$ and thus $q_{sr}(i) = E_r/(N_r^c \overline{p}_r)$ by (2). Behrens and Murata (2007, 2012a) focus on such a symmetric case. In contrast, this paper considers firm heterogeneity, so that not only the average price $\overline{p}_r$ but the entire price distribution matter for the demand $q_{sr}(i)$. The differential entropy h_r in (2) captures the latter price dispersion.

Each firm i that enters in region r discovers its marginal labor requirement $m_r(i) \geq 0$ only after making this irreversible entry decision. We assume that $m_r(i)$ is drawn from a known, continuously differentiable distribution G_r.[6] Shipments from region r to region s are subject to trade costs $\tau_{rs} > 1$ for all r and s, which firms incur in terms of labor. Since entry costs are sunk, firms will survive (i.e., operate) provided they can charge prices $p_{sr}(i)$ above marginal costs $\tau_{rs} m_r(i) w_r$ in at least one region. The surviving firms operate in the same region where they enter.

We assume that product markets are segmented, i.e., resale or third-party arbitrage is sufficiently costly, so that firms are free to price discriminate between regions. The operating profit of a firm i located in region r is then as follows:

$$\pi_r(i) = \sum_s \pi_{rs}(i) = \sum_s L_s q_{rs}(i) \left[p_{rs}(i) - \tau_{rs} m_r(i) w_r \right], \tag{5}$$

where $q_{rs}(i)$ is given by (3). Each surviving firm maximizes (5) with respect to its prices $p_{rs}(i)$ separately. Since there is a continuum of firms, no individual firm has any impact on p_r^d, so that the first-order conditions for (operating) profit maximization are given by:

$$\ln \left[\frac{p_s^d}{p_{rs}(i)} \right] = \frac{p_{rs}(i) - \tau_{rs} m_r(i) w_r}{p_{rs}(i)}, \quad \forall i \in \Omega_{rs}. \tag{6}$$

A price distribution satisfying (6) is called a *price equilibrium*. Equations (3) and (6) imply that $q_{rs}(i) = (1/\alpha)[1 - \tau_{rs} m_r(i) w_r / p_{rs}(i)]$. Thus, the minimum output that a firm in r may sell in market s is given by $q_{rs}(i) = 0$ at $p_{rs}(i) = \tau_{rs} m_r(i) w_r$. This, by (6), implies 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 to s, whereas all firms in r with draws below m_{rs}^x are productive enough to sell to 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:

$$m_{rs}^x = \frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d. \tag{7}$$

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

$$N_r^c = \sum_s N_s^E G_s(m_{sr}^x), \tag{8}$$

which is the sum of all firms that are productive enough to sell to market r.[7]

[6] Differences in F_r and G_r thus reflect production amenities such as startup costs and local knowledge that are not transferable across space. Firms take those differences into account when making their entry decisions.

[7] Expression (7) reveals an interesting relationship how trade costs and wage differences affect firms' abilities to break into different markets. In particular, when wages are equalized across regions ($w_r = w_s$) and internal trade

Since all firms in each region differ only by their marginal labor requirements, we can express all firm-level variables in terms of m. Specifically, solving (6) 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, are given by:

$$p_{rs}(m) = \frac{\tau_{rs} m w_r}{W}, \quad q_{rs}(m) = \frac{1}{\alpha}(1 - W), \quad \pi_{rs} = \frac{L_s \tau_{rs} m w_r}{\alpha}(W^{-1} + W - 2), \tag{9}$$

where W denotes the Lambert W function with argument $\mathrm{e}m/m^x_{rs}$, which we suppress to alleviate notation (see Appendix A for more details). Since $W(0) = 0$, $W(\mathrm{e}) = 1$ and $W' > 0$ for all non-negative arguments, we have $0 \leq W \leq 1$ if $0 \leq m \leq m^x_{rs}$. The expressions in (9) show that a firm in r with a draw m^x_{rs} charges a price equal to marginal cost, faces zero demand, and earns zero operating profits in market s. Furthermore, using the properties of W', we readily obtain $\partial p_{rs}(m)/\partial m > 0$, $\partial q_{rs}(m)/\partial m < 0$, and $\partial \pi_{rs}(m)/\partial m < 0$. In words, firms with higher productivity (lower m) charge lower prices, sell larger quantities, and earn higher operating profits. These properties are similar to those of the Melitz (2003) model with CES preferences. Yet, our specification with variable demand elasticity also features higher markups for more productive firms. Indeed, the *origin-destination markup* for a firm located in the origin region r and a consumer located in the destination region s,

$$\Lambda_{rs}(m) \equiv \frac{p_{rs}(m)}{\tau_{rs} m w_r} = \frac{1}{W} \tag{10}$$

implies that $\partial \Lambda_{rs}(m)/\partial m < 0$. Unlike Melitz and Ottaviano (2008), who use quasi-linear preferences, we incorporate this feature into a full-fledged general equilibrium model with income effects for varieties.

2.3 Equilibrium

2.3.1 Single region case

To illustrate our model, we first consider the single region case. There are two equilibrium conditions: zero expected profits and labor market clearing. These two conditions can be solved for the domestic cutoff m^d and the mass of entrants N^E, which completely characterize the equilibrium. For notational convenience, we drop the regional subscript and normalize the internal trade costs to one. Using (5), the zero expected profit condition is given by:

$$L \int_0^{m^d} [p(m) - mw]\, q(m) \mathrm{d}G(m) = Fw, \tag{11}$$

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 local firms in s, which is the standard case in the literature (e.g., Melitz, 2003; Melitz and Ottaviano, 2008). However, in the presence of wage differences and internal trade costs, the domestic cutoff need not be larger than the export cutoff in equilibrium. The usual ranking $m^d_s > m^x_{rs}$ prevails only when $\tau_{ss} w_s < \tau_{rs} w_r$.

which, combined with (9), can be rewritten as a function of m^d only:

$$\frac{L}{\alpha}\int_0^{m^d} m\left(W^{-1}+W-2\right)\mathrm{d}G(m)=F. \tag{12}$$

As the left-hand side of (12) is strictly increasing in m^d from 0 to ∞, there always exists a unique equilibrium cutoff. Furthermore, the labor market clearing condition is given by:[8]

$$N^E\left[L\int_0^{m^d} mq(m)\mathrm{d}G(m)+F\right]=L, \tag{13}$$

which, combined with (9), can be rewritten as a function of m^d and N^E:

$$N^E\left[\frac{L}{\alpha}\int_0^{m^d} m\left(1-W\right)\mathrm{d}G(m)+F\right]=L. \tag{14}$$

Given the equilibrium cutoff m^d from (12), equation (14) can be uniquely solved for N^E.

How does population size affect entry and firms' survival probabilities? Using the equilibrium conditions (12) and (14), we can show that a larger L leads to more entrants N^E and a smaller cutoff m^d, respectively. 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^p, which is equivalent to consumption diversity N^c in the single region case, cannot be signed for a general distribution G. However, under the commonly made assumption that firms' productivity draws $1/m$ follow a Pareto distribution[9]

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

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

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

where κ_1 and κ_2 are positive constants that solely depend on k, and $\mu^{\max}\equiv\left[\alpha F(m^{\max})^k\right]/\kappa_2$.[10] The term $\mu^{\max}$ can be interpreted as a measure of 'technological possibilities': the lower the fixed labor requirement for entry F or the lower the upper bound for the firms' draws $m^{\max}$, the lower

[8]Note that by using (11) 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 (13), yields $E=w$ in equilibrium. The expenditure of the representative consumer thus depends only on the wage rate.

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

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

the value of $\mu^{\max}$ and, hence, the better the technological possibilities. As can be seen from (15), the cutoff m^d is decreasing in L and increasing in $\mu^{\max}$. Since $\overline{m} = [k/(k+1)]m^d$ holds when productivity follows a Pareto distribution, a larger population or better technological possibilities map into higher average productivity $1/\overline{m}$. The mass of surviving firms is given by

$$N = \frac{1}{\kappa_1 + \kappa_2} \frac{\alpha}{m^d} = \frac{\alpha}{\kappa_1 + \kappa_2} \left(\frac{L}{\mu^{\max}} \right)^{\frac{1}{k+1}}, \tag{16}$$

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

We next turn to the issue of markups. It follows from (10) that origin-destination markups in the single region case are defined as $1/W(\mathrm{e}m/m^d)$. It is worth pointing out that firm productivity $1/m$ and average productivity $1/\overline{m} = 1/\{[k/(k+1)]m^d\}$ have opposite effects: more productive firms (smaller m) tend to charge higher markups, whereas tougher selection in the market (lower m^d) puts downward pressure on markups. To derive a measure for the consumers' exposure to market power, we need to aggregate those markups across all origins. The simple (unweighted) average is not adequate for this purpose, however, as consumers have different expenditure shares across varieties produced with different productivity. We hence define the expenditure share-weighted average of markups as follows:

$$\overline{\Lambda}^c \equiv \frac{1}{G(m^d)} \int_0^{m^d} \frac{p(m)q(m)}{E} \Lambda(m) \mathrm{d}G(m) = \frac{\kappa_3 m^d}{\alpha}, \tag{17}$$

where κ_3 is a positive constant that solely depends on k.[11] Note that the weighted average of markups is proportional to the cutoff. It thus follows from (16) and (17) that our model displays pro-competitive effects, since $\overline{\Lambda}^c$ decreases with the mass N of competing firms:

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

Note further that expression (17), together with (15), shows that $\overline{\Lambda}^c$ is smaller in larger markets or markets with better technological possibilities, as more firms compete in these markets.[12]

Finally, the indirect utility in the single region case is given by

$$U = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\alpha}{m^d} = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\kappa_3}{\overline{\Lambda}^c}, \tag{18}$$

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

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

[12] A similar result can be obtained when using the weighted average of Lerner indices $[p(m) - mw]/p(m)$ as an alternative measure of market power. In that case, $\overline{\Lambda}^c$ is given by $(\kappa_2 m^d)/\alpha$, which is also decreasing in L and increasing in $\mu^{\max}$. An *unweighted* average of markups (or Lerner indices) would be a constant in our model, as in other models with heterogeneous firms (e.g., Bernard *et al.*, 2003; Melitz and Ottaviano, 2008).

be seen from (15)–(18), larger markets or markets with better technological possibilities allow for higher utility because of tougher selection, tougher competition, and greater consumption diversity.

2.3.2 Multi-region case

In the multi-region economy, there are three sets of equilibrium conditions. For each region, zero expected profits and labor market clearing can be written analogously as in the single region setup. In addition, trade must be balanced for each region, which requires that the total value of exports equals the total value of imports. The zero expected profit condition in region r now reads as

$$\sum_s L_s \int_0^{m_{rs}^x} \left[p_{rs}(m) - \tau_{rs} m w_r\right] q_{rs}(m) \mathrm{d}G_r(m) = F_r w_r, \tag{19}$$

and the labor market clearing condition becomes

$$N_r^E \left[\sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m q_{rs}(m) \mathrm{d}G_r(m) + F_r\right] = L_r. \tag{20}$$

Last, the trade balance condition for region r is given by

$$N_r^E \sum_{s \neq r} L_s \int_0^{m_{rs}^x} p_{rs}(m) q_{rs}(m) \mathrm{d}G_r(m) = L_r \sum_{s \neq r} N_s^E \int_0^{m_{sr}^x} p_{sr}(m) q_{sr}(m) \mathrm{d}G_s(m). \tag{21}$$

As in the single region case, we assume Pareto distributions for productivity draws. The shape parameter $k \geq 1$ is assumed to be identical, but the upper bounds are allowed to vary across regions, i.e., $G_r(m) = (m/m_r^{\max})^k$. Under this parametrization, and making use of (7) and (9), the equilibrium conditions can be simplified. In particular, the zero expected profit, labor market clearing and trade balance conditions can be rewritten as follows:

$$\mu_r^{\max} = \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d\right)^{k+1}, \tag{22}$$

$$N_r^E \left[\frac{\kappa_1}{\alpha \left(m_r^{\max}\right)^k} \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d\right)^{k+1} + F_r\right] = L_r, \tag{23}$$

$$\frac{N_r^E w_r}{\left(m_r^{\max}\right)^k} \sum_{s \neq r} L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d\right)^{k+1} = L_r \sum_{s \neq r} \tau_{sr} \frac{N_s^E w_s}{\left(m_s^{\max}\right)^k} \left(\frac{\tau_{rr}}{\tau_{sr}} \frac{w_r}{w_s} m_r^d\right)^{k+1}, \tag{24}$$

where $\mu_r^{\max} \equiv \left[\alpha F_r \left(m_r^{\max}\right)^k\right]/\kappa_2$ denotes technological possibilities. Note that $\mu_r^{\max}$ is region-specific, and thus captures the local production amenities that are not transferable across space.

The $3 \times K$ general equilibrium conditions (22)–(24) depend on $3 \times K$ 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 (7). Combining (22) and (23), we can immediately show that

$$N_r^E = \frac{\kappa_2}{\kappa_1 + \kappa_2} \frac{L_r}{F_r}, \tag{25}$$

which implies that more firms choose to enter in larger markets and in markets with lower entry requirements. Adding the term in r that is missing on both sides of (24), and using (22) and (25), we obtain the following relationship:

$$\frac{1}{(m_r^d)^{k+1}} = \sum_s L_s \tau_{rr} \left(\frac{\tau_{rr}}{\tau_{sr}} \frac{w_r}{w_s} \right)^k \frac{1}{\mu_s^{\max}}. \tag{26}$$

The $2 \times K$ conditions (22) and (26) summarize how wages and cutoffs are related in general equilibrium, given the regional population sizes, technological possibilities, and trade costs.

Using these expressions, we can furthermore show that – in equilibrium – the mass of varieties consumed in region r is inversely proportional to the domestic cutoff, while the (expenditure share) weighted average of markups that consumers face is proportional to the cutoff in that region,

$$N_r^c = \frac{1}{\kappa_1 + \kappa_2} \frac{\alpha}{\tau_{rr} m_r^d}, \tag{27}$$

$$\overline{\Lambda}_r^c \equiv \frac{\sum_s N_s^E \int_0^{m_{sr}^x} \frac{p_{sr}(m) q_{sr}(m)}{E_r} \Lambda_{sr}(m) \mathrm{d}G_s(m)}{\sum_s N_s^E G_s(m_{sr}^x)} = \frac{\kappa_3 \tau_{rr} m_r^d}{\alpha}. \tag{28}$$

It thus follows from (27) and (28) that there are pro-competitive effects on the consumers' side, as in the single region case, since $\overline{\Lambda}_r^c$ decreases with the mass of competing firms in region r, $\overline{\Lambda}_r^c = [\kappa_3/(\kappa_1 + \kappa_2)](1/N_r^c)$. Finally, the indirect utility can be expressed as

$$U_r = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\alpha}{\tau_{rr} m_r^d} = \left[\frac{1}{(\kappa_1 + \kappa_2)(k+1)} - 1 \right] \frac{\kappa_3}{\overline{\Lambda}_r^c}, \tag{29}$$

which implies that tougher selection (lower m_r^d) and tougher competition (lower $\overline{\Lambda}_r^c$) translate into higher welfare in region r. Alternatively, we have $U_r = [1/(k+1) - (\kappa_1 + \kappa_2)] N_r^c$, i.e., the indirect utility is proportional to the mass of varieties consumed.[13]

2.4 Some comparative statics

In order to build intuition for the counterfactual experiment, we illustrate some comparative statics results for the two-region case. Using (22)–(25), 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 $\omega \equiv w_1/w_2$):

[13] 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 the cutoff m_r^d. This is in accordance 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.

$$\mu_1^{\max} = L_1\tau_{11}\left(m_1^d\right)^{k+1} + L_2\tau_{12}\left(\frac{\tau_{22}}{\tau_{12}}\,\frac{1}{\omega}\,m_2^d\right)^{k+1} \tag{30}$$

$$\mu_2^{\max} = L_2\tau_{22}\left(m_2^d\right)^{k+1} + L_1\tau_{21}\left(\frac{\tau_{11}}{\tau_{21}}\,\omega\,m_1^d\right)^{k+1} \tag{31}$$

$$\omega^{2k+1} = \rho\left(\frac{\tau_{21}}{\tau_{12}}\right)^k\left(\frac{\tau_{22}}{\tau_{11}}\right)^{k+1}\left(\frac{m_2^d}{m_1^d}\right)^{k+1}, \tag{32}$$

where $\rho \equiv \mu_2^{\max}/\mu_1^{\max}$. When $\rho > 1$, region 1 has better technological possibilities than region 2. Equations (30) and (31) can readily be solved for the cutoffs as a function of the relative wage:

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

Substituting these cutoffs into (32) yields after some simplification

$$\text{LHS} \equiv \omega^k = \rho\frac{L_1}{L_2}\left(\frac{\tau_{21}}{\tau_{12}}\right)^k\frac{\rho\tau_{11}^{-k} - \tau_{21}^{-k}\omega^{k+1}}{\tau_{22}^{-k}\omega^{k+1} - \rho\tau_{12}^{-k}} \equiv \text{RHS}. \tag{33}$$

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 (33) 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. In what follows, assume further that trade costs are symmetric ($\tau_{12} = \tau_{21} = \tau$ and $\tau_{11} = \tau_{22} = t$, where $\tau > t$). We can then establish the following three results.[14]

(a) Differences in market size or technological possibilities. Suppose that the two regions are identical, except that region 1 is larger or has better technological possibilities than region 2. Noting that the RHS of (33) is decreasing in ω, we can prove the following results. For given levels of trade costs, region 1 has the higher wage ($\omega^* > 1$), the lower cutoff ($m_1^d < m_2^d$), the higher average productivity ($1/\overline{m}_1 > 1/\overline{m}_2$), the greater consumption diversity ($N_1^c > N_2^c$), the lower (expenditure share) weighted average of markups ($\overline{\Lambda}_1^c < \overline{\Lambda}_2^c$), and the higher welfare ($U_1 > U_2$) than region 2. 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

[14] If we relax the assumption of symmetric trade costs, we can also prove that 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 (expenditure share) weighted average of markups, greater consumption diversity and higher welfare than region 2.

in region 1 requires, in equilibrium, a higher wage and a lower cutoff in that region. The latter then maps into more consumption diversity, tougher competition and higher welfare as shown in (27)–(29).

(b) The impacts of trade integration – symmetric regions. To study the impacts of trade integration, we first consider the simplest case of two symmetric regions with $L_1 = L_2 = L_{sym}$ and $\mu_1^{\max} = \mu_2^{\max} = \mu_{sym}^{\max}$, so that $\rho = 1$. In that case, it is easy to see from (30)–(32), as well as (33), that $\omega^* = 1$ and that $m_1^d = m_2^d = m_{sym}^d$, with

$$m_{sym}^d = \left[\frac{1}{1 + (t/\tau)^k} \frac{\mu_{sym}^{\max}}{t L_{sym}} \right]^{\frac{1}{k+1}}. \tag{34}$$

Since m_{sym}^d is increasing in τ, trade integration induces tougher selection. As the mass of entrants is constant by (25), there is, hence, exit of the least productive firms in both regions. We also know that the export cutoff in (7) can be written as $m_{sym}^x = (t/\tau) m_{sym}^d$, which is shown to be decreasing in τ. The export cutoff thus rises as bilateral trade barriers fall, so that the share of exporters goes up. Consumption diversity N_{sym}^c increases as τ falls by (27), that is, the displacement of domestic varieties is more than compensated by newly imported varieties in both regions.

Turning to the issue of markups, all firms that remain active after trade liberalization charge a lower markup on the domestic market since $W(\mathrm{e}m/m_{sym}^d)$ in (10) rises. However, exporting firms raise their markup on the foreign market since $W(\mathrm{e}m/m_{sym}^x)$ decreases due to trade liberalization. Note that the markup of each exporting firm is a function of its relative position in the firm productivity distribution m/m_{sym}^x, so that the firm with productivity $1/m$ can raise its markup after the border removal due to the larger m_{sym}^x. Intuitively, the reduction of τ induces less productive firms to export. This, in turn, allows the existing more productive exporters to charge a higher markup. From the perspective of a single firm, trade liberalization may therefore lead to a lower or to a higher average markup, depending on the importance of the export market for the firm's total sales. From the consumers' perspective, however, it is clear from (28) that there are pro-competitive effects of trade liberalization, since the (expenditure share) weighed average of markups $\overline{\Lambda}_{sym}^c$ decreases. Finally, (29) implies that trade integration leads to welfare gains for both regions.

(c) The impacts of trade integration – asymmetric regions. Now suppose that region 1 is larger or has better technological possibilities than region 2, as in **(a)**. Then, as interregional trade costs τ decrease, it can be shown that wages and cutoffs converge between the two regions. That is, both the relative wage $\omega^* > 1$ and the relative cutoff $m_2^d/m_1^d > 1$ become smaller, which then directly implies that (expenditure share) weighted averages of markups, consumption diversity and welfare also converge. These results thus suggest that bilateral trade liberalization tends to attenuate regional economic differences.

To our knowledge, there exists so far no trade model with heterogeneous firms that can provide such a rich portrait of the impacts of trade liberalization in a unifying framework. 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 qualitatively similar conclusions about the impacts on wages, cutoffs and welfare.[15] However, markups are fixed in CES models because the price elasticity of demand is constant by construction. This result thus sharply contradicts the abundant empirical evidence discussed above, which shows that markups depend on productivity and local market size, and in particular, that markups respond to trade liberalization (e.g., Feenstra and Weinstein, 2010; De Loecker, 2011; and De Loecker *et al.*, 2012). Second, models based on the Melitz and Ottaviano (2008) framework assume quasi-linear preferences without income effects of demand for varieties. Their trade equilibrium displays FPE, and therefore precludes differential wage responses to trade liberalization across asymmetric regions. In Bernard *et al.* (2003) there are no endogenous reponses of wages and markups triggered by trade liberalization.

3 Estimation

We now take the model with K asymmetric regions to the data. To this end, we first derive a *gravity equation system* – a gravity equation with general equilibrium conditions. Using Canada-US regional data, we then structurally estimate trade friction parameters as well as other variables of the model. Finally, we assess the performance of our estimated framework by comparing its predictions to several empirical facts that were not used in the quantification procedure.

3.1 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 (7), (9), (25), and the Pareto distribution for $G_r(m)$, we obtain the gravity equation:

$$\frac{X_{rs}}{L_r L_s} = \tau_{rs}^{-k} \tau_{ss}^{k+1} (w_s/w_r)^{k+1} w_r (m_s^d)^{k+1} (\mu_r^{\max})^{-1}. \tag{35}$$

As can be seen from (35), 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

[15] 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 equilibrium. Hence, both models cannot cope with endogenous wage responses driven by trade integration.

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 (22) and (26) give us the following general equilibrium conditions:

$$\mu_r^{\max} = \sum_v L_v \tau_{rv}^{-k} \tau_{vv}^{k+1} \left(\frac{w_v}{w_r} \right)^{k+1} (m_v^d)^{k+1} \quad r = 1, 2. \dots K \tag{36}$$

$$\frac{1}{(m_s^d)^{k+1}} = \sum_v L_v \tau_{vs}^{-k} \tau_{ss}^{k+1} \left(\frac{w_s}{w_v} \right)^{k} (\mu_v^{\max})^{-1} \qquad s = 1, 2, \dots K. \tag{37}$$

The *gravity equation system* consists of the gravity equation (35) and the $2 \times K$ general equilibrium conditions (36) and (37) that summarize the interactions between the endogenous variables, namely the K wages and the K cutoffs. The latter conditions are reminiscent of those 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.[16]

Interestingly, the gravity equation system (35)–(37) is, indeed, akin to that in Anderson and van Wincoop (2003). To see this, let $Y_r = w_r L_r$ be the labor income of country r. Define the world labor income as $Y_W = \sum_r w_r L_r$ and the labor income share of country r as $\sigma_r = Y_r / Y_W$. Also define the multilateral resistance terms as follows:

$$\Phi_r^{-k} = \left(\frac{w_r L_r}{Y_W} \right)^{k+1} \mu_r^{\max} L_r^{-k-1} = \sigma_r^{k+1} \mu_r^{\max} L_r^{-k-1} \tag{38}$$

$$\Psi_s^{-k} = \left(\frac{w_s L_s}{Y_W} \right)^{-k} \tau_{ss}^{-k-1} (m_s^d)^{-k-1} L_s^k = \sigma_s^{-k} \tau_{ss}^{-k-1} (m_s^d)^{-k-1} L_s^k. \tag{39}$$

Then, our gravity equation system (35)–(37) can be rewritten as

$$X_{rs} = \frac{Y_r Y_s}{Y_W} \left(\frac{\tau_{rs}}{\Phi_r \Psi_s} \right)^{-k} \tag{40}$$

$$\Phi_r^{-k} = \sum_v \sigma_v \left(\frac{\tau_{rv}}{\Psi_v} \right)^{-k} \tag{41}$$

$$\Psi_s^{-k} = \sum_v \sigma_v \left(\frac{\tau_{vs}}{\Phi_v} \right)^{-k}, \tag{42}$$

[16]Note that expression (35) is also similar to gravity equations that have been derived in previous models with heterogeneous firms. Those 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). One exception is Balistreri and Hillberry (2007), who allow regional incomes to respond to trade liberalization. However, their model abstracts from firm heterogeneity. See Balistreri *et al.* (2011) for a structural estimation of a CES model with heterogeneous firms and fixed markups.

which is the same as the gravity equation system (9)–(11) in Anderson and van Wincoop (2003), except that their exponent capturing the elasticity of substitution is replaced by the shape parameter k of the Pareto distributions. Assuming that $\tau_{rs} = \tau_{sr}$ as in Anderson and van Wincoop (2003), we know that (41) and (42) yield a solution $\Phi_r = \Psi_r$ with

$$\Phi_r^{-k} = \sum_v \sigma_v \tau_{rv}^{-k} \Phi_v^k, \tag{43}$$

which we will use in the subsequent analysis as it greatly simplifies our estimation procedure.

Notice that this equivalence of the gravity equation systems does *not* mean that our model makes the same predictions about the impacts of trade integration as that of Anderson and van Wincoop (2003).[17] The reason is that the changes in the multilateral resistance terms, as well as those in the income shares, are model specific. More importantly, our model structure is richer than that in Anderson and van Wincoop (2003), and allows for many additional margins of adjustment. Therefore, we can provide a more detailed portrait of the effects of trade integration on different economic variables such as productivity and markups. We return to these points in Section 4 below.

3.2 Data and estimation procedure

Our sample consists of 10 Canadian provinces, 50 US states plus the District of Columbia, and 23 rest-of-the-world (ROW) countries (OECD countries plus Mexico). All data is for 1993. To measure L_r, we use data on employment. We use hourly average wage rates in US dollars for w_r. The GDP is then constructed as $Y_r = w_r L_r$, from which we construct the labor income shares σ_r. For the specification of trade costs τ_{rs} we stick to standard practice and assume that $\tau_{rs} \equiv d_{rs}^{\gamma} \mathrm{e}^{\theta b_{rs}}$, where d_{rs} stands for distance between r and s. Following Anderson and van Wincoop (2003), bilateral distances d_{rs} are computed using the great circle formula. Internal distances are measured either by $d_{rr} = (2/3)\sqrt{\text{surface}_r/\pi}$ as in Redding and Venables (2004) using surface data, or by $d_{rr} = (1/4)\min_{s \neq r}\{d_{rs}\}$ as in Anderson and van Wincoop (2003). As shown below, results are not sensitive to that choice. All distances are measured in thousands of kilometers. The term b_{rs} is a border dummy valued 1 if r and s are not in the same country and 0 otherwise.

Bilateral trade flows X_{rs} for 10 Canadian provinces and 30 US states are publicly available (see Anderson and van Wincoop, 2003; Feenstra, 2004). When estimating the trade friction parameters γ and θ, we deal with the fact that bilateral trade flows among the 40 regions are also affected by

[17] Anderson and van Wincoop (2003) find that, on average, the border increases trade between Canadian provinces by a factor of 10.7 when compared to trade with US states. The corresponding number for the US is 2.24. Using their definition of the border effect, our model and calibration predict less dissimilar effects of removing the border on trade flows: 6.18 for Canada and 2.89 for the US, respectively.

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. We address this issue by including further regions and countries into (41)–(42), even if we lack their bilateral trade flow data.[18]

Our estimation procedure for the gravity equation system (40)–(42) can be described as follows:

1. Given our specification of trade costs $\tau_{rs} \equiv d_{rs}^{\gamma} \mathrm{e}^{\theta b_{rs}}$, the gravity equation (40) can be rewritten in log-linear stochastic form:

$$\ln\left(\frac{X_{rs}}{Y_r Y_s}\right) = c - k\gamma \ln d_{rs} - k\theta b_{rs} + \zeta\chi_{rs} + k\ln\Phi_r + k\ln\Psi_s + \varepsilon_{rs}, \tag{44}$$

   where c is a constant; χ_{rs} is a zero flow dummy such that $\chi_{rs} = 0$ if $X_{rs} > 0$ and $\chi_{rs} = 1$ otherwise; and ε_{rs} is an error term with the usual properties.[19] We estimate (44) using origin and destination fixed effects, which for a given value of k yields estimates of the parameters $(c, \gamma, \theta, \zeta)$.

2. Given the starting values $(c, \gamma, \theta, \zeta)$ obtained from the fixed-effects regression, and using (43), we then solve the problem as in Anderson and van Wincoop (2003) as follows:

$$\min_{c,\gamma,\theta,\zeta} \sum_r \sum_{s\neq r} \left[\ln\left(\frac{X_{rs}}{Y_r Y_s}\right) - c + k\gamma \ln d_{rs} + k\theta b_{rs} - \zeta\chi_{rs} - k\ln\Phi_r - k\ln\Phi_s\right]^2$$
$$\text{subject to} \quad \Phi_r^{-k} = \sum_v \sigma_v d_{rv}^{-\gamma k} \mathrm{e}^{-k\theta b_{rv}} \Phi_v^k, \quad \text{for} \quad r = 1, 2, \ldots, K.$$

   Given k, for any given (γ, θ), these constraints can be solved for $\Phi_r = \Phi_r(\gamma, \theta)$. The above minimization problem thus boils down to

$$\min_{c,\gamma,\theta,\zeta} \sum_r \sum_{s\neq r} \left[\ln\left(\frac{X_{rs}}{Y_r Y_s}\right) - c + k\gamma \ln d_{rs} + k\theta b_{rs} - \zeta\chi_{rs} - k\ln\Phi_r(\gamma,\theta) - k\ln\Phi_s(\gamma,\theta)\right]^2,$$

[18] See Table 3 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, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, 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).

[19] We add one dollar to zero flows in the data and include a zero-flow dummy in the regression as there is no generally agreed-upon methodology to deal with this problem (see, e.g., Anderson and van Wincoop, 2004; Disdier and Head, 2008). When focusing on 40 regions, there are 49 zero trade flows out of 1,600 observations. Yet, our zeros are unlikely to be 'true zeros', as this would imply no aggregate manufacturing trade between several US states.

which yields the parameter estimates $(\widehat{c}, \widehat{\gamma}, \widehat{\theta}, \widehat{\zeta})$ and a set of multilateral resistance terms $\{\widehat{\Phi}_r\}_{r=1}^{40}$ for 'in gravity sample'. Furthermore, given k and $(\widehat{\gamma}, \widehat{\theta})$, we use the constraints to compute $\{\widehat{\Phi}_r\}_{r=41}^{84}$ for 'out of gravity sample'.

3. Plugging $\{\widehat{\Phi}_r\}_{r=1}^{84}$ into (38) and using the data on the income share and population, we compute, for a given value of k, technological possibilities $\{\widehat{\mu}_r^{\max}\}_{r=1}^{84}$ that depend on the unobservables α, F_r and $m_r^{\max}$. Similarly, from (39) we obtain the cutoffs $\{\widehat{m}_r^d\}_{r=1}^{84}$, given k.

4. Finally, using the estimates thus obtained under the value of k, we compute the productivity advantage of US exporters from a random sample of firms drawn from the fitted productivity distributions of our model (see Appendix F for more details). We repeat this procedure for different values of k until our sample matches the 33% productivity advantage of US exporters in 1992, which is reported by Bernard *et al.* (2003).

3.3 Estimation results

Our estimation procedure yields $k = 8.5$, which we henceforth take as our benchmark. Our choice of k is well in line with previous studies. For example, Feenstra (2010b, p.53) states that $k = 2.9$ is "too low", and chooses $k = 8$ by referring to Eaton and Kortum (2002) who find values in the range between 6 and 12 with the most robust estimate around 8.

Insert Table 1 about here.

Our estimation results for the gravity equation system are summarized in Table 1, where we report bootstrapped standard errors in parentheses.[20] Column 1 presents the benchmark case with $K = 84$ areas and $k = 8.5$, whereas columns 2-4 contain alternative specifications used as robustness checks. 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 -1.3049, which implies that $\widehat{\gamma} = 0.1535$. The border coefficient estimate is -1.4484, which implies that $\widehat{\theta} = 0.1704$.

Column 2 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 2 are not statistically different from each other as the corresponding 95% confidence intervals overlap. However, 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. Column 3 reports results using an alternative measure for internal distance as proposed by Anderson and van Wincoop

[20]To this end, we proceed as follows. First, we randomly permute $\widehat{\varepsilon}_{rs}$ obtained after estimation step 1 to get $\widehat{\varepsilon}_{rs}^b$. We then compute $\widehat{X}_{rs}^b \equiv X_{rs} + \widehat{\varepsilon}_{rs}^b$ that are consistent with the permutation to obtain new initial values $(c^b, \gamma^b, \theta^b, \zeta^b)$, and apply step 2. By repeating this procedure, we end up with a distribution for $(\widehat{c}^b, \widehat{\gamma}^b, \widehat{\theta}^b, \widehat{\zeta}^b)$ from which we compute standard errors. We use 200 replications.

(2003), whereas column 4 presents results when we exclude Alaska, Hawaii, and the District of Columbia. Again, the coefficient of the border dummy does not vary much across the different specifications.

3.4 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 domestic cutoffs $\widehat{m}_r^d$, obtained from our estimation, with the observed cutoffs. The latter are constructed as follows. 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 GDP per employee, a standard measure of labor productivity. In our benchmark case, the correlation between computed and observed cutoffs is 0.88 when including the ROW, and 0.63 when focusing only on Canadian provinces and US states. Thus, the predicted cutoffs match observed ones fairly well.

Insert Table 2 about here.

Although our main focus is on regional aggregates, we can also assess the fit of our estimated model by using well-established firm-level facts, namely, the share of exporters and the distribution of export intensities (see Appendix F for more details on how to construct firm samples). 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 5.19% across all US firms, which is reasonably close to the observed number. The corresponding share of Canadian exporters is given by 15.81%. 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 seen from the third column in Table 2, 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 the fourth column. Not surprisingly, Canadian firms have higher export intensity as they are less reliant on the small domestic market.

4 Counterfactual analysis

Having estimated the gravity equation system and having assessed the performance of the estimated model, we now move to the counterfactual experiment. In particular, we consider a scenario where all trade barriers generated by the Canada-US border are completely eliminated. In

Section 4.1 we explore how this hypothetical trade integration would affect various regional economic aggregates such as wages, productivity, consumption diversity, and welfare. In Section 4.2 we focus on the impacts on markups both from the firms' and the consumers' perspective. Finally, we address the impacts on bilateral trade flows and provide a decomposition of the border effects in Section 4.3.

Let variables with a tilde refer to values in the hypothetical "borderless" scenario. Formally, to compute the counterfactual equilibrium, we set $\widetilde{b}_{rs} = 0$ for all r and s, holding the shape parameter k, the estimated technological possibilities $\widehat{\mu}_r^{\max}$, and trade frictions $(\widehat{\gamma}, \widehat{\theta})$ constant. Our procedure can then be summarized as follows.

1. We first eliminate Ψ_v from (41) by substituting (42) to obtain:

$$\Phi_r^{-k} = \sum_v \frac{\sigma_v \widetilde{\tau}_{rv}^{-k}}{\sum_s \sigma_s \left(\frac{\widetilde{\tau}_{sv}}{\Phi_s}\right)^{-k}}, \tag{45}$$

   where $\widetilde{\tau}_{rv} = d_{rs}^{\widehat{\gamma}} \mathrm{e}^{\widehat{\theta}\widetilde{b}_{rs}} = d_{rs}^{\widehat{\gamma}}$. Plugging (38) into both sides of (45) yields a system of equations that depends on the labor income shares σ_r only as follows:

$$\sigma_r^{k+1} \widehat{\mu}_r^{\max} L_r^{-k-1} = \sum_v \frac{\sigma_v \widetilde{\tau}_{rv}^{-k}}{\sum_s \sigma_s^{-k} \widetilde{\tau}_{sv}^{-k} (\widehat{\mu}_s^{\max})^{-1} L_s^{k+1}}.$$

   Since the system is not independent, we drop one of the equations and impose the constraint that the labor income shares sum to one: $\sum_v \sigma_v = 1$. We solve that system for the counterfactual labor income shares $\widetilde{\sigma}_r$ that would prevail in the borderless world.

2. Plugging the counterfactual labor income shares $\widetilde{\sigma}_r$ thus obtained into (38), we compute $\widetilde{\Phi}_r^{-k} = \widetilde{\sigma}_r^{k+1} \widehat{\mu}_r^{\max} L_r^{-k-1}$. Combining the resulting expressions and (42) yields $\widetilde{\Psi}_r^{-k}$. This, together with (39), generates the counterfactual cutoffs $\widetilde{m}_r^d$.

3. Noting (27)–(29), we obtain the (expenditure share) weighted average of markups $\widetilde{\overline{\Lambda}}_r^c$, the mass of varieties consumed $\widetilde{N}_r^c$ and the welfare $\widetilde{U}_r$ in each region. Using these new values, we can also compute the counterfactual firm-level variables for our random sample of firms.

Insert Tables 3–4 about here.

4.1 The impacts on wages, productivity, consumption diversity and welfare

Table 3 reports the predicted percentage changes in regional aggregate variables that would occur after eliminating all trade barriers generated by the border.

(a) Wages. First, we discuss the changes in labor income shares, which are proportional to the regional wage changes. For Canadian provinces the changes in labor income shares range from 2.18% for Newfoundland to 6.56% for British Columbia. In the US those changes can be positive or negative, and are much smaller in absolute values.

To further explore the patterns of these hypothetical wage changes, we use a simple approach and regress these changes on two crucial exogenous regional characteristics: geography and size. The former dimension is captured by the distance from region r to the closest foreign region, and the latter by population size L_r. We further include a US dummy to pick up overall differences between the two countries. Such a multivariate OLS regression analysis allows us to address, in a simple descriptive way, whether the border removal would mainly affect regions closer to the border or smaller regions. Table 4 reports the results.

The first specification in Table 4 confirms the 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.[21] We also find that regions further away from the border tend to experience smaller income share increases, and that there is a positive and significant relationship with population size. In the second specification we consider interaction terms of our proxies for *geography* and *size* with the US dummy in order to capture parameter heterogeneity. This specification shows that the income share increases are stronger in larger Canadian provinces, thus favoring wage divergence across them, whereas for US states there is virtually no relationship between size and wage changes. The elasticity of changes in income shares 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.

(b) Productivity. Predicted changes in the domestic cutoffs m_r^d are negative for all Canadian provinces and for almost all US states, which shows that removing the border induces tougher selection and increases average productivity almost everywhere.[22] The national productivity gain in Canada, with weights given by regions' shares of surviving firms, would be 8.03%, whereas in the US it is much smaller and amounts to only 1.02%. Clearly, since Canada is the smaller economy with less selection prior to the border removal, trade integration has more substantial consequences there. Furthermore, across Canadian provinces we find stronger productivity gains in larger regions and in regions closer to the border.

Predicted changes in export cutoffs m_{rs}^x display a richer pattern than changes in domestic

[21] This is in line with the comparative static results in Section 2.4, where we have shown that trade liberalization leads to wage convergence across two asymmetric regions.

[22] The only exceptions are Alaska, Arizona, and New Mexico.

cutoffs. From the definition of the export cutoff, we have

$$\frac{\widetilde{m}_{rs}^x}{m_{rs}^x} = \left(\frac{\widetilde{\tau}_{ss}/\widetilde{\tau}_{rs}}{\tau_{ss}/\tau_{rs}}\right)\left(\frac{\widetilde{\sigma}_s/\widetilde{\sigma}_r}{\sigma_s/\sigma_r}\right)\frac{\widetilde{m}_s^d}{m_s^d} = \mathrm{e}^{\widehat{\theta} b_{rs}}\left(\frac{\widetilde{\sigma}_s/\widetilde{\sigma}_r}{\sigma_s/\sigma_r}\right)\frac{\widetilde{m}_s^d}{m_s^d}.$$

Using the initial and the counterfactual values of σ_r and m_r^d, along with $\widehat{\theta} = 0.1704$, we can compute the percentage changes in those cutoffs. Their density is depicted in Figure 1. The left panel of Figure 1 shows that when regions r and s are in Canada, cutoff changes are always negative, i.e. $\widetilde{m}_{rs}^x < m_{rs}^x$, with an average of -8.11%. Hence, stronger competition due to trade integration with the US makes it harder for Canadian firms to serve the provinces. In the case of US firms selling to the US, the percentage changes in the cutoffs m_{rs}^x are also almost always negative, with an average of -0.80%.[23]

Insert Figure 1 about here.

The right panel of Figure 1 depicts the density of the percentage changes in the cutoffs m_{rs}^x associated with cross-border sales. As one can see, the export cutoffs increase in all cases. When region r is in Canada and region s is in the US, the percentage changes in m_{rs}^x range from 9.28% to 15.87%, with a mean of 12.97%. In the opposite case, the changes range from 9.89% to 16.24%, with an average of 13.45%. Thus, changes in m_{rs}^x for cross-border operations are comparable in both directions. The reason is that although domestic cutoffs m_s^d decrease more in Canada – thus reducing m_{rs}^x from the US to Canada as compared to those from Canada to the US – Canadian incomes shares σ_s rise to roughly offset the former effect.

(c) Consumption diversity and welfare. Last, we know from our model that regional welfare is proportional to regional consumption diversity, and that a decrease in the cutoff translates into welfare gains. As can be seen from Table 3, variety and welfare gains due to the border removal are larger in Canada than in the US, and they range from about 5.09% to about 13.78%.[24] The corresponding range for the US is from -0.67% to 2.90%. The effects are again more pronounced in Canada, and welfare gains in the US are, on average, larger in regions closer to the border.

As for the welfare impacts of trade liberalization, it is also possible to relate our theoretical framework to the recent results by Arkolakis *et al.* (2012). They show that in a wide class of models, including Krugman (1980), Eaton and Kortum (2002), and Melitz (2003), the welfare effects of a change in trade costs can be expressed by a simple statistic that captures the import

[23]The only exceptions are Alaska, Arizona, and New Mexico. For those destinations we may have $\widetilde{m}_{rs}^x > m_{rs}^x$. This is largely due to the fact that their domestic cutoffs m_s^d increase.

[24]Table 3 reports cardinal percentage changes in welfare. Therefore, they are sensitive to a monotonic transformation of the utility function. However, their ranking reported in the last column of the table is invariant to such a transformation. According to this ranking, British Columbia would gain the most from the border removal, followed by Ontario, Manitoba, and Québec.

penetration rate and the trade elasticity. We show in Appendix G that a similar – though not identical – formula holds in our case.

4.2 The impacts on markups

The aim of this subsection is threefold. First, in part **(a)** we investigate how origin-destination markups $\Lambda_{rs}(m)$, charged by firms located in r selling to market s, react to trade integration. We then consider aggregation of those markups. In part **(b)**, for each consumer in a particular destination, we aggregate origin-destination markups across all origins, using expenditure shares as weights. This yields a measure of consumers' exposure to market power, which is a sufficient statistic for welfare. Finally, in part **(c)**, for each firm in a particular origin, we aggregate $\Lambda_{rs}(m)$ across all destinations, using sales shares as weights. This yields a measure of firms' market power that allows us to make detailed predictions about how firm-level markups change along various margins with trade integration.

(a) Origin-destination markups. The markup charged by a firm located in r and selling to s with productivity $1/m$ is given by expression (10), where $W \equiv W(\mathrm{e}m/m_{rs}^x)$. Hence, conditional on being active in r and selling to s before and after the border removal, $\widetilde{\Lambda}_{rs}(m) \gtreqless \Lambda_{rs}(m)$ if and only if $\widetilde{m}_{rs}^x \gtreqless m_{rs}^x$. The foregoing condition being independent of m, all firms in r selling to s will either increase or decrease their markups in s, irrespective of their productivity.

We have provided a detailed analysis of cutoff changes in Section 4.1. Since the direction of changes in origin-destination markups solely depends on that in cutoffs, it immediately follows that all Canadian firms reduce their markups on domestic sales after the border removal. For US firms, changes in markups on domestic sales are also almost always negative. Put differently, trade integration has pro-competitive effects for domestic transactions. Quite surprisingly, however, trade integration *increases* the markups firms charge on cross-border sales. The intuition is that, by construction of $W(\mathrm{e}m/m_{rs}^x)$, the rise in the export cutoff m_{rs}^x driven by trade integration is equivalent to a productivity gain of a firm with marginal labor requirement m relative to the other firms selling in region s. This result already suggests that markups at the firm-level may change in complicated ways, depending on which markets the firm sells to and depending on the relative importance of domestic versus foreign markets.

Note, finally, that the decrease in markups on domestic sales and the increase in markups on cross-border sales are in accordance with our comparative static results in Section 2.4.

(b) Markups on the consumers' side. To derive a measure of consumers' exposure to market power, we aggregate origin-destination markups $\Lambda_{rs}(m)$ across all origins for each consumer in a particular destination, using expenditure shares as weights. As shown in (28), the resulting measure $\overline{\Lambda}_s^c$ is proportional to m_s^d, which together with the results on the domestic cutoffs in Section 4.1, implies that markups on the consumers' side decrease in all Canadian provinces and in almost all US states. Table 3 reports that markup changes on the consumers' side are substantially smaller in the US than in Canada. The intuition is that the US is a much larger and already more competitive market before the border removal.

Note that the decrease in $\overline{\Lambda}_s^c$ is, again, in line with our comparative static results. Note also that, as can be seen from equation (29), the change in $\overline{\Lambda}_s^c$ is a sufficient static for assessing the welfare change in region s.

(c) Markups on the firms' side. To derive a measure of market power at the firm level, we now aggregate origin-destination markups across all destinations for each firm in a particular origin, using sales shares as weights. This yields firm-level markups, $\overline{\Lambda}_r^p(m)$. Analyzing these markups is important given that recent microeconometric studies look for the pro-competitive effects of international trade at the firm-level (e.g., De Loecker et al., 2012). As available firm-level data typically does not allow to identify markups across different destinations, $\overline{\Lambda}_r^p(m)$ would be the markup the econometrician estimates for each firm.[25]

There are at least two reasons why looking at firm-level markups through the lens of our multi-region model is interesting. First, we can keep track of *compositional effects* that are hard to sort out empirically. Firm-level markups change due to changes in origin-destination markups, changes in the number of markets served, and changes in sales shares across different markets. We quantify all those changes using our model.

Insert Figures 2 and 3 about here.

Second, contrary to pro-competitive effects on the consumers' side, we show that changes in firm-level markups are ambiguous.[26] Interestingly, the correlation between both types of markups at the regional level is shown to be virtually zero. A similar finding holds for the correlation of counterfactual changes in these two types of markups.[27] Our results thus suggest that markups

[25]Although some recent datasets allow to observe the quantity of goods sold to and prices charged in a specific destination, they do not contain information about the value or quantity of inputs used in the production of goods sold to a specific market. Unless one is willing to make strong assumptions, it is thus impossible to identify markups across different destinations. Simonovska (2011) uses, for the case of a Spanish apparel firm, prices of 180 finely defined products sold in eighteen European countries. Although such data allows to observe prices and markup *differences* across markets, it is not informative about markups *levels* due to the lack of cost data.

[26]Zhelobodko *et al.* (2012) obtain similar findings by imposing some restrictions on marginal costs and consumers' relative risk aversion. In our case, the 'anti-competitive' behavior is triggered by trade cost reductions.

[27]The correlation between $\overline{\Lambda}_r^c$ and $\overline{\Lambda}_r^p$ for the 61 regions in the US and Canada before removing the border is -0.0169 and nowhere near statistical significance. The correlation between percentage changes of $\overline{\Lambda}_r^c$ and counterpart percentage changes on the firms' side (computed on stayers only) for the 61 regions in the US and Canada is 0.0009. It is again statistically insignificant.

on the firms' side, which are the focus of a number of recent microeconometric studies (e.g., De Loecker *et al.*, 2012), provide a very different piece of information than markups on the consumers' side, which are central to any welfare statement.

We compute firm-level markups $\overline{\Lambda}_r^p(m)$ for a large sample of firms before and after the border removal (see Appendix H for details). Figure 2 shows that the densities of firm-level markups for Canadian and US firms change little after the border removal. Yet, those densities are not derived from the same sample of firms. The density after the border removal applies to 'stayers' only, i.e., the most productive firms that remain active. In contrast, the density before the border removal includes the 'exiters', i.e., the less productive firms that stop operating after the trade integration. We thus depict in Figure 3 the density of firm-level markups before and after the border removal for stayers only. As can be seen, the distribution shifts to the left after the border removal, and that shift is more pronounced in Canada.

Although Figure 3 suggests that firm-level markups fall 'on average', it provides no information on whether markups go down for all firms and, if not, for which firms they eventually go up. Figure 4 depicts firm-level markups, sales levels and shares, and the number of markets served by firms located in Québec before the border removal. The top pair of graphs reports the relationship between firm productivity and the firm-level markup (left panel), as well as the relationships between firm productivity, sales, and the number of markets served (right panel). Both sales and the number of markets served increase with firm productivity, a straightforward property of our model. However, the left panel reveals that firm-level markups do not necessarily increase with productivity. As productivity rises, a firm sells more and charges higher markups in those markets where it was already selling, i.e., markets for which $m < m_{rs}^x$. This pushes towards an increase of the firm-level markup. Yet, as productivity increases, the firm also starts selling to new markets where it previously could not sell, i.e., markets for which $m > m_{rs}^x$. The markups charged in these new markets are low, hence pushing down the firm-level markup. The relative speed at which sales increase in the old and in the new markets – and the speed at which new markets are added – determine sales shares, which are the weights used for computing $\overline{\Lambda}_r^p(m)$.

Insert Figures 4 and 5 about here.

Going from top to bottom, the remaining three pairs of graphs in Figure 4 show what happens in the three sets of markets Québec firms might be selling to: Canada, the US, and the ROW. For each set we compute the sales-share weighted averages of markups which, by definition, add up to the average firm-level markup when using Canadian, US, and ROW sales shares. The relationships between firm productivity, average markups by set of markets, and sales shares are depicted in the left panel of Figure 4 while the right panel shows the link between firm productivity, sales, and the number of markets served.

Observe that within each set of markets, the relationship between average markups and productivity is more monotonous than across all markets together. As productivity increases, the markups on domestic, on US, and on ROW sales generally increase. A firm in Québec will also typically charge a lower markup on sales to the more distant US markets, and an even lower markup on sales to the even more distant ROW markets. Firms in the middle-to-low productivity range are those with more scope for expansion outside of Canada, and within that productivity range an increase in $1/m$ maps into a large increase in the number of US and ROW markets served, as well as a sharp rise in the share of sales going to these markets. Since firms charge lower markups in these markets, the relationship between $\overline{\Lambda}_r^p(m)$ and productivity is negative within that productivity range.

Turning to our counterfactual results, Figure 5 provides the same information as that contained in the left panel of Figure 4, yet adds two twists. First, we now also consider a US region, New York. Second, we depict firm-level markups and sales shares both before and after the border removal. The top left graph of Figure 5 depicts the relationship between firm productivity and firm-level markups in Québec. As can be seen, the productivity range over which firm-level markups are defined is not the same before and after the border removal. Selection gets tougher after the border removal and the least productive firms leave the market. Comparing the markups of stayers before and after trade integration, i.e., for a given value of productivity, reveals that the most productive firms actually increase their average firm-level markup. Firms in the medium-to-low productivity range instead decrease their firm-level markup.

The top right graph of Figure 5 depicts the same relationships for New York. Compared to Québec, all stayers in New York see their average firm-level markups slightly decrease. The key to understand the difference between the two regions is market size. As can be seen from Figure 5, for firms in New York the US market represents the bulk of their sales. Consequently, the increase in cross-border markups and in the share of sales to Canada are not enough to compensate for the fall of domestic markups. For firms in Québec, the domestic market represents a much smaller share, i.e., they raise their markups on a larger fraction of their sales after the border removal. This is particularly true for the most productive firms, who have a higher share of sales to the US.

Insert Table 5 about here.

The foregoing examples of Québec and New York do reflect deeper regularities of our model. Those are summarized in Table 5, which reports OLS estimates of a simple regression where the dependent variable is the counterfactual percentage change in $\overline{\Lambda}_r^p(m)$. In the first two columns we consider all stayers, whereas in the other two columns we restrict our estimations to stayers that sell to at least one non-domestic market before removing the border ('exporting stayers'). In columns 1 and 3, we only consider firm productivity, along with a set of region dummies, as covariates. In columns 2 and 4, we further consider the change in the number of markets served as well as the share of sales to non-domestic markets. Columns 1 and 3 generalize the insights from Figure 5 that more productive firms are relatively better off in terms of markup changes. This is true for stayers and, especially, for exporting stayers. Columns 2 and 4 also confirm the intuition that firms which increase more the number of markets they serve and which have a lower share of sales to non-domestic markets are worse off in terms of markups they can charge. The importance of the share of sales to non-domestic markets is stronger for exporting stayers (column 4).

4.3 The impacts on regional trade flows

Having investigated the impacts of the border removal on regional aggregates and firm-level markups, we finally turn to the impact of the border removal on inter-regional trade flows. Doing so allows us to revisit the issue of measuring 'border effects' between Canada and the US in a model where firms are heterogeneous and where incomes change with trade integration. To this end, we follow Anderson and van Wincoop (2003) and 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:

$$B_{rs} \equiv \frac{\dfrac{\widetilde{X}_{rs}}{\widetilde{Y}_r \widetilde{Y}_s / \widetilde{Y}_W}}{\dfrac{\widehat{X}_{rs}}{Y_r Y_s / Y_W}} = \mathrm{e}^{k\widehat{\theta} b_{rs}} \left(\frac{\widetilde{\Phi}_r}{\widehat{\Phi}_r} \right)^k \left(\frac{\widetilde{\Phi}_s}{\widehat{\Phi}_s} \right)^k, \tag{46}$$

where size is controlled for by dividing the bilateral trade numbers by $Y_r Y_s / Y_W$. Note that (46) is equivalent to bilateral border effects in Anderson and van Wincoop (2003) except that their elasticity of substitution is replaced with the shape parameter k of the Pareto distribution. Furthermore, expressions (38) and (39) allow us to rewrite the multilateral resistance terms as functions of the labor income shares and the cutoff as follows:

$$B_{rs} = \mathrm{e}^{k\widehat{\theta} b_{rs}} \left(\frac{\widetilde{\sigma}_r}{\sigma_r} \right)^{-k-1} \left(\frac{\widetilde{\sigma}_s}{\sigma_s} \right)^k \left(\frac{\widetilde{m}_s^d}{\widehat{m}_s^d} \right)^{k+1}. \tag{47}$$

Using this expression and the previous results in Section 4.1, we can 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.[28]

What drives bilateral border effects? As can be seen from (47), B_{rs} can be decomposed into the following four components: (i) a *pure border effect* $\mathrm{e}^{k\widehat{\theta} b_{rs}}$; (ii) an *origin income share effect* $(\widetilde{\sigma}_r/\sigma_r)^{-k-1}$; (iii) a *destination income share effect* $(\widetilde{\sigma}_s/\sigma_s)^k$; and (iv) a *selection effect* $(\widetilde{m}_s^d/\widehat{m}_s^d)^{k+1}$. Tables 6 and 7 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 6 and 7 about here.

[28] We could compute 84×84 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.

(a) Canada-US bilateral trade. Table 6 lists the components of B_{rs} for exports from Québec (QC) to all Canadian provinces and US states. Consider, for example, the bilateral border effect with New York (NY). The pure border effect corresponds to the predicted change in bilateral trade flows that would prevail if endogenous changes in income shares and cutoffs were not taken into account. In this example, it states that the value of exports from QC to NY would rise by a factor of 4.2568. Yet, the wage in QC rises relative to that in NY after the border removal, and QC firms thus become less competitive in NY due to relatively higher production costs. This change is captured by the income share effects which decrease the export value from QC to NY by a factor of 0.6659 (0.6478×1.0279). Put differently, neglecting the endogenous wage responses to the border removal leads to overstating the bilateral border effect by about 33%. Finally, there is a selection effect. The border removal reduces the cutoff marginal labor requirement that firms need to match to survive in NY. In other words, trade integration induces tougher selection and makes it harder for QC firms to sell in NY. This selection effect decreases the export value by a factor of 0.8679, i.e., it further reduces the bilateral border effect by about 13%. Putting together the different components, the bilateral border effect is then given by $4.2568 \times 0.6478 \times 1.0279 \times 0.8679 = 2.4601$, which is about 42% less than the pure border effect without endogenous wage and productivity responses.

The top-left panel of Figure 6 depicts the densities of incomes share and selection effects for 510 bilateral trade flows from Canadian provinces to US states. The solid line is the product of the origin and destination income share effects, whereas the dashed line corresponds to the selection effect. Whenever the effects are greater (less) than one they incude larger (smaller) bilateral border effects, and the further away they are from one the larger is their impact on trade flows. In accordance with the QC-NY example, the top-left panel of Figure 6 shows that the income share effects strongly dampen export values because Canadian provinces experience significant wage increases relative to US states. The selection effects are somewhat weaker because the border removal induces little selection in the already competitive US markets.

Insert Figure 6 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 QC to Ontario (ON) in Table 6. There is, of course, no pure border effect for this intranational trade flow, but due to the endogenous changes in the income shares and the cutoff we find a bilateral border effect equal to $1 \times 0.6476 \times 1.6914 \times 0.3023 = 0.3312$. The border removal thus reduces the value of exports from QC to ON by 66.9%. Note that the income share in QC falls relative to that in ON, which provides QC firms with a cost advantage and per se increases exports to ON by around 9.6%. The main effect at work is, however, the tougher selection in ON due to the increased presence of more productive US firms. This makes it much harder for QC firms to sell in ON and reduces the export value by about 70%.[29]

The top-right panel of Figure 6 shows the densities of income share and selection effects across all 100 bilateral trade flows within Canada. It conveys a general message that is similar to the specific QC-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. 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 7 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 $4.2568 \times 0.9698 \times 1.4747 \times 0.4041 = 2.4601$. In this example, the income share in QC rises relative to that 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 6, which shows the densities of income share and selection effects across all 510 trade flows from US states to Canadian provinces, confirms this pattern. Put differently, endogenous wage responses magnify bilateral trade flows from the US to Canada, whereas endogenous productivity responses reduce them.

(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 7 change little after the border removal ($1 \times 0.9698 \times 0.9902 \times 0.9392 = 0.9019$, i.e., a 10% decrease). The explanation is that CA is large and far away from the border, so that little additional selection is induced there, while the income shares in NY and CA change only slightly. The bottom-right panel of Figure 6 confirms that both income share and selection effects are quite small for the 2601 flows within the US. Whereas selection tends to reduce trade flows overall, the impacts of endogenous wage responses can go either way.

To summarize, both income share and selection effects are crucial for assessing how trade flows would change after the border removal. We find that the expansion of Canada-US flows is reduced, on average, by 31.7% due to relative wage increases in Canada as compared to a world where wages would remain unchanged. Furthermore, the expansion in US-Canada flows is reduced, on average, by 55.3% due to productivity increases driven by firm selection in Canada as compared to a world without selection. Neglecting those two empirically important margins of adjustment leads to systematic and strong biases in predicted changes in trade flows and, therefore, in measured border effects.

[29] The induced selection effects are also visible in the bilateral border effect of QC with itself. The value of local sales by QC firms drops by 59.6% due to the tougher selection.

5 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 that consumers face 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 illustrate this advantage, we have structurally estimated a gravity equation subject to general equilibrium conditions, and have simulated the impacts of removing all trade barriers generated by the Canada-US border. We have also exploited the micro-structure of our model to look in depth at changes in average firm-level markups. Contrary to pro-competitive effects on the consumers' side, changes in firm-level markups are ambiguous. This result suggests that markups on the firms' side, which are the focus of a number of recent microeconometric studies, provide a very different piece of information than markups on the consumers' side, which are central to any welfare statement.

Although this paper has focused on the border removal as one specific counterfactual exercise, 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 allows to recover 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 multiple sectors as in Behrens and Murata (2012a) or differential factor proportions in order to cope with a broader international setting including North-South trade. It would also be interesting to embed consumer heterogeneity as in Behrens and Murata (2012b) into the present framework with firm heterogeneity. Another possible extension would be to endogenize regional populations by allowing for interregional and international migration based on utility maximization as in Behrens *et al.* (2011). Taking this road will give rise to a new generation of spatial economics in which theory, structural estimation, and counterfactual experiments are tightly linked.

Acknowledgements. We thank the editor Jan Eeckhout, three anonymous referees, as well as Thomas Chaney, Robert C. Johnson, Alain Gabler, Marc Melitz, Gianmarco Ottaviano, Esteban Rossi-Hansberg, and numerous seminar and conference participants at various institutions for helpful comments and suggestions. Behrens is the 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) and from the Standard Research Grants Program of SSHRC. Mion gratefully acknowledges financial support from the UK Economic and Social Research Council (ESRC). Murata gratefully acknowledges financial support from the Japan Society for the Promotion of Science (17730165, 23730254) and MEXT.ACADEMIC FRONTIER (2006-2010). The usual disclaimer applies.

References

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

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

[3] Arkolakis C., A. Costinot, and A. Rodríguez-Clare (2012) New trade models, same old gains?, *American Economic Review* 102, 94-130.

[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] Atkeson, A. and A. Burstein (2008), Pricing-to-market, trade costs, and international relative prices, *American Economic Review* 98, 1998-2031.

[6] 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.

[7] 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.

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

[9] Balistreri, E.J., R.H. Hillberry and T.F. Rutherford (2011) Structural estimation and solution of international trade models with heterogeneous firms, *Journal of International Economics* 83, 95-108.

[10] Behrens, K, G. Mion, Y. Murata and J. Südekum (2011) Spatial frictions, *CEPR Discussion Paper* #8572.

[11] Behrens, K. and Y. Murata (2012a) Trade, competition, and efficiency, *Journal of International Economics* 87, 1-17.

[12] Behrens, K. and Y. Murata (2012b) Globalization and individual gains from trade, *Cahiers de recherche* 1218, CIRPEE.

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

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

[15] 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.

[16] 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.

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

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

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

[20] 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.

[21] De Loecker, J. (2011) Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity, *Econometrica* 79, 1407-1451.

[22] De Loecker, J., P.K. Goldberg, A.K. Khandelwal and N. Pavcnik (2012) Prices, markups and trade reform, *NBER Working Paper* #17925.

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

[24] 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.

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

[26] Feenstra, R.C. (2010a) Measuring the gains from trade under monopolistic competition, *Canadian Journal of Economics* 43, 1-28.

[27] Feenstra, R.C. (2010b) *Product Variety and the Gains from International Trade*. Cambridge, Mass: MIT Press.

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

[29] Feenstra, R.C. (2003) A homothetic utility function for monopolistic competition models, without constant price elasticity, *Economics Letters* 78, 79-86.

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

[31] 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.

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

[33] Holmes, T.J., W.-T. Hsu and S. Lee (2012) Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade, *mimeo*, University of Minnesota.

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

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

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

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

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

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

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

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

[42] Simonovska, I. (2011) Income differences and prices of tradables, *mimeo*, Princeton University.

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

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

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

[46] 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.

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

[48] Zhelobodko, E., S. Kokovin, M. Parenti and J.-F. Thisse (2012) Monopolistic Competition: Beyond the CES, *Econometrica*, forthcoming.

Appendix

The Appendix is structured as follows: **Appendix A** shows how to derive the demand functions (2) and the firm-level variables (9) using the Lambert W function. In **Appendix B** we provide integrals involving the Lambert W function and derive the terms $\{\kappa_1, \kappa_2, \kappa_3, \kappa_4\}$ that are used in the paper. **Appendix C** contains proofs and computations for the single region case. In **Appendix D** we derive the equilibrium conditions (22)–(24) and provide further derivations for the multi-region case. **Appendix E** deals with the two-region case and establishes the comparative statics results. **Appendix F** discusses how we calibrate the parameter k. **Appendix G** relates our theoretical framework to the recent approach by Arkolakis *et al.* (2012). Finally, **Appendix H** provides our definitions for markups on the firms' side.

Appendix A: Demand functions and firm-level variables

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

$$\alpha \mathrm{e}^{-\alpha q_{sr}(i)} = \lambda p_{sr}(i), \quad \forall i \in \Omega_{sr} \tag{48}$$

and the budget constraint $\sum_s \int_{\Omega_{sr}} p_{sr}(k) q_{sr}(k) \mathrm{d}k = E_r$. Taking the ratio of (48) for $i \in \Omega_{sr}$ and $j \in \Omega_{vr}$ yields

$$q_{sr}(i) = q_{vr}(j) + \frac{1}{\alpha} \ln \left[\frac{p_{vr}(j)}{p_{sr}(i)} \right] \quad \forall i \in \Omega_{sr}, \forall j \in \Omega_{vr}.$$

Multiplying this expression by $p_{vr}(j)$, integrating with respect to $j \in \Omega_{vr}$, and summing across all origin regions v we obtain

$$q_{sr}(i) \sum_v \int_{\Omega_{vr}} p_{vr}(j) \mathrm{d}j = \underbrace{\sum_v \int_{\Omega_{vr}} p_{vr}(j) q_{vr}(j) \mathrm{d}j}_{\equiv E_r} + \frac{1}{\alpha} \sum_v \int_{\Omega_{vr}} \ln \left[\frac{p_{vr}(j)}{p_{sr}(i)} \right] p_{vr}(j) \mathrm{d}j. \tag{49}$$

Using $\overline{p}_r \equiv (1/N_r^c) \sum_v \int_{\Omega_{vr}} p_{vr}(j) \mathrm{d}j$, expression (49) can be rewritten as follows:

$$\begin{aligned} q_{sr}(i) &= \frac{E_r}{N_r^c \overline{p}_r} - \frac{1}{\alpha} \ln p_{sr}(i) + \frac{1}{\alpha N_r^c \overline{p}_r} \sum_v \int_{\Omega_{vr}} \ln \left[p_{vr}(j) \right] p_{vr}(j) \mathrm{d}j \\ &= \frac{E_r}{N_r^c \overline{p}_r} - \frac{1}{\alpha} \ln \left[\frac{p_{sr}(i)}{N_r^c \overline{p}_r} \right] + \frac{1}{\alpha} \sum_v \int_{\Omega_{vr}} \ln \left[\frac{p_{vr}(j)}{N_r^c \overline{p}_r} \right] \frac{p_{vr}(j)}{N_r^c \overline{p}_r} \mathrm{d}j, \end{aligned}$$

which, given the definition of h_r, yields (2).

A.2. Derivation of the firm-level variables (9) and properties of W. Using $p_s^d = m_{rs}^x \tau_{rs} w_r$, the first-order conditions (6) can be rewritten as

$$\ln\left[\frac{m_{rs}^x \tau_{rs} w_r}{p_{rs}(m)}\right] = 1 - \frac{\tau_{rs} m w_r}{p_{rs}(m)}.$$

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

$$\mathrm{e}\frac{m}{m_{rs}^x} = \frac{\tau_{rs} m w_r}{p_{rs}(m)} \mathrm{e}^{\frac{\tau_{rs} m w_r}{p_{rs}(m)}}.$$

Noting that the Lambert W function is defined as $\varphi = W(\varphi)\mathrm{e}^{W(\varphi)}$ and setting $\varphi = \mathrm{e}m/m_{rs}^x$, we obtain $W(\mathrm{e}m/m_{rs}^x) = \tau_{rs} m w_r / p_{rs}(m)$, which implies $p_{rs}(m)$ as given in (9). The expression for quantities $q_{rs}(m) = (1/\alpha)\left[1 - \tau_{rs} m w_r / p_{rs}(m)\right]$ and the expression for operating profits $\pi_{rs}(m) = L_s q_{rs}(m)\left[p_{rs}(m) - \tau_{rs} m w_r\right]$ 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 yields

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

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 $W(\mathrm{e}) = 1$.

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

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

The subscript r can be dropped in the single region case. 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 (expenditure share) weighted average of 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: Equilibrium in the single region case

C.1. 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 (12) 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 (12) is strictly increasing. This uniquely determines the equilibrium cutoff m^d, because

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

C.2. Market size, the equilibrium cutoff, and the mass of entrants. Differentiating (12) 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 (14) with respect to L yields

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

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

C.3. Indirect utility. To derive the indirect utility in the single region case, we first compute the (unweighted) average price across all varieties. Multiplying both sides of (6) by $p(i)$ while dropping the regional subscript, integrating over Ω, and using (3), 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 (4) can be rewritten as $U = N - (\alpha + N\overline{m})/m^d$. When combined with $\overline{m} = [k/(k+1)]m^d$, and with (16) and (17), we obtain the expression for U as given in (18).

Appendix D: Equilibrium in the multi-region case

D.1. Equilibrium conditions using the Lambert W function. We restate the equilibrium conditions for the multi-region case using the Lambert W function.

First, plugging (9) into (19), zero expected profits can be rewritten as

$$\frac{1}{\alpha} \sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right)^{-1} + W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right) - 2\right] \mathrm{d}G_r(m) = F_r. \tag{50}$$

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

Then, using (9), the labor market clearing condition (20) becomes

$$N_r^E \left\{\frac{1}{\alpha} \sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[1 - W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right)\right] \mathrm{d}G_r(m) + F_r\right\} = L_r. \tag{51}$$

Finally, with (9) the trade balance condition (21) is given by

$$\begin{aligned} N_r^E w_r \sum_{s \neq r} L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{rs}^x}\right)^{-1} - 1\right] \mathrm{d}G_r(m) \\ = L_r \sum_{s \neq r} N_s^E \tau_{sr} w_s \int_0^{m_{sr}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{sr}^x}\right)^{-1} - 1\right] \mathrm{d}G_s(m). \end{aligned} \tag{52}$$

Applying the region-specific Pareto distributions $G_r(m) = (m/m_r^{\max})^k$ to (50)–(52) yields, after some algebra and using the results of Appendix B, expressions (22)–(24) given in the main text.

D.2. The mass of varieties consumed. Using N_r^c as defined in (8), and the export cutoff and the mass of entrants as given by (7) and (25), 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 (26), we can then express the mass of varieties consumed in region r as given in (27).

D.3. The (expenditure share) weighted average of markups. Plugging (9) and (10) into the definition (28), the (expenditure share) weighted average of markups in the multi-region case can be rewritten as

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

where the argument $\mathrm{e}m/m_{sr}^x$ of the Lambert W function is suppressed to alleviate notation. As shown in Appendix B, 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 (7) and $E_r = w_r$ yields the expression in (28).

D.4. Indirect utility. To derive the indirect utility, we first compute the (unweighted) average price across all varieties sold in each market. Multiplying both sides of (6) 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 for region r can be rewritten as follows

$$\overline{p}_r = \frac{1}{N_r^c}\sum_s \tau_{sr} w_s \left(\frac{k}{k+1}\right) m_{sr}^x N_s^E G_s(m_{sr}^x) + \frac{\alpha E_r}{N_r^c} = \left(\frac{k}{k+1}\right) p_r^d + \frac{\alpha E_r}{N_r^c}, \tag{53}$$

where we have used (8) and $p_r^d = \tau_{sr} w_s m_{sr}^x$. Plugging (53) into (4) and using (7), the indirect utility is then given by

$$U_r = \frac{N_r^c}{k+1} - \frac{\alpha}{\tau_{rr} m_r^d}, \tag{54}$$

which together with (27) and (28) yields (29).

Appendix E: Some comparative statics

E.1. Existence and uniqueness of equilibrium in the two-region case. Under our assumption that intraregional trade is less costly than interregional trade, i.e., $\tau_{11} < \tau_{21}$ and $\tau_{22} < \tau_{12}$, the RHS of (33) is non-negative if and only if $\underline{\omega} < \omega < \overline{\omega}$, where $\underline{\omega} \equiv \rho^{1/(k+1)} \left(\tau_{22}/\tau_{12}\right)^{k/(k+1)}$ and $\overline{\omega} \equiv \rho^{1/(k+1)} \left(\tau_{21}/\tau_{11}\right)^{k/(k+1)}$. Furthermore, the RHS is strictly decreasing in $\omega \in (\underline{\omega}, \overline{\omega})$ with $\lim_{\omega \to \underline{\omega}+}$ RHS $= \infty$ and $\lim_{\omega \to \overline{\omega}-}$ RHS $= 0$. The LHS of (33) is, on the contrary, strictly increasing in $\omega \in (0, \infty)$. Hence, there exists a unique equilibrium relative wage $\omega^* \in (\underline{\omega}, \overline{\omega})$.

E.2. Differences in market size or technological possibilities. Recalling that $\tau_{12} = \tau_{21} = \tau$ and $\tau_{11} = \tau_{22} = t$ with $t < \tau$, we can prove the comparative statics results in part **(a)** of Section 2.4.

Assume that $\rho = 1$. The equilibrium relative wage ω^* is increasing in L_1/L_2 as an increase in L_1/L_2 raises the RHS of (33) 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 (32), 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 (27)–(29) that a lower cutoff maps into greater consumption diversity, lower (expenditure share) weighted average of markups and higher welfare.

Assume next that $L_1 = L_2$. Since $t < \tau$ holds, the RHS of (33) 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 (33) 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 (32) 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 (27)–(29) that $m_1^d < m_2^d$ implies $N_1^c > N_2^c$, $\overline{\Lambda}_1^c < \overline{\Lambda}_2^c$, and $U_1 > U_2$.

E.3. The impacts of trade integration – symmetric regions. Using (34) and (7) it can be shown that the export cutoff is decreasing in τ:

$$\frac{\partial m_{sym}^x}{\partial \tau} = -\frac{t}{\tau^2(k+1)} \frac{1 + k + (t/\tau)^k}{1 + (t/\tau)^k} m_{sym}^d < 0.$$

E.4. The impacts of trade integration – asymmetric regions. To prove the results in part **(c)** of Section 2.4, we first verify by using (33) that

$$\frac{\partial(\text{RHS})}{\partial\tau} = -\frac{k\rho t^k}{\tau^{k+1}}\frac{L_1}{L_2}\frac{\rho^2-\omega^{2(k+1)}}{[\omega^{k+1}-\rho(t/\tau)^k]^2}\left\{\begin{array}{c} > \\ = \\ < \end{array}\right\}0 \quad \text{for} \quad \left\{\begin{array}{c} \underline{\omega} < \rho^{\frac{1}{k+1}} < \omega^* < \overline{\omega} \\ \underline{\omega} < \omega^* = \rho^{\frac{1}{k+1}} < \overline{\omega} \\ \underline{\omega} < \omega^* < \rho^{\frac{1}{k+1}} < \overline{\omega} \end{array}\right\}. \tag{55}$$

When regions are of equal size, but have different technological possibilities ($\rho > 1$), the first case of (55) applies since $\omega^* > \rho^{1/(k+1)}$ as shown in E.2. Hence, lower interregional trade costs τ 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 (55) 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 cutoffs, and thus (conditional) convergence of consumption diversity, (expenditure share) weighted averages of markups, and welfare between the two regions.

Appendix F: 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 productivity $\widehat{m}_r^d$; 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 wages, 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 share of Canadian exporters is defined in an analogous way.

Export intensity. Let $\widehat{\chi}_{rs}(m) = 1$ if $m < \widehat{m}_{rs}^x$ and $\widehat{\chi}_{rs}(m) = 0$ 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

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

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

Note that the 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:

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

which is again independent of α.

We can now compute the productivity advantage of exporters. To make the sample representative, we draw firms in each region 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 $I = \text{CAN}, \text{US}$ 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 unobserved parameter α. For sample sizes N_{CAN} and N_{US}, we randomly draw int$(\widehat{\theta}_s N_{CAN})$ firms for each Canadian province and int$(\widehat{\theta}_r N_{US})$ firms for each US state from the region-specific productivity distribution, where 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).

We keep the same sample of firms for constructing firm variables after the border removal. Doing so allows us to more cleanly assess the impacts of trade on firms, especially continuing ones. Observe that some firms in the initial sample will no longer operate after the border removal. For those firms, which we call 'exiters', we cannot compute counterfactual sales, markups and export intensities.

Appendix G: The formula for welfare gains

Let $\lambda_{rs} = X_{rs}/Y_s$ be the share of imports from region r in the total expenditure in region s. Since $Y_s = \sum_r X_{rs} = \sum_r X_{sr}$ from the equality of expenditure and income in region s we can write

$$\lambda_{ss} = \frac{X_{ss}}{Y_s} = \frac{X_{ss}}{\sum_v X_{sv}}. \tag{56}$$

Notice that λ_{ss} is the share of internal absorption in region s. Plugging our gravity equation (35) into (56), we get

$$\lambda_{ss} = \frac{L_s L_s \tau_{ss}^{-k} \tau_{ss}^{k+1} w_s (m_s^d)^{k+1} (\mu_s^{\max})^{-1}}{\sum_v L_s L_v \tau_{sv}^{-k} \tau_{vv}^{k+1} w_s (w_v/w_s)^{k+1} (m_v^d)^{k+1} (\mu_s^{\max})^{-1}} = \frac{L_s \tau_{ss} (m_s^d)^{k+1}}{\sum_v L_v \tau_{sv}^{-k} \tau_{vv}^{k+1} (w_v/w_s)^{k+1} (m_v^d)^{k+1}}.$$

Using the equilibrium condition (36), the internal absorption share can be rewritten as $\lambda_{ss} = L_s \tau_{ss} (m_s^d)^{k+1}/\mu_s^{\max}$, from which we immediately obtain $d \ln \lambda_{ss} = (k+1)\, d \ln m_s^d$. Since indirect utility in region s is inversely proportional to the cutoff in that region by (29), we can establish the following relationship: $d \ln U_s = -[1/(k+1)] d \ln \lambda_{ss}$. Let $\epsilon < 0$ denote the trade elasticity, i.e., the elasticity of trade flows with respect to variable trade costs. Notice from (35) that this trade elasticity is equal to $-k$ in our model. Hence, we have

$$\underline{\textit{Our model}} \qquad d \ln U = \frac{1}{\epsilon - 1}\, d \ln \lambda,$$

where we have dropped the subscript s. In the class of models studied by Arkolakis *et al.* (2012), where markups are constant, welfare changes can be expressed by the following formula:

$$\underline{\textit{Arkolakis et al.}} \qquad d \ln U = \frac{1}{\epsilon}\, d \ln \lambda.$$

Thus, our formula for welfare gains under variable demand elasticity and endogenous markups differs from that in Arkolakis *et al.* (2012). In particular, consider a trade liberalization exercise. Conditional on the *same* trade elasticity $\epsilon < 0$, and conditional on the *same* change in the internal absorption share $d \ln \lambda < 0$, our model predicts *smaller* welfare gains than those in Arkolakis *et al.* (2012), despite the presence of pro-competitive effects as an additional source of gains from trade.

Appendix H: Definitions of markups on the firms' side

We provide analytical expressions for the case of Canadian firms only, with mirror expressions holding for US firms. Consider a firm with productivity $1/m$, located in a Canadian province r. Let $\mathcal{S}_r(m)$ be the set of markets the firm sells its product to, including the local market r. Assume that the firm is selling to some Canadian provinces, exports to some US states, and exports to some ROW countries. Some of the sets may be empty, in which case the associated shares of sales are equal to zero. Redefine $\mathcal{S}_r(m)$ as $\mathcal{S}_r^{\text{CAN}}(m) \cup \mathcal{S}_r^{\text{US}}(m) \cup \mathcal{S}_r^{\text{ROW}}(m)$. The (revenue-weighted) average firm-level markup *on domestic sales* is defined as:

$$\overline{\Lambda}_r^{\text{CAN}}(m) \equiv \sum_{s \in \mathcal{S}_r^{\text{CAN}}(m)} \theta_{rs}^{\text{CAN}} \Lambda_{rs}(m), \quad \text{with} \quad \theta_{rs}^{\text{CAN}}(m) \equiv \frac{p_{rs}(m) q_{rs}(m)}{\sum_{v \in \mathcal{S}_r^{\text{CAN}}(m)} p_{rv}(m) q_{rv}(m)}.$$

Analogously, the average firm-level markup *on exports to US states* can be defined as:

$$\overline{\Lambda}_r^{\text{US}}(m) \equiv \sum_{s \in \mathcal{S}_r^{\text{US}}(m)} \theta_{rs}^{\text{US}} \Lambda_{rs}(m), \quad \text{with} \quad \theta_{rs}^{\text{US}}(m) \equiv \frac{p_{rs}(m) q_{rs}(m)}{\sum_{v \in \mathcal{S}_r^{\text{US}}(m)} p_{rv}(m) q_{rv}(m)},$$

with the equivalent expression for *exports to the ROW countries* being:

$$\overline{\Lambda}_r^{\text{ROW}}(m) \equiv \sum_{s \in \mathcal{S}_r^{\text{ROW}}(m)} \theta_{rs}^{\text{ROW}} \Lambda_{rs}(m), \quad \text{with} \quad \theta_{rs}^{\text{ROW}}(m) \equiv \frac{p_{rs}(m) q_{rs}(m)}{\sum_{v \in \mathcal{S}_r^{\text{ROW}}(m)} p_{rv}(m) q_{rv}(m)}.$$

The average firm-level markup on all markets, which we simply refer to as the average firm-level markup, can thus be defined and decomposed as follows:

$$\begin{aligned}
\overline{\Lambda}_r^{p}(m) &\equiv \sum_{s \in \mathcal{S}_r(m)} \frac{p_{rs}(m) q_{rs}(m)}{\sum_{s \in \mathcal{S}_r(m)} p_{rs}(m) q_{rs}(m)} \Lambda_{rs}(m) \\
&= \theta_r^{\text{CAN}}(m) \overline{\Lambda}_r^{\text{CAN}}(m) + \theta_r^{\text{US}}(m) \overline{\Lambda}_r^{\text{US}}(m) + \theta_r^{\text{ROW}}(m) \overline{\Lambda}_r^{\text{ROW}}(m), \qquad (57)
\end{aligned}$$

with $\theta_r^{\text{CAN}}(m)$, $\theta_r^{\text{US}}(m)$, and $\theta_r^{\text{ROW}}(m)$ being the share of sales going to Canada, to the US, and to ROW countries, respectively, and $\theta_r^{\text{CAN}}(m) + \theta_r^{\text{US}}(m) + \theta_r^{\text{ROW}}(m) = 1$. Let $\chi_{rs}(m) = 1$ if $m < m_{rs}^x$ and $\chi_{rs}(m) = 0$ otherwise. Denoting firm sales by $\text{sls}_r(m) \equiv \sum_s \chi_{rs}(m) L_s p_{rs}(m) q_{rs}(m)$ we can also define the (sales share) weighted average of firm-level markups in region r, i.e., the counterpart of $\overline{\Lambda}_r^c$ on the firms' side as:

$$\overline{\Lambda}_r^{p} \equiv \frac{1}{G(m_r^d)} \int_0^{\max_s m_{rs}^x} \frac{\text{sls}_r(m)}{\text{sls}_r} \overline{\Lambda}_r^{p}(m) \mathrm{d}G_r(m), \quad \text{with} \quad \text{sls}_r \equiv \int_0^{\max_s m_{rs}^x} \text{sls}_r(m) \mathrm{d}G_r(m).$$

As highlighted by equation (57), the average firm-level markup depends on: (i) the set of markets a firm sells its product to; (ii) the share of sales to those markets; and (iii) the markups charged in each destination. These three elements are each affected by the border removal, and are idiosyncratic to firms' productivity and location. Using the random sample of firms drawn from the fitted productivity distributions of our model (see Appendix F for more details) we compute average firm-level markups, as well as sales levels and shares and the numbers of markets served, both before and after the border removal. Counterfactual values and changes in those quantities are defined over the subsample of firms which are still selling to at least one market after the border removal ('stayers'). Indeed, given that most firm-level datasets do not provide full information for the universe of firms, and given that an empirical analysis typically requires the use of firm fixed effects, i.e., comparing the same firms before and after, the sample of 'stayers' resembles most closely what an econometrician would typically use for policy evaluation.

Table 1: Estimation of the gravity equation system

	Benchmark(1)	Fixed Effects(2)	Robustness(3)	Robustness(4)
Regions (in gravity)	84 (40)	84 (40)	84 (40)	81 (40)
Trade flows	1560	1560	1560	1560
k	8.5	—	8.5	7.5
Internal distance	Surface	Surface	AvW	Surface
Estimation procedure	AvW	OLS	AvW	AvW
$\ln d_{rs}$	-1.3049^a	-1.2412^a	-1.2801^a	-1.3040^a
	(0.0401)	(0.0419)	(0.0397)	(0.0333)
(implied value of γ)	0.1535^a		0.1506^a	0.1739^a
	(0.0047)		(0.0047)	(0.0044)
b_{rs}	-1.4484^a	-1.4509^a	-1.4493^a	-1.4463^a
	(0.0580)	(0.0670)	(0.0625)	(0.0573)
(implied value of θ)	0.1704^a		0.1705^a	0.1928^a
	(0.0069)		(0.0074)	(0.0076)
$0-$ dummy	-17.1974^a	-16.9760^a	-17.3819^a	-17.1982^a
	(0.1482)	(0.1501)	(0.1718)	(0.1513)
constant	-28.8338^a	-25.7841^a	-28.7435^a	-28.8373^a
	(0.0306)	(0.2342)	(0.0375)	(0.0292)
Shares of firms:				
US exporters	5.19%		2.69%	6.76%
CA exporters	15.81%		4.28%	11.11%
Productivity advantage:				
US exporters	33.45%		39.45%	34.01%
CA exporters	26.56%		33.25%	38.76%
Avg. export intensity:				
US exporters	9.19%		7.55%	12.07%
CA exporters	29.01%		21.24%	26.46%

Notes: Bootstrapped standard errors (with 200 replications) of the parameters computed using the procedure by Anderson and van Wincoop (2003, henceforth AvW) are reported in parentheses. All specifications exclude intra-regional flows X_{rr} and include the same 40 states and provinces as in AvW. 'Surface' refers to the surface-based measure of internal distance, following Redding and Venables (2004), whereas AvW refers to Anderson and van Wincoop's (2003) measure. Coefficients significant at 10% level (^{c}), 5% level (^{b}), and 1% level (^{a}). The productivity advantage of exporters is computed from a random sample drawn representatively from all states and provinces in the calibrated model (200,000 US firms and 2,000 Canadian firms; see Appendix F). Productivity is measured by revenue (i.e., value added in our model) per employee. Export intensity is computed conditional on exporting.

Table 2: Export intensity distribution for Benchmark(1)

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	69.99	29.15
10-20	16	19	18.48	9.14
20-30	7.7	4.2	6.77	11.22
30-40	4.4	0.0	1.33	22.38
40-50	2.4	0.0	1.24	15.52
50-60	1.5	0.0	1.61	3.51
60-70	1.0	0.0	0.58	2.94
70-80	0.6	0.0	0.0	1.83
80-90	0.5	0.0	0.0	2.28
90-100	0.7	0.0	0.0	2.02

Notes: Export intensity is defined as in Appendix F as the firm's share of export revenue in total revenue, conditional upon exporting something. Figures in columns 1 and 2 are provided by Bernard *et al.* (2003, henceforth BEJK). Column 1 reports the observed distribution using 1992 Census of Manufactures data, whereas column 2 provides the simulation results obtained by BEJK. Columns 3 and 4 provide our own simulation results for Benchmark(1) with $k = 8.5$ for the US and for Canada.

Table 3: Impacts of removing all trade barriers generated by the Canada-US border

	Income shares $\Delta\sigma_r\%$	Cutoffs and Markups $\Delta m_r^d\%$ and $\Delta\overline{\Lambda}_r^c\%$	Varieties and Welfare $\Delta N_r^c\%$ and $\Delta U_r^*\%$	Rank of $\Delta U_r^*\%$
States/Provinces	**In Gravity sample**			
Alberta	3.2171	-6.6470	7.1202	6
British Columbia	6.5555	-12.1108	13.7796	1
Manitoba	4.9904	-9.6118	10.6339	3
New Brunswick	3.0668	-6.3887	6.8248	7
Newfoundland	2.1811	-4.8455	5.0922	10
Nova Scotia	2.9674	-6.2175	6.6297	8
Ontario	6.3780	-11.8327	13.4208	2
Prince Edward Island	2.6528	-5.6722	6.0133	9
Quebec	4.6759	-9.0965	10.0068	4
Saskatchewan	4.3287	-8.5225	9.3165	5
Alabama	-0.2077	-0.4834	0.4858	43
Arizona	-0.4726	0.0190	-0.0190	59
California	-0.1152	-0.6579	0.6623	36
Florida	-0.1185	-0.6518	0.6561	37
Georgia	-0.1717	-0.5513	0.5544	38
Idaho	-0.3567	-0.2013	0.2017	54
Illinois	-0.0499	-0.7808	0.7870	30
Indiana	-0.0250	-0.8278	0.8347	29
Kentucky	-0.1025	-0.6818	0.6865	34
Louisiana	-0.2839	-0.3392	0.3404	47
Maine	0.6351	-2.0567	2.0999	13
Maryland	0.2168	-1.2805	1.2972	21
Massachusetts	0.3954	-1.6131	1.6395	16
Michigan	0.1615	-1.1774	1.1914	25
Minnesota	-0.1743	-0.5465	0.5495	40
Missouri	-0.1946	-0.5083	0.5108	42
Montana	-0.2911	-0.3256	0.3267	48
New Hampshire	0.3718	-1.5692	1.5942	17
New Jersey	0.3368	-1.5041	1.5271	19
New York	0.3237	-1.4798	1.5021	20
North Carolina	-0.1138	-0.6606	0.6650	35
North Dakota	-0.2543	-0.3953	0.3969	46
Ohio	0.0999	-1.0620	1.0734	27
Pennsylvania	0.1910	-1.2325	1.2479	23
Tennessee	-0.1735	-0.5481	0.5511	39
Texas	-0.3597	-0.1955	0.1959	55
Vermont	0.5875	-1.9688	2.0083	14
Virginia	-0.0193	-0.8384	0.8455	28
Washington	0.8188	-2.3945	2.4532	12
Wisconsin	-0.0985	-0.6895	0.6943	33
States/Provinces	**Out of Gravity sample**			
Alaska	-0.8131	0.6706	-0.6661	61
Arkansas	-0.3520	-0.2102	0.2107	52
Colorado	-0.3559	-0.2028	0.2032	53
Connecticut	0.3678	-1.5618	1.5866	18
Delaware	0.2166	-1.2802	1.2968	22
Hawaii	0.1764	-1.2051	1.2198	24
Iowa	-0.2195	-0.4612	0.4633	44
Kansas	-0.3080	-0.2937	0.2946	49
Mississippi	-0.3191	-0.2726	0.2733	51
Nebraska	-0.3164	-0.2777	0.2784	50
Nevada	0.1557	-1.1664	1.1802	26
New Mexico	-0.5386	0.1447	-0.1445	60
Oklahoma	-0.3724	-0.1715	0.1718	56
Oregon	-0.0899	-0.7056	0.7106	32
Rhode Island	0.4297	-1.6767	1.7053	15
South Carolina	-0.1893	-0.5183	0.5210	41
South Dakota	-0.3798	-0.1574	0.1576	57
Utah	-0.4259	-0.0699	0.0699	58
West Virginia	-0.0715	-0.7403	0.7458	31
Wyoming	-0.2510	-0.4015	0.4031	45
District of Columbia	1.0513	-2.8196	2.9014	11

Notes: See Section 4 for details on computations.

Table 4: Determinants of changes in regional aggregates

	Dependent variable		
	Income shares $\Delta\sigma_r\%$	Cutoffs and Markups $\Delta m_r^d\%$ and $\Delta\overline{\Lambda}_r^c\%$	Varieties and Welfare $\Delta N_r^c\%$ and $\Delta U_r^*\%$
Regressor	**Estimated coefficients specification 1**, (All regions, $N = 61$)		
DISTANCE TO BORDER (log)	-0.6035^a	1.0762^a	-1.1732^a
	(0.1667)	(0.2842)	(0.3289)
SIZE (log)	0.1871^b	-0.3115^b	0.3713^b
	(0.0849)	(0.1428)	(0.1682)
US-dummy	-3.9207^a	6.8885^a	-7.6535^a
	(0.3759)	(0.6186)	(0.7486)
Constant	5.1392^a	-10.2596^a	10.7928^a
	(1.3228)	(2.2655)	(2.6053)
R^2	0.8958	0.9038	0.8931
Regressor	**Estimated coefficients specification 2**, (All regions, $N = 61$)		
DISTANCE TO BORDER (log)	-1.2452^a	2.0745^a	-2.4706^a
	(0.3983)	(0.6417)	(0.7986)
DISTANCE TO BORDER (log) $\times$ US-dummy	0.8412^b	-1.3176^c	1.6978^b
	(0.4115)	(0.6704)	(0.8226)
SIZE (log)	0.6547^a	-1.0865^a	1.3005^a
	(0.2242)	(0.3642)	(0.4482)
SIZE (log) $\times$ US-dummy	-0.6405^a	1.0578^a	-1.2739^a
	(0.2296)	(0.3759)	(0.4582)
US-dummy	0.0317	-0.0819	0.0556
	(5.0650)	(8.2789)	(10.1106)
Constant	2.4290	-5.3686	5.5448
	(4.9858)	(8.1100)	(9.9652)
R^2	0.9441	0.9452	0.9437

Notes: See Section 4 for additional details on computations. Coefficients significant at 10% level (^{c}), 5% level (^{b}), and 1% level (^{a}). Robust standard errors in parenthesis.

Table 5: Determinants of changes in average firm-level markups

	Dependent variable $\Delta\overline{\Lambda}_r^p(m)\%$			
	All stayers		Exporting stayers	
Productivity	0.0071[a]	0.0050[a]	0.0138[a]	-0.0264[a]
	(0.0002)	(0.0002)	(0.0010)	(0.0011)
Change in the number of markets served		-0.0009[a]		-0.0010[a]
		(0.0000)		(0.0000)
Share of sales to non-domestic markets		0.0265[a]		0.0565[a]
		(0.0013)		(0.0015)
Constant	-0.0066[a]	-0.0052[a]	-0.0184[a]	0.0234[a]
	(0.0001)	(0.0001)	(0.0011)	(0.0015)
Region dummies	Yes	Yes	Yes	Yes
Observations	193,577	193,577	13,535	13,535
R^2	0.7009	0.7494	0.5807	0.7550

Notes: Coefficients significant at 10% level ([c]), 5% level ([b]), and 1% level ([a]). Robust standard errors in parenthesis.

Table 6: Decomposition of bilateral border effects with Québec as exporter

	Pure border $e^{\widehat{k\theta} b_{rs}}$	Origin income share $(\widetilde{\sigma}_r/\sigma_r)^{-k-1}$	Destination income share $(\widetilde{\sigma}_s/\sigma_s)^k$	Selection $(\widetilde{m}_s^d/\widehat{m}_s^d)^{k+1}$	Bilateral border B_{rs}
Importer:	**In Gravity sample**				
Alberta	1.0000	0.6478	1.3089	0.5203	0.4411
British Columbia	1.0000	0.6478	1.7155	0.2934	0.3260
Manitoba	1.0000	0.6478	1.5128	0.3829	0.3752
New Brunswick	1.0000	0.6478	1.2927	0.5341	0.4473
Newfoundland	1.0000	0.6478	1.2013	0.6238	0.4855
Nova Scotia	1.0000	0.6478	1.2822	0.5434	0.4514
Ontario	1.0000	0.6478	1.6914	0.3023	0.3312
Prince Edward Island	1.0000	0.6478	1.2493	0.5742	0.4647
Quebec	1.0000	0.6478	1.4747	0.4041	0.3861
Saskatchewan	1.0000	0.6478	1.4336	0.4290	0.3985
Alabama	4.2568	0.6478	0.9825	0.9550	2.5874
Arizona	4.2568	0.6478	0.9605	1.0018	2.6536
California	4.2568	0.6478	0.9902	0.9392	2.5648
Florida	4.2568	0.6478	0.9900	0.9398	2.5656
Georgia	4.2568	0.6478	0.9855	0.9488	2.5786
Idaho	4.2568	0.6478	0.9701	0.9810	2.6244
Illinois	4.2568	0.6478	0.9958	0.9282	2.5489
Indiana	4.2568	0.6478	0.9979	0.9241	2.5429
Kentucky	4.2568	0.6478	0.9913	0.9371	2.5617
Louisiana	4.2568	0.6478	0.9761	0.9682	2.6063
Maine	4.2568	0.6478	1.0553	0.8208	2.3887
Maryland	4.2568	0.6478	1.0186	0.8848	2.4852
Massachusetts	4.2568	0.6478	1.0341	0.8569	2.4435
Michigan	4.2568	0.6478	1.0138	0.8936	2.4982
Minnesota	4.2568	0.6478	0.9853	0.9493	2.5792
Missouri	4.2568	0.6478	0.9836	0.9527	2.5842
Montana	4.2568	0.6478	0.9755	0.9695	2.6081
New Hampshire	4.2568	0.6478	1.0320	0.8605	2.4490
New Jersey	4.2568	0.6478	1.0290	0.8659	2.4571
New York	4.2568	0.6478	1.0279	0.8679	2.4601
North Carolina	4.2568	0.6478	0.9904	0.9390	2.5644
North Dakota	4.2568	0.6478	0.9786	0.9631	2.5989
Ohio	4.2568	0.6478	1.0085	0.9035	2.5129
Pennsylvania	4.2568	0.6478	1.0164	0.8889	2.4913
Tennessee	4.2568	0.6478	0.9854	0.9491	2.5790
Texas	4.2568	0.6478	0.9698	0.9816	2.6252
Vermont	4.2568	0.6478	1.0511	0.8279	2.3995
Virginia	4.2568	0.6478	0.9984	0.9231	2.5415
Washington	4.2568	0.6478	1.0718	0.7943	2.3477
Wisconsin	4.2568	0.6478	0.9917	0.9364	2.5607
Importer:	**Out of Gravity sample**				
Alaska	4.2568	0.6478	0.9330	1.0656	2.7414
Arkansas	4.2568	0.6478	0.9705	0.9802	2.6233
Colorado	4.2568	0.6478	0.9701	0.9809	2.6242
Connecticut	4.2568	0.6478	1.0317	0.8611	2.4499
Delaware	4.2568	0.6478	1.0186	0.8848	2.4852
Hawaii	4.2568	0.6478	1.0151	0.8912	2.4947
Iowa	4.2568	0.6478	0.9815	0.9570	2.5903
Kansas	4.2568	0.6478	0.9741	0.9724	2.6123
Mississippi	4.2568	0.6478	0.9732	0.9744	2.6150
Nebraska	4.2568	0.6478	0.9734	0.9739	2.6144
Nevada	4.2568	0.6478	1.0133	0.8945	2.4996
New Mexico	4.2568	0.6478	0.9551	1.0138	2.6704
Oklahoma	4.2568	0.6478	0.9688	0.9838	2.6284
Oregon	4.2568	0.6478	0.9924	0.9349	2.5586
Rhode Island	4.2568	0.6478	1.0371	0.8516	2.4356
South Carolina	4.2568	0.6478	0.9840	0.9518	2.5829
South Dakota	4.2568	0.6478	0.9682	0.9851	2.6302
Utah	4.2568	0.6478	0.9644	0.9934	2.6418
West Virginia	4.2568	0.6478	0.9939	0.9318	2.5541
Wyoming	4.2568	0.6478	0.9789	0.9625	2.5981
District of Columbia	4.2568	0.6478	1.0930	0.7621	2.2969

Notes: Border effects are decomposed as indicated by (47).

Table 7: Decomposition of bilateral border effects with New York as exporter

	Pure border $e^{\widehat{k\theta} b_{rs}}$	Origin income share $(\widetilde{\sigma}_r/\sigma_r)^{-k-1}$	Destination income share $(\widetilde{\sigma}_s/\sigma_s)^k$	Selection $(\widetilde{m}_s^d/\widehat{m}_s^d)^{k+1}$	Bilateral border B_{rs}
Importer:			**In Gravity sample**		
Alberta	4.2568	0.9698	1.3089	0.5203	2.8110
British Columbia	4.2568	0.9698	1.7155	0.2934	2.0774
Manitoba	4.2568	0.9698	1.5128	0.3829	2.3910
New Brunswick	4.2568	0.9698	1.2927	0.5341	2.8502
Newfoundland	4.2568	0.9698	1.2013	0.6238	3.0937
Nova Scotia	4.2568	0.9698	1.2822	0.5434	2.8764
Ontario	4.2568	0.9698	1.6914	0.3023	2.1106
Prince Edward Island	4.2568	0.9698	1.2493	0.5742	2.9613
Quebec	4.2568	0.9698	1.4747	0.4041	2.4601
Saskatchewan	4.2568	0.9698	1.4336	0.4290	2.5390
Alabama	1.0000	0.9698	0.9825	0.9550	0.9099
Arizona	1.0000	0.9698	0.9605	1.0018	0.9332
California	1.0000	0.9698	0.9902	0.9392	0.9019
Florida	1.0000	0.9698	0.9900	0.9398	0.9022
Georgia	1.0000	0.9698	0.9855	0.9488	0.9068
Idaho	1.0000	0.9698	0.9701	0.9810	0.9229
Illinois	1.0000	0.9698	0.9958	0.9282	0.8964
Indiana	1.0000	0.9698	0.9979	0.9241	0.8942
Kentucky	1.0000	0.9698	0.9913	0.9371	0.9008
Louisiana	1.0000	0.9698	0.9761	0.9682	0.9165
Maine	1.0000	0.9698	1.0553	0.8208	0.8400
Maryland	1.0000	0.9698	1.0186	0.8848	0.8739
Massachusetts	1.0000	0.9698	1.0341	0.8569	0.8593
Michigan	1.0000	0.9698	1.0138	0.8936	0.8785
Minnesota	1.0000	0.9698	0.9853	0.9493	0.9070
Missouri	1.0000	0.9698	0.9836	0.9527	0.9088
Montana	1.0000	0.9698	0.9755	0.9695	0.9172
New Hampshire	1.0000	0.9698	1.0320	0.8605	0.8612
New Jersey	1.0000	0.9698	1.0290	0.8659	0.8641
New York	1.0000	0.9698	1.0279	0.8679	0.8651
North Carolina	1.0000	0.9698	0.9904	0.9390	0.9018
North Dakota	1.0000	0.9698	0.9786	0.9631	0.9140
Ohio	1.0000	0.9698	1.0085	0.9035	0.8837
Pennsylvania	1.0000	0.9698	1.0164	0.8889	0.8761
Tennessee	1.0000	0.9698	0.9854	0.9491	0.9069
Texas	1.0000	0.9698	0.9698	0.9816	0.9232
Vermont	1.0000	0.9698	1.0511	0.8279	0.8438
Virginia	1.0000	0.9698	0.9984	0.9231	0.8937
Washington	1.0000	0.9698	1.0718	0.7943	0.8256
Wisconsin	1.0000	0.9698	0.9917	0.9364	0.9005
Importer:			**Out of Gravity sample**		
Alaska	1.0000	0.9698	0.9330	1.0656	0.9641
Arkansas	1.0000	0.9698	0.9705	0.9802	0.9225
Colorado	1.0000	0.9698	0.9701	0.9809	0.9228
Connecticut	1.0000	0.9698	1.0317	0.8611	0.8615
Delaware	1.0000	0.9698	1.0186	0.8848	0.8740
Hawaii	1.0000	0.9698	1.0151	0.8912	0.8773
Iowa	1.0000	0.9698	0.9815	0.9570	0.9109
Kansas	1.0000	0.9698	0.9741	0.9724	0.9186
Mississippi	1.0000	0.9698	0.9732	0.9744	0.9196
Nebraska	1.0000	0.9698	0.9734	0.9739	0.9194
Nevada	1.0000	0.9698	1.0133	0.8945	0.8790
New Mexico	1.0000	0.9698	0.9551	1.0138	0.9391
Oklahoma	1.0000	0.9698	0.9688	0.9838	0.9243
Oregon	1.0000	0.9698	0.9924	0.9349	0.8998
Rhode Island	1.0000	0.9698	1.0371	0.8516	0.8565
South Carolina	1.0000	0.9698	0.9840	0.9518	0.9083
South Dakota	1.0000	0.9698	0.9682	0.9851	0.9250
Utah	1.0000	0.9698	0.9644	0.9934	0.9290
West Virginia	1.0000	0.9698	0.9939	0.9318	0.8982
Wyoming	1.0000	0.9698	0.9789	0.9625	0.9137
District of Columbia	1.0000	0.9698	1.0930	0.7621	0.8077

Notes: Border effects are decomposed as indicated by (47).

Figure 1: Density of changes in the cutoffs m_{rs}^x across Canadian provinces and US states

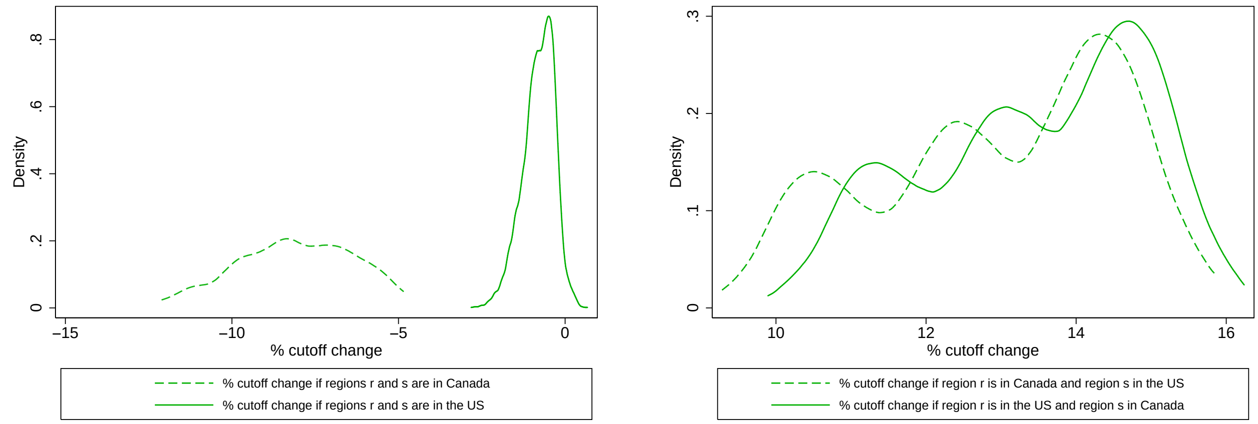

Figure 2: Density of average firm-level markups in Canada and the US before and after the border removal (all active firms).

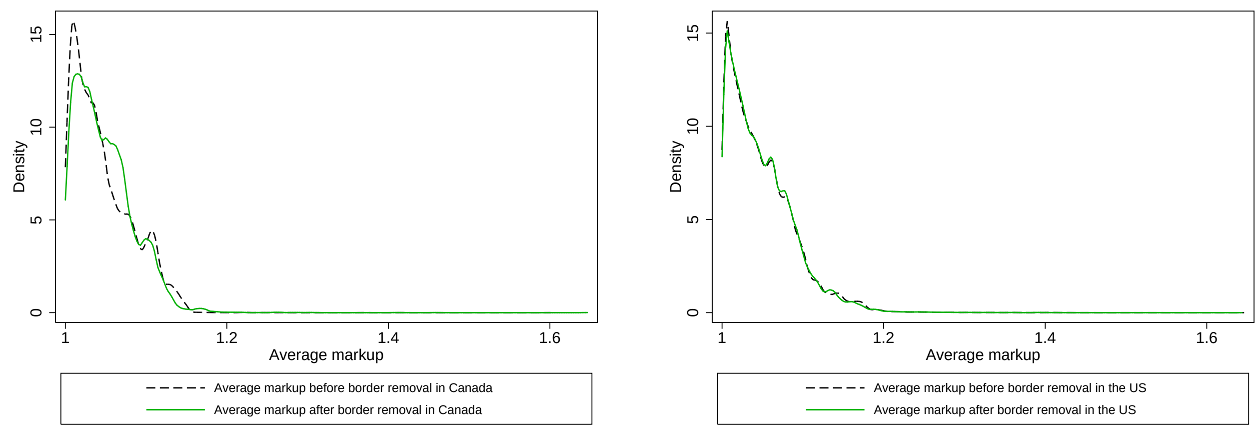

Figure 3: Density of average firm-level markups in Canada and in the US before and after the border removal (stayers only).

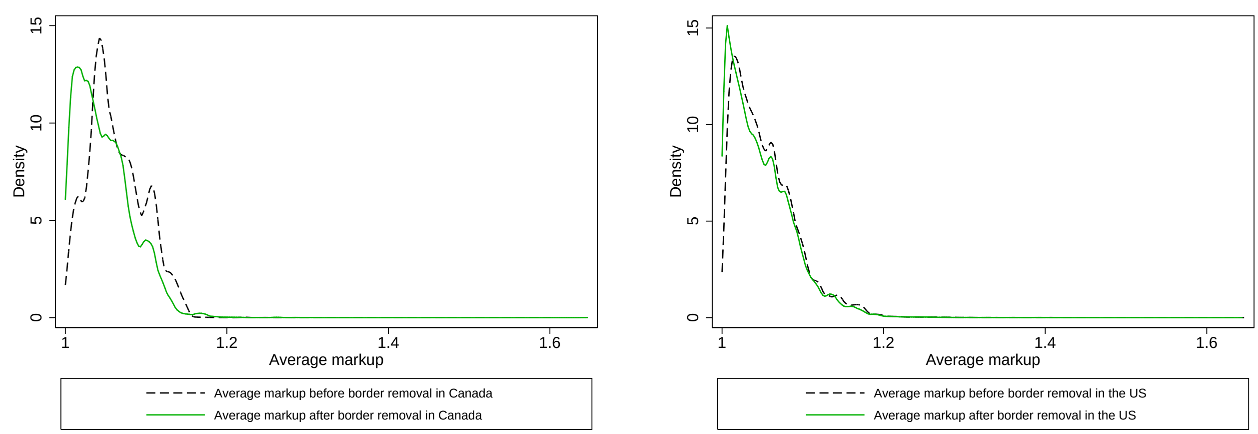

Figure 4: Markups, sales levels and shares, and the number of markets served by firms located in Québec before the border removal.

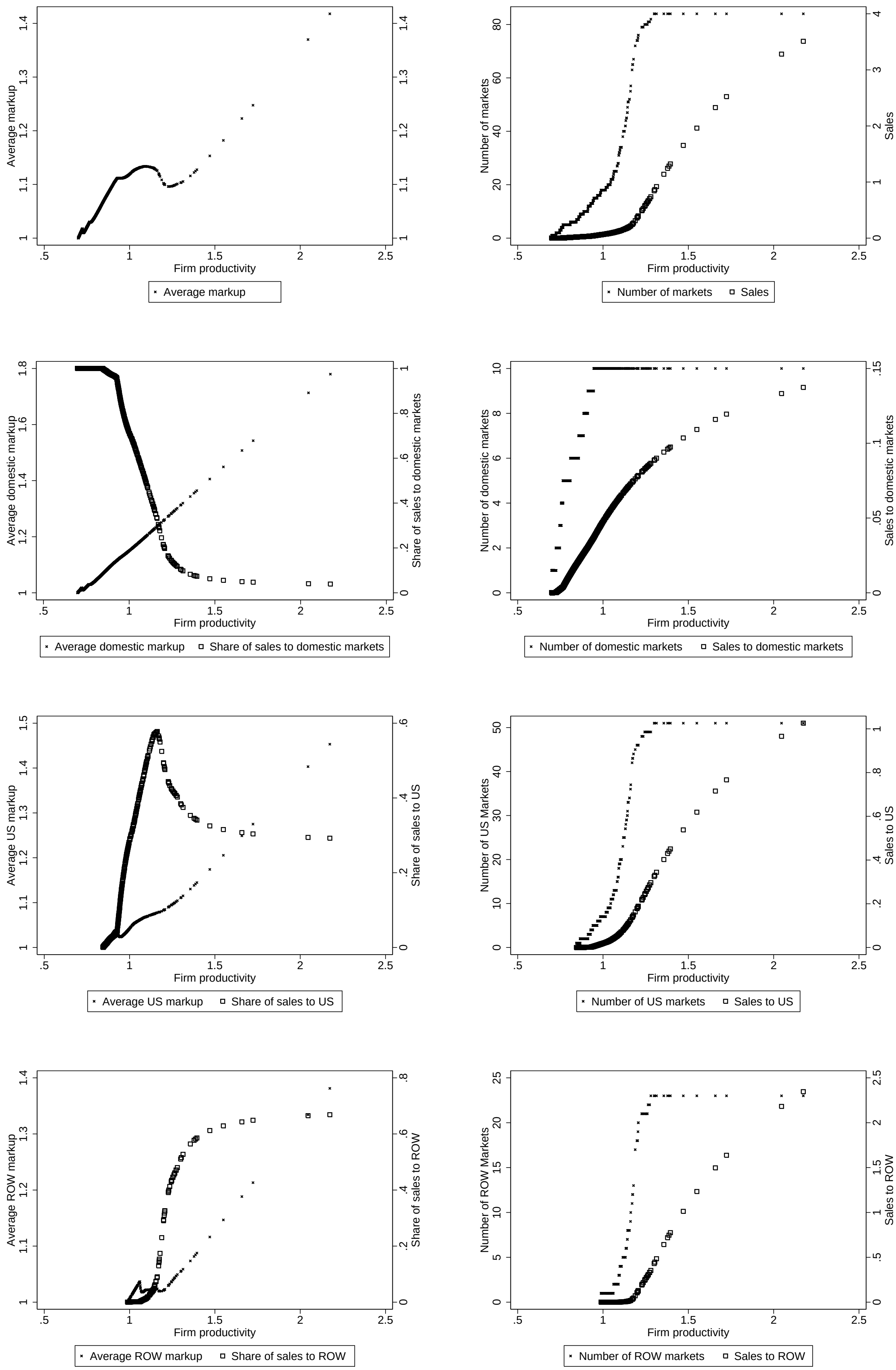

Figure 5: Markups and sales shares for firms located in Québec and in New York before and after the border removal (all active firms).

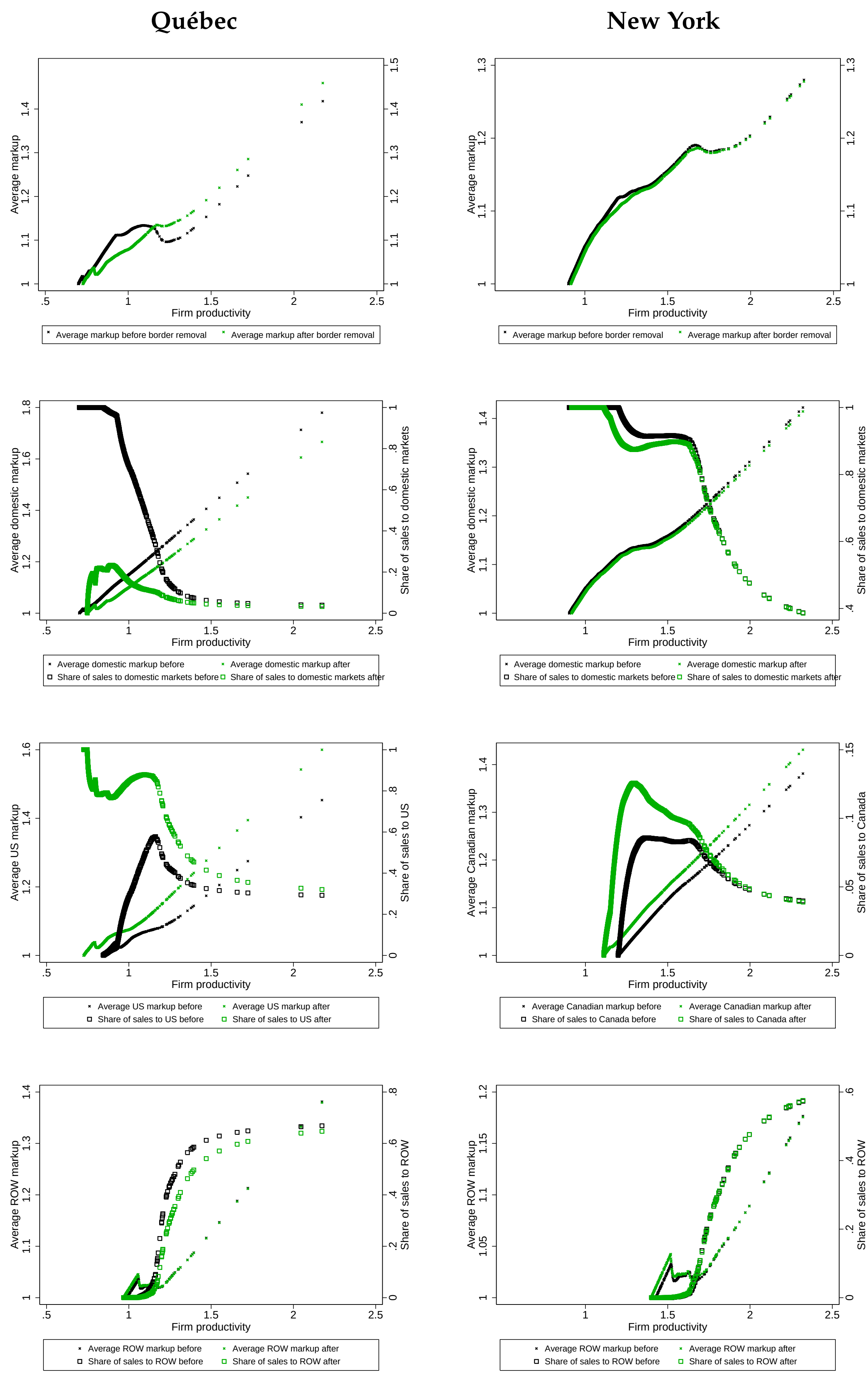

Figure 6: Density of income share and selection effects