Heterogeneous Internal Trade Cost and Its Implications in Trade

Han Zheng*

September 2, 2022

Abstract

Quantitative models of international trade typically assume that trade cost within a country (internal trade cost) is the same across countries. This paper presents evidence for heterogenous internal trade costs across countries and incorporates them into a standard quantitative trade model. Allowing heterogeneous internal trade costs improves the model's ability to predict prices and technology in data. This model also offers a novel answer to the question why small countries export less than large ones. That is small countries trade more with themselves because of lower internal trade costs they have.

Keywords: Heterogeneous internal trade cost, Domestic trade friction, Ratio type gravity estimation

JEL Codes: F10, F14, F17

*Hitotsubashi Institute for Advanced Study, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Email address: hanzheng211@gmail.com.

1 Introduction

A convenient assumption in international trade literature is that the internal trade cost is the same across countries, whereas little evidence is shown to warrant this assumption and on the contrary, intuitions and both anecdotal and academic evidences hint at the opposite—advanced countries generally have better quality of infrastructure and goods have to travel longer distance in large countries. Many studies also analyze the presence and implication of the internal trade cost, such as Anderson and Van Wincoop (2003), Donaldson and Hornbeck (2016) and Donaldson (2018).[1] A common assumption shared by these three studies is that the trade flow between sub-national regions also follow gravity equation structure and trade cost between them is shown to be heterogeneous across both country border and time dimension. Provided that, it is natural to deduce that country's internal trade cost should be heterogeneous across countries. Furthermore, this paper documents a new stylized fact for European Union countries that internal trade cost tends to increase with country size. All the above deduction and evidence contradict to the traditional country level gravity equation estimation practice assuming internal trade cost is the same across countries and normalized as one.

To resolve this discrepancy, this paper develops a model of aggregation to show that the internal trade cost is heterogeneous across countries and it is an increasing function in country's size, a point aligning well with the stylized fact. Guided by the theory, then this paper extends the standard country level gravity estimation to accommodate internal trade cost and yields an internal trade cost measure aligns well with theory prediction. Using the estimated internal trade cost measure and commonly available data, this paper shows not only this model with internal trade cost can reconcile with price data as emphasized in Waugh (2010), but also outperform the exporter specific trade cost model advocated by Waugh (2010) in fitting the R&D data.

The contributions of this study are severalfold. It contributes to the literature of aggregation in trade. Ramondo et al. (2016) show that an Eaton and Kortum (2002) (thereafter EK) structure on the state level can be successfully aggregated into country level under some innocuous assumptions. Similarly, Coughlin and Novy (2021) aggregate an Armington type gravity structure on a finer sub-region level into region level and examine the border effect associated with the cross country border trade flow between two regions relative to the trade flows between regions within country border. As their theory predicts, they find larger regions are systematically associated with smaller border effects. Inspired by both Ramondo et al. (2016) and Coughlin and Novy (2021), this paper defines a basic level and characterizes the conditions under which EK type gravity structure can

[1] There are many other notable studies analyzing intra-country trade cost with a wide variety of methods, *e.g.*, Allen and Arkolakis (2014), Ramondo et al. (2016), Kerem Coşar and Fajgelbaum (2016), Atkin and Donaldson (2017), Wrona (2018), Ma and Tang (2020), Coughlin and Novy (2021) and Allen and Arkolakis (2022).

be consistently aggregated to any level higher than the basic level, providing theoretical foundations for the validity of gravity researches at various levels. Different from aforementioned studies examining the consistency of gravity model under spatial aggregation, Redding and Weinstein (2019) investigate the consistency under sector aggregation.

In addition, this paper extends the vast strand of literature exploring how international borders adversely affect the international trade flow. Inspired by the seminal paper McCallum (1995), Anderson and Van Wincoop (2003) devise a theory motivated gravity estimation method qualifying and in part confirming McCallum (1995)'s results—international borders do hinder the trade flows. It is from then on that it becomes standard practice to include border dummy in the structural gravity estimation in researches exploiting the variations between trade flows across and within country borders.[2] Another strand of literature strives to identify the source of border effect. For example, Combes et al. (2005) find significantly lower magnitude of border effect after accounting for social and business networks, a result confirmed and further reinforced by Garmendia et al. (2012) showing that the border effect could vanish when using the value of trade flows and accounting for the social and business network at the same time. A particularly noteworthy piece is Coughlin and Novy (2021) where authors show both theoretically and empirically the estimated international border effect is region specific and heterogeneous across regions. In particular, larger size regions tend to have smaller size border effect in absolute value, indicating the level of aggregation in the region of interest could mechanically affect the border effect estimated, even though the underlying economic situation remains unchanged. Unlike the traditional gravity equation estimation method, Santamaria et al. (2020) use the causal inference framework to find significant causal effect of country borders on trade flows in Europe where small border effect is expected because countries in Europe are considered to be highly integrated with each other. Different from all the previous studies, this paper adapts the methodology used in Coughlin and Novy (2021) to country level trade flows and emphasizes the heterogeneity in the internal trade conditions across countries, using a country specific border dummy to capture the combined effects from the different level of aggregation due to difference in country size and heterogeneous internal trade impediments inside each country. Consistent with prediction of the model developed in this paper, the estimated border effect decreases in the country size. The intuition beneath this result is the following. As a country grows larger, trading within a country becomes harder on average and trading with partners across borders becomes relatively more appealing. Therefore, the estimated border effect specific to that country becomes smaller.

On another front, this paper offers new insights on literature attempting to fit the trade model to data and papers trying to structurally interpret the fixed effect in the

[2]Some notable studies include Chen (2004), Anderson and Yotov (2010), Arkolakis et al. (2018) and Wrona (2018).

ratio type gravity model. Assuming that the technology parameter is proportional to the labor within a country in an EK framework, Alvarez and Lucas (2007) firstly show that a parsimonious EK framework with intermediate input trading can match data moment quite well. One step further, Waugh (2010) argues that a traditional EK model with importer specific trade cost is unable to match the pattern of price data, whereas an EK model with exporter specific trade cost can resolve this conflict. Meanwhile, he shows that due to the existence of exporter specific trade cost the importer side fixed effect in a ratio type gravity model is a better measure for country's competitiveness. Based on Waugh (2010)'s exporter specific trade cost model, Simonovska and Waugh (2014) estimate the trade elasticity using simulated method of moment and they show that the estimated trade elasticity is insensitive to the trade cost interpretation. Though motivated differently, Ramondo et al. (2016) build an EK type model with internal trade cost and show theoretically that their model is observationally equivalent to Waugh (2010)'s model in generating equilibrium wage, trade flows and aggregate price level. Based on Waugh (2010)'s interpretation of fixed effect in ratio type gravity estimation, Levchenko and Zhang (2016) back out the technology parameters by industries in EK type model and find there exists convergence in technology across sectors within countries. Inspired by Coughlin and Novy (2021), this paper recovers an internal trade cost measure in an augmented gravity equation estimation which are consistent with the theory developed in this paper. Using this theory consistent internal trade cost, this paper quantitatively shows that an EK type model with internal trade cost can generate patterns matched with aggregate price data, just as Ramondo et al. (2016) predict theoretically. Further distinction between an EK type model with internal trade cost and a model with exporter specific trade cost endorsed by Waugh (2010) relies on their fit to R&D data. The correlation between R&D value generated under model with internal trade cost is 0.86 whereas the corresponding correlation under model with exporter specific trade cost is around 0.5. This comparison is important for the following two reasons. First, so far there is no satisfying credible data measure to approximate the technology parameter in the EK model. This comparison can help to testify whether R&D data is suitable to be one candidate. Second, those two models have distinctive implications on which side of fixed effects should be used as country's competitiveness measure to recover technology parameter and those two sides of fixed effects exhibit dissimilar patterns. Methods similar to the one used in Levchenko and Zhang (2016) are expected to yield different results using either side of fixed effects estimated in a ratio type gravity equation. Other than the studies using EK framework, heterogeneous firm models based on Melitz (2003) are also used to match data moments as well. Some notable examples include Arkolakis (2010) trying to explain excessive entry of small firms into a market; Eaton et al. (2011) trying to replicate the sorting pattern of firms entry into markets; Antràs et al. (2017) trying to match the sorting pattern in the firms' sourcing decisions.

The following section provides the stylized fact regarding size-adjusted internal trade to motivate the theoretical framework ensues. Section 3 lays out a theory characterizing the internal trade cost and provides a new welfare formula different from Arkolakis et al. (2012) highlighting the importance of the internal trade cost. Section 4 provides empirical tests to two major predictions made in theory section. Section 5 represents the empirical regularities regarding trade shares in Waugh (2010) and offers new empirical observation which supports the model with internal trade cost over other alternatives. Section 6 conducts quantification analysis and shows that the model with internal trade cost aligns very well with data in various dimensions and outperforms its competitor—the model with exporter specific trade cost in matching price index data and R&D data. The final section concludes.

2 Size-adjusted Internal Trade

Hummels and Klenow (2005) mention that almost all the theories of international trade predict larger regions export more in absolute terms than smaller regions and this stylized fact is generally testified to be true by the data. But one mystery remains—what causes this covariation? Does it come from the scale effect of one country's economic size or bigger countries are systematically associated with lower international trade cost? To this end, we utilize the structure of the popular structural gravity framework and document a new stylized fact that in a cross-sectional setting, conditional on the size of the region, the size-adjusted internal trade decreases with the size variables hinting that the internal trade cost is very likely to vary across countries, which contradicts to the convenient assumption that internal trade cost is homogeneous and can be normalized as one. This systematically higher internal trade cost for larger regions forces them to trade more with their international partners than themselves.

As Head and Mayer (2014) state, many types of structural gravity model converge into the following equation:

$$X_{ij} = C\frac{Y_i}{\Omega_i}\frac{Y_j}{\Phi_j}\phi_{ij} \quad \forall i, j, \tag{1}$$

where X_{ij} is the trade flow from region i to region j; C is a constant (very often could be the production of the whole world); Y_i is the expenditure of region i; Ω_i and Φ_j are corresponding multilateral resistance terms; ϕ_{ij} is bilateral trade cost.[3] Moving the country size terms to left, we arrive at the size-adjusted trade on the left, a notion

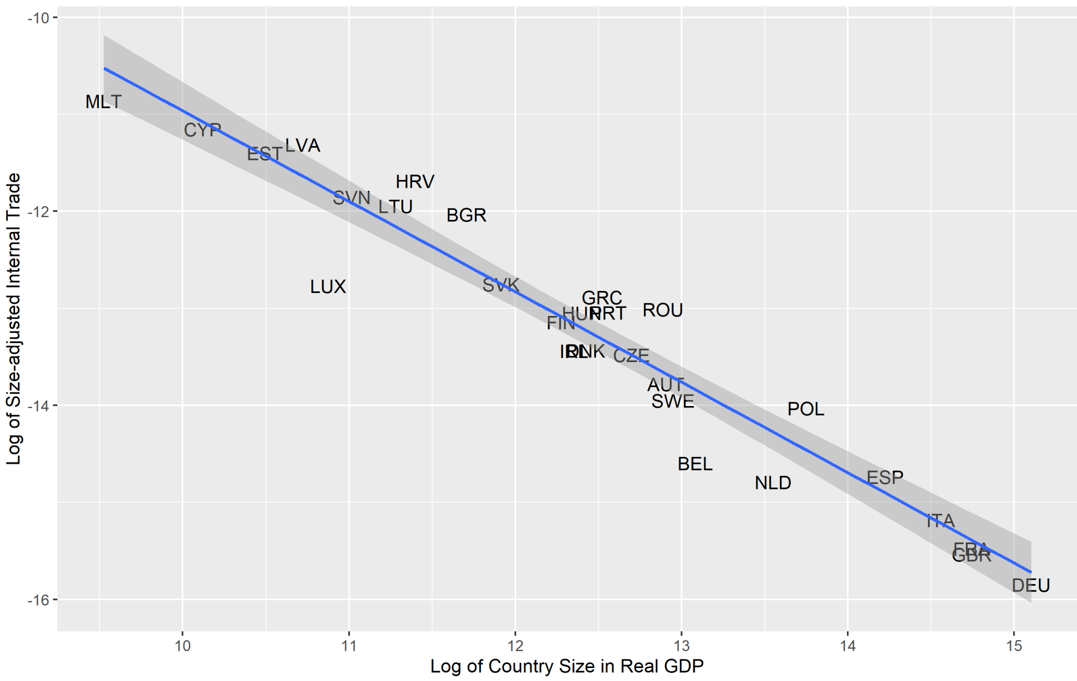
[3] Anderson and Van Wincoop (2003) arrive at the gravity structure from an Armington assumption; Chaney (2008) reaches the same equation using heterogeneous firm framework; the later part of this paper will show the same equation can be arrived from a Ricardian framework.

introduced by Anderson and Van Wincoop (2003):

$$\frac{X_{ij}}{Y_i Y_j} = C \frac{\phi_{ij}}{\Omega_i \Phi_j} \quad \forall i, j. \tag{2}$$

In a cross-sectional setting, the equation (2) means that conditional on the bilateral partners' sizes, all the variations in the size-adjusted trade should come from either bilateral resistance or multilateral resistance and this should hold even when $i = j$. Equation (2) provides a convenient channel through which a sneak peek into the variations of ϕ_{ii} becomes possible. To obviate the potential heterogeneity in the multilateral resistance terms, the highest integrated country block—the European Union is chosen as the sample so that the multilateral resistance terms can be roughly viewed as the same across countries. Data for both X_{ii} and Y_i are from ITPD-E database whose details are given in the section 4. To ensure consistency with roundabout production framework, Y_i takes the value of gross expenditure. The Figure 1 shows the relationship of the size-adjusted internal trade with respect to country size for EU countries.

Figure 1: Size-adjusted Internal Trade across Country Size

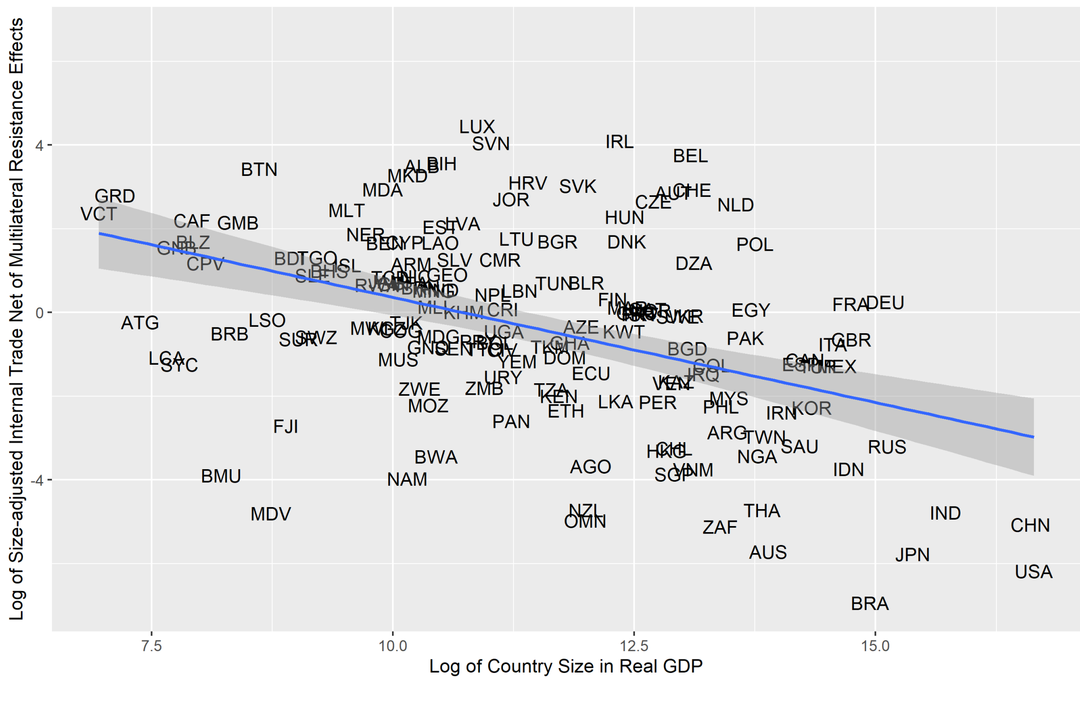


Note: The year for this figure is 2014 when the Great Britain was still part of EU.

Figure 1 shows a tight downward sloping relationship between size-adjusted internal trade and country size, meaning conditional on the size of the country, larger countries tend to trade less with themselves therefore more with their international partners. Assuming the multilateral resistance terms are roughly the same in EU, this stark downward relation implies larger countries tend to have larger internal trade cost. This relation largely persists when using other country size measures, except adding more noise, which is shown in appendix B.1.

To assess whether the same pattern persists worldwide regardless of the economic integration level, we compute the multilateral resistance terms following Donaldson and Hornbeck (2016) as $\Omega_i = \Phi_i = \sum_{j \neq i} \phi_{ij} Y_j$ where ϕ_{ij} is recovered from a standard trade level gravity model. These multilateral resistance terms summarize one country's access to international market. When moving Ω_i and Φ_i to the left of equation (2), the resulting statistics on the right is the size-adjusted internal trade net of multilateral resistance effects, which according to structural gravity theory reflects the internal trade cost. Figure 2 illustrates the cross country variations of the size-adjusted internal trade net of multilateral resistance effects for more than 100 countries worldwide. It shows that even after controlling the multilateral resistance terms there is clear evidence that, as Figure 1 suggests, the internal trade cost is heterogeneous across countries and increases with the size of the country. This downward sloping trend persists when population size is used as country size measure.

Figure 2: Size-adjusted Internal Trade Net of Multilateral Resistance across Country Size



Note: The year for this figure is 2014.

This stylized fact offers a new lens to why larger countries export more than smaller countries. Because of the higher internal trade costs, larger countries find relatively less convenient to trade with themselves, forcing them to trade more with international partners. The following section will show that this covariation between internal trade cost and country size will naturally arise when the gravity structure is imposed onto a finer geographical level.

3 Theoretical Framework

This section provides a theoretical framework that identifies the conditions under which gravity structure holds at the basic geographical units (will be defined in this section) can be consistently aggregated into any arbitrary levels above, giving credence to studies conveniently applying gravity structure to any aggregation level of interest. This framework also shows that the internal trade cost at country level should be heterogeneous across countries because of two reasons: first, it comes from heterogeneous trade cost between basic geographical units across countries; second, the different size of countries implies different number of basic geographical units each country contains, which leads to heterogeneous internal trade cost at country level through aggregation. The theoretical framework with internal trade cost naturally yields a welfare formula different from Arkolakis et al. (2012) (thereafter ACR)—the conventional gains from trade from change of absorption rate has to be adjusted by the change of internal trade cost.

3.1 Trade Between Basic Geographical Units

To set a starting point for aggregation, we define the basic geographical units satisfying the following properties.

Definition. A basic geographical unit is a swath of land segmented in such a way that:

1. any basic geographical units within a country are symmetric in the sense that each unit has the same amounts of labor l_i where i is the country index;
2. the technology parameter associated with any basic geographical unit in country i takes the parametric form $t_i = \phi_i l_i$, where ϕ_i is the country specific innovation intensity parameter.

The parametric form of technology parameter associated with basic geographical unit is inherited from Ramondo et al. (2016) where detailed discussion is provided. This parametric form allows aggregation conveniently across basic geographical units because of the property of Fréchet distribution—$\max\{x, y\}$ of two Fréchet random variables with shape parameter θ and scale parameter t_x and t_y is Fréchet with shape parameter θ and scale parameter $t_x + t_y$. Later in the quantification section, it will be shown that the model generated innovation intensity through this parametric form in fact fits the R&D data quite well.

The trade flow between two basic geographical units $m \in \Xi_i$ in country i and $n \in \Xi_j$ in country j satisfy the regular expression as in Eaton and Kortum (2002):

$$x_{i,m;j,n} = \frac{t_i c_i^{-\theta} d_{mn}^{-\theta}}{\sum_s t_s c_s^{-\theta} d_{sn}^{-\theta}} y_{j,n}, \tag{3}$$

where Ξ_i denotes the set of basic geographical units contained in country i; $x_{i,m;j,n}$ represents the trade flow from unit m in country i to unit n in country j; t_i is the technology parameter for unit m in country i; d_{mn} denotes the trade cost from m to n; $y_{j,n}$ is the total expenditure spent by labors in unit n. To avoid confusion at once keep compact notations, a notional convention is kept throughout this paper. The letter after the comma represents the basic geographical unit in the country whose index is given before the comma, *e.g.*, $y_{j,n}$. The letter after the comma will be omitted if that value is the same across all the units in a typical country, t_i.

Defining the price index at unit n in country j, $p_{j,n}^{-\theta} := \sum_{j\in\Omega}\sum_{s\in\Xi_j} t_s c_s^{-\theta} d_{sn}^{-\theta}$ where Ω is the set of countries in the world and using the fact that unit m's income comes from its sales $\sum_n x_{mn} = y_{i,m}$, we have the following:

$$y_{i,m} = t_i c_i^{-\theta} \sum_j \sum_n \frac{d_{mn}^{-\theta}}{p_{j,n}^{-\theta}} y_{j,n}. \tag{4}$$

To simplify notions, we define:

$$\pi_{i,m}^{-\theta} := \sum_j \sum_n \frac{d_{mn}^{-\theta}}{p_{j,n}^{-\theta}} y_{j,n}. \tag{5}$$

Plug equation (5) back to equation (4) to solve for $t_s c_s^{-\theta}$ which then is plugged back into equation (3) yield the following:

$$x_{i,m;j,n} = y_{i,m} y_{j,n} \frac{d_{mn}^{-\theta}}{p_{j,n}^{-\theta} \pi_{i,m}^{-\theta}},$$

where $p_{j,n}^{-\theta}$ is simplified as:

$$p_{j,n}^{-\theta} = \sum_i \sum_m \frac{d_{mn}^{-\theta}}{\pi_{i,m}^{-\theta}} y_{i,m}. \tag{6}$$

Following the convention in trade literature, the trade cost between any two units in the world is the same regardless of its direction, *i.e.*, $d_{mn} = d_{nm}$. Then one solution to the system of equations (5) and (6) is:

$$p_{i,m} = \pi_{i,m}.\text{[4]} \tag{7}$$

To allow aggregation into higher level, the following condition is needed.

Assumption. *$p_{i,m} = p_{i,n}$ for any $m, n \in \Xi_i$.*

[4] Strictly speaking, the necessary condition for the solutions of the equation system is $p_{i,m} = \lambda\pi_{i,m}$ for $\lambda > 0$. However, the inclusion of λ does not change the structure of the theoretical model and its interpretation, because it will be subsumed into the constant term. Therefore, we set $\lambda = 1$ to elude unnecessary complications.

This assumption says that the price index in each basic geographical unit is equalized within each country, which is not very restrictive given that the basic geographical units are defined symmetrically. This assumption can be satisfied when the trade cost between any two basic geographical units within a country is the same as in Ramondo et al. (2016) or a certain topology is imposed on the basic geographical units. For example, if the basic geographical units are arranged in tandem, those units can be thought as the symmetric arc of a circle as in Coughlin and Novy (2021), and the distribution center of goods moving across international border locates at the center of the circle. Or if the basic geographical units are placed on a plain and they are uniformly distributed over the surface of a sphere, meanwhile the distribution center of goods across international border locates at the center of the sphere. In the real world, this assumption is more likely to hold when the segmentation to define basic geographical unit is sufficiently small. For robustness, the appendix A.1 discusses how the main results in the theoretical section will be affected if the above assumption is relaxed.

3.2 Aggregation of Basic Units to Country Level

The trade flow within a country border is the summation of trade flows between any two basic geographical units within this country:

$$X_{ii} = \sum_{m \in \Xi_i} \sum_{n \in \Xi_i} x_{i,m;i,n} = \sum_{m \in \Xi_i} \sum_{n \in \Xi_i} y_i y_i \frac{d_{mn}^{-\theta}}{p_i^{-\theta} p_i^{-\theta}}, \tag{8}$$

where X_{ii} is the flow of goods country i trades with itself which is defined as:

$$X_{ii} \coloneqq Y_i Y_i \frac{D_{ii}^{-\theta}}{P_i^{-\theta} P_i^{-\theta}}, \tag{9}$$

where Y_i is the national income of country i and due to symmetry of basic geographical units $Y_i = N_i y_i$ where N_i is the number of basic geographical units country i contains. D_{ii} is the country level equivalent internal trade cost after aggregation. P_i is the country level price index which is conventionally defined as:

$$P_i^{-\theta} \coloneqq \sum_{j \in \Omega} T_j c_j^{-\theta} d_{ji}^{-\theta}, \tag{10}$$

where T_j is the technology parameter at country level which will be a composite of variables and parameters at basic geographical unit level to ensure aggregation consistency; d_{ji} is the country level trade cost. Then equation (8) can be transformed as:

$$Y_i Y_i \frac{D_{ii}^{-\theta}}{P_i^{-\theta} P_i^{-\theta}} = \sum_{m \in \Xi_i} y_i y_i \frac{d_{mm}^{-\theta}}{p_i^{-\theta} p_i^{-\theta}} + \sum_{n \in \Xi_i} \sum_{m \in \Xi_i, m \neq n} y_i y_i \frac{d_{mn}^{-\theta}}{p_i^{-\theta} p_i^{-\theta}}. \tag{11}$$

Using symmetry of basic geographical units and normalization $d_{mm} = 1$, it can be further simplified as:

$$\frac{D_{ii}^{-\theta}}{P_i^{-\theta} P_i^{-\theta}} = \frac{1}{N_i} \left(\frac{1}{p_i^{-\theta} p_i^{-\theta}} \right) + \left(\frac{1}{N_i} \right)^2 \sum_{n \in \Xi_i} \sum_{m \in \Xi_i, m \neq n} \frac{d_{mn}^{-\theta}}{p_i^{-\theta} p_i^{-\theta}}. \tag{12}$$

We define:

$$D_{ii}^{-\theta} := \frac{1}{N_i} + \frac{1}{N_i} \sum_{m \in \Xi_i, m \neq n} d_{mn}^{-\theta}, \tag{13}$$

then equation (12) becomes :

$$P_i^{-\theta} = p_i^{-\theta}, \tag{14}$$

meaning the country level price index is the same as the basic geographical unit level price index.

Under the mild assumption $\forall m \in \Xi_i, n \in \Xi_j$, $D_{ij}^{-\theta} = d_{mn}^{-\theta}$, namely trade cost between any two cross border basic geographical units are the same and equals to the country level trade cost,[5] and equation (13), it can be shown that the country level technology parameter has to satisfy:

$$T_i = \phi_i L_i, \tag{15}$$

to ensure the aggregated P_i is consistent with its definition, whose proof is delegated into appendix A.2. L_i is the total amount of labor which satisfies $L_i = N_i l_i$. The expression (15) shows that the gravity structure on the basic geographical unit level is consistently aggregated into the country level. Given the definition of t_i, this result can be intuitively interpreted as the best technology across all the basic geographical units within a country border will be used as production technology exclusively throughout this country.

More importantly equation (13) shows that the country level equivalent internal trade cost is a function of the number of the basic geographical units it contains and the innate trade cost between the basic geographical units in it, meaning the internal trade cost at country level is heterogeneous. This heterogeneity does not only come from the innate difference in the domestic trade conditions but also is affected by the size of the country.

To further illustrate how those two sources of heterogeneity affect country level equivalent internal trade cost, d_{mn} is assumed to be the same across basic geographical units within country and equals to value $\delta > 1$. Then equation (13) collapses to:

$$D_{ii}^{-\theta} := \frac{1}{N_i} + \frac{N_i - 1}{N_i} \delta^{-\theta}, \tag{16}$$

[5] This assumption is more likely to hold when the international trade cost is much lager than internationally traded goods' domestic distribution cost, such as for a small country. Or one could conveniently define D_{ij} equals to the average of d_{mn} between any basic geographical units pair across these two countries.

which is the same as the formula in Ramondo et al. (2016). This means the internal trade cost is a weighted average of one and $\delta^{-\theta}$. Keeping the number of basic geographical units the same, the higher the value of d meaning the higher the trade cost between two basic geographical units, the higher the country level internal trade cost will be. Suppose another topology which satisfies the equivalence of price index across basic geographical units is imposed within a country, which is closer to the real world. The symmetric basic geographical units are arranged in tandem on the circumference of a circle. The trade cost between any adjacent units is $\delta > 1$ and trade between any two units which are one unit apart has to travel through the adjacent unit meaning the trade cost will be δ^2. And the trade flow has to be clockwise on the circumference of the circle.[6] Then the equation (13) is transformed as:

$$D_{ii}^{-\theta} = \frac{1}{N_i} \left(\frac{1 - \left(\delta^{-\theta}\right)^{N_i}}{1 - \delta^{-\theta}} \right). \tag{17}$$

As $\left(\delta^{-\theta}\right)^{N_i}$ reduces to 0 relatively fast, D_{ii} generally is an increasing function in N_i, meaning the country level equivalent internal trade cost increases with the size of the country.[7] This feature of the model echos the key message of the stylized fact mentioned in the previous section and it will be one of the key predictions from the theory that empirical analysis strives to test in the later part of this paper.

3.3 Prediction on Ratio Type Gravity Estimation

The inclusion of heterogeneous internal trade cost does not modify much of the structure and interpretation in the trade level type gravity equation estimation. On the contrary, it significantly affects the interpretation of the country specific fixed effects in a ratio type gravity equation estimation which, unfortunately, is often used in studies need structural interpretation of country specific fixed effects.

Following the logic in the previous subsection, one can easily show the consistency of aggregation from basic geographical unit level to country level for international trade flows:

$$X_{ij} = \sum_{m \in \Xi_i} \sum_{n \in \Xi_j} x_{mn} = Y_i Y_j \frac{D_{ij}^{-\theta}}{P_i^{-\theta} P_j^{-\theta}}. \tag{18}$$

Together with the formula for internal trade flows, the traditional ratio type gravity

[6]The clockwise assumption and the circle positioning of basic geographical units is for simplification purpose. Alternative travel path like the shortest arc route and the surface positioning of a sphere will not qualitatively change the result.

[7]In fact, Ramondo et al. (2016) use distances between provinces to construct a proxy for internal trade cost at country level and show that this measure do increase with the size of the country. However, later the empirical part of this paper will show that distance alone is insufficient to be an appropriate measure of internal trade cost.

equation becomes:

$$\frac{X_{ij}}{X_{jj}} = \left(\frac{Y_i}{P_i^{-\theta}}\right)\left(\frac{P_j^{-\theta}}{Y_j D_{jj}^{-\theta}}\right) D_{ij}^{-\theta}. \tag{19}$$

The first parenthesis is exporter specific and the second parenthesis is importer specific. The third term will be approximated by traditional bilateral trade proxies. Therefore, when using fixed effect specification to estimate the above equation, the importer specific fixed effect effectively contains the internal trade cost term. And the exporter specific fixed effect divides the importer specific fixed effect for the same country will yield the country level internal trade cost estimate. Whereas, the traditional ratio type gravity equation specification implies that the cross country variation in the exporter specific fixed effects should be roughly the same as the cross country variation in the importer specific fixed effects and any difference from these two should be driven by totally random errors. This will be another key theoretical prediction that empirical analysis will attempt to test in the rest half of the paper.

3.4 General Equilibrium and Welfare Implication

To close the model in a tractable at the same time realistic way, this paper follows the setting in Waugh (2010) that the tradable sectors produce the intermediate inputs used in the homogeneous final goods production which are not traded. The trade share equation from country i to country j can be written as:

$$\Pi_{ij} = Y_i \frac{D_{ij}^{-\theta}}{P_i^{-\theta} P_j^{-\theta}} = \frac{T_i c_i^{-\theta} D_{ij}^{-\theta}}{\sum_{k\in\Omega} T_k c_k^{-\theta} D_{kj}^{-\theta}}. \tag{20}$$

The unit input bundle is assumed to be produced using a nested Cobb-Douglas technology such that its unit cost satisfies:

$$c_i = A\left(w_i^{\alpha_i} r_i^{1-\alpha_i}\right)^{\beta} p_i^{1-\beta}, \tag{21}$$

where A is a collection of constant parameters and gamma function. Similarly, the final goods production technology is assumed to be nested Cobb-Douglas technology and sold competitively:

$$p_F = B\left(w_i^{\alpha_i} r_i^{1-\alpha_i}\right)^{\gamma} p_i^{1-\gamma}, \tag{22}$$

where B is a collection of constant parameters and gamma function. Trade balance condition implies what one country expends equals to its sales:

$$\sum_j X_{ji} = Y_i = \sum_j X_{ij} = \sum_j \Pi_{ij} Y_j. \tag{23}$$

The Cobb-Douglas production technology of input bundle and final goods implies the total expenditure spent on intermediates goods equals its demands from intermediate goods production and final goods production:

$$Y_i = (1-\beta)\, Y_i + (1-\gamma)\, I_i, \tag{24}$$

where $I_i = w_i L_i + r_i K_i$. Meanwhile the labor market clears implies that total labor income equals the expenditure accrued to labor from tradable sectors and final sectors:

$$Y_i \beta \alpha_i + I_i \gamma \alpha_i = w_i L_i. \tag{25}$$

By Walras's Law the capital market clears as well. Equations (23), (24) and (25) together implies the followings: first, α_i share of total income comes from labor,

$$I_i = \frac{w_i L_i}{\alpha_i} = \frac{r_i K_i}{1-\alpha_i}. \tag{26}$$

Second, total income is proportional to intermediate input sales:

$$Y_i = \frac{1-\gamma}{\beta} I_i. \tag{27}$$

Third, labor income in country i is purely a weighted sum of all the other country's labor income:

$$w_i L_i = \sum_j \frac{\alpha_i}{\alpha_j} \Pi_{ij} w_j L_j. \tag{28}$$

The above equation can be used to pin down the relative wage level given parameters, trade share and labor size.

Though the inclusion of heterogeneous internal trade cost seems to be a small addition to the conventional trade models, it unexpectedly introduces additional components into the real income formula. After several rounds of substitution and rearranging, the real income formula can be shown as follows, whose proof is delegated into appendix A.3:

$$\frac{I_i/L_i}{p_F} = C \left(\frac{T_i}{\Pi_{ii}}\right)^{\frac{1-\gamma}{\beta\theta}} \left(\frac{1}{D_{ii}}\right)^{\frac{1-\gamma}{\beta}} \left(\frac{K_i}{L_i}\right)^{1-\alpha_i}, \tag{29}$$

where C is a constant composite consisting of A, B, and α_i. Clearly this model modifies ACR formula in two different ways—the inclusion of heterogeneous internal trade cost brings in the second term and inclusion of capital as another production factor create the third term. The impact of internal trade cost change on the real income per capita can be intuitively illustrated by the following example. Suppose an island country's infrastructure system is disrupted by a natural disaster, forcing them to rely more on the foreign sources of goods. According to the original ACR formula, reduction of absorption rate will necessarily increase the real income which is at odds with the fact that this country received a negative shock. Whereas in this heterogeneous internal trade cost model the traditional gains from trade would have to be adjusted by the internal trade condition and the real income will decrease in this situation which aligns with the intuition.[8] The third term says that higher capital level leads to higher real income, which is not surprising because the more capital one person owns, the more return he received from capital.

4 Empirical Tests

This section mainly conducts empirical tests against two predictions made in the theory section:

1. The internal trade cost measure is increasing in country size.
2. The ratio of exporter fixed effect over importer fixed effect is roughly the same size of the internal trade cost measure in ratio type gravity model.

Using a panel data of trade flows, this section introduces a country specific border effect to capture all the unobserved components of internal trade cost other than distance. Using this country specific border effect plus distance as the internal trade cost measure, this section confirms the theoretical predictions and shows that comparing to international trade cost, the dispersion of internal trade cost is not small enough to be homogenized.

4.1 Data Descriptions

Trade flow data — The trade flow data is from the ITPD-E Database.[9] Its coverage spans 17 years, from 2000 to 2016, including 243 countries and 170 industries. Relative to the other trade flow database, one advantage is that it contains internal trade flows whenever possible and its computation uses gross production rather than GDP as the reference sum which is more consistent with trade theory with roundabout production. The database's construction does not rely on an imputation using gravity method, therefore, the database is suitable for the gravity estimation. However, its merit can be its curse from the flip side—ITPD-E is not suitable for general equilibrium simulation, as most of the data is not regularized. As an alternative, we use trade flow data from GTAP database for general equilibrium simulation.

[8] One may argue that the traditional ACR formula already takes this example into account because T_i in the original ACR formula could be viewed a composite and the change in the interal trade cost could be deemed as a change in T_i, which is sensible enough. However, this kind of argument does not help in unraveling what constitutes the technology composite T_i in original ACR formula. In fact, later in the empirical sections, this paper will show after separating the effect from internal trade cost, this technology parameter T_i can be very well approximated by the R&D data.

[9] Its link is given as https://www.usitc.gov/data/gravity/itpde.htm.

Bilateral Trade Proxies — The proxies come from the CEPII website. One merit of the proxies constructed there is that it contains distance within country. Therefore, the distance component of the internal trade cost can be accounted for and the internal distance is also used to approximate the internal trade cost in several studies.[10]

Regional Trade Agreement — The regional trade agreement dummy comes from the Mario Larch's Regional Trade Agreements Database. Dummies separating the type of trade agreements, such as currency union, free trade agreement, economic integration agreement, are also included.

Country Characteristics — Country characteristics such as labor forces, capital stock, popoluation size, real GDP come from Version 8 Penn World Tables. In this new version, it distinguishes real GDP from expenditure side measuring standard of living and production side. Country's innovation intensity is measured by the number of personnel engaged in creation of knowledge per million people, which is obtainable from World Bank's World Development Indicators.

4.2 Internal Trade Cost Measure and Its Regularities

The structural gravity model predicts internal trade flows follows the same structure as international trade flows. However, internal trade flows are seldom used in gravity equation estimation. One reason could be credible data source of internal trade flow is not available until recently. Now, with the credible internal trade flow data from ITPD-E, we are able to conduct gravity estimation with internal trade flows. Table 1 shows the results.

The column (1) is the standard gravity model with common bilateral trade proxies, which confirms the validity of the data set we employed as all the estimates locate in the ranges specified in Head and Mayer (2014). However, when internal trade flows are mechanically pooled together with international trade flows the estimates start to bend astray—the coefficient on distance increases by 50%, the estimates on border continuity and RTA become insignificant. Nevertheless, as column (3) shows when a common border dummy which takes value 1 for international trade flow is added to the regression, all the estimates go back to its normal values as in column (1). This small exercise shows that there are unobserved components structurally exist in internal trade cost that cannot be captured by distance only. Just as Egger and Nigai (2015) convincingly show, failing to account for unobserved trade cost components could potentially bias all the estimates through trade balance conditions and this border effect on trade is almost 10 times higher than RTA and 6 times larger than common language, which should not be ignored. So it

[10]See Ramondo et al. (2016), Yotov (2012), and Heid et al. (2021).

Table 1: The Common Border Effect

	(1) Foreign	(2) Domestic_without	(3) Domestic_with
ln_dist	-1.169***	-1.503***	-1.117***
	(0.034)	(0.051)	(0.043)
lang_ethno	0.593***	0.427***	0.632***
	(0.070)	(0.075)	(0.072)
contig	0.462***	0.096	0.507***
	(0.132)	(0.137)	(0.131)
colony	0.691***	0.760***	0.686***
	(0.124)	(0.133)	(0.121)
rta	0.243***	-0.010	0.273***
	(0.052)	(0.058)	(0.054)
border			-3.604***
			(0.280)
_cons	14.299***	17.323***	17.444***
	(0.310)	(0.455)	(0.402)
N	101185	102450	102450
r2	0.771	0.768	0.777

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: This table provides the results of gravity estimates using OLS. The trade proxies are from the CEPII Database. Column (1) shows the standard fixed effects gravity results. Column (2) shows pooled results with both international flow and internal flow. Column (3) provides pooled results with both international flow and internal flow together with a border dummy added.

is reasonable to parameterize the trade cost function for the international trade flow as:

$$\log D_{ij,t}^{-\theta} = \alpha \log Dist_{ij} + Border \times \beta + BTP_{ij,t} \times \eta + \epsilon_{ij,t}, \tag{30}$$

where the β is common across country pairs. This means following conventional literature, we presume the border effect on international trade flows per se is homogeneous. On the other hand, the theory section shows that D_{jj} is directly associated with the number of the basic geographical units the country j contains, the trade cost function for internal trade flows are:

$$\log D_{jj,t}^{-\theta} = \alpha \log Dist_{jj} + \Lambda_j \times \gamma + BTP_{jj,t} \times \eta + \epsilon_{jj,t}, \tag{31}$$

where Λ_j is country specific and captures the heterogeneous internal trade cost components. Combining the above two equations together, we have the overall trade cost function:

$$\log D_{ij,t}^{-\theta} = \alpha \log Dist_{ij} + Border \times (\beta - \Lambda_j \times \gamma) + \Lambda_j \times \gamma + BTP_{ij,t} \times \eta + \epsilon_{ij,t} \quad \forall i, j. \tag{32}$$

Under the ceteris paribus assumption,

$$\frac{\Delta X_{ij,t}}{\Delta Border} = \beta - \Lambda_j \times \gamma, \tag{33}$$

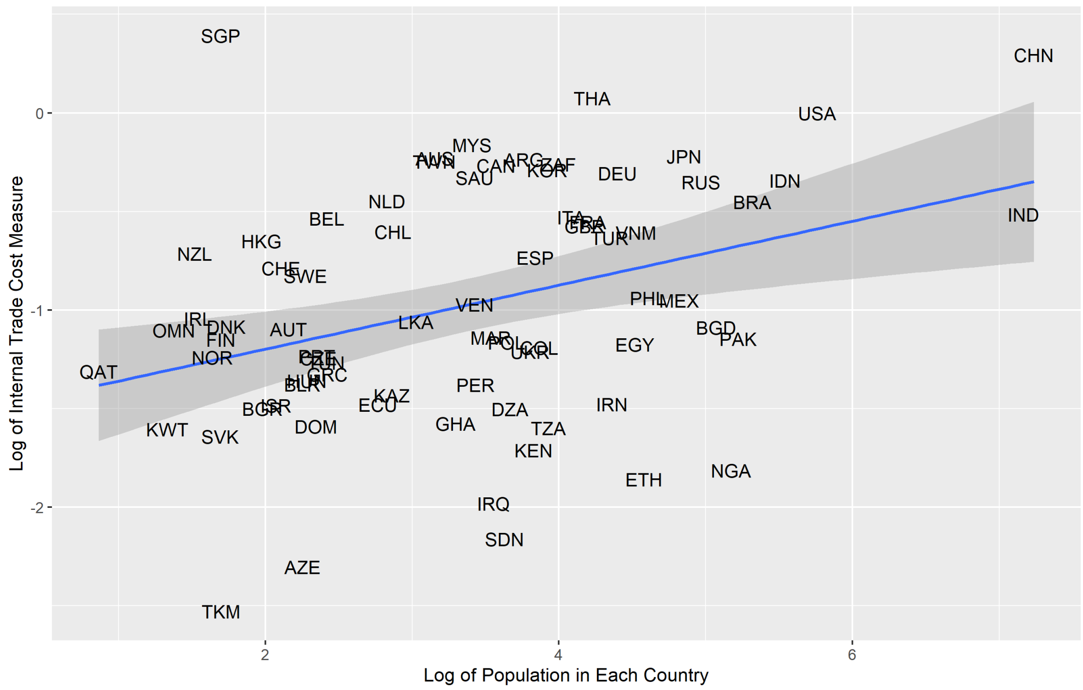
which shows that the true partial effect of borders on trade flows is in fact heterogeneous across countries. Though we presume a common border effect on international trade flows, because of the heterogeneous internal trade cost, the border effect of international trade flows relative to internal trade flows is heterogeneous across countries.

Therefore, to estimate country specific border effect, the following variant of trade cost function is assumed:

$$D_{ij,t}^{-\theta} = Dist_{ij}^{\alpha} \times \exp\left((1 - Border) \times \beta + Border \times \Lambda_j \times \gamma + BTP_{ij,t} \times \eta + \epsilon_{ij,t} \forall i, j\right), \tag{34}$$

where *border* takes value 1 when it is international trade flow. Plugging back this trade cost function specification into gravity equation in levels with internal trade flows, we conduct a gravity equation estimation with panel data spanning from year 2000 to year 2016. The internal trade cost measure is constructed combining the effect of country specific border dummies and distances. The following Figure 3 displays the relationship between the internal trade cost measure and country size.

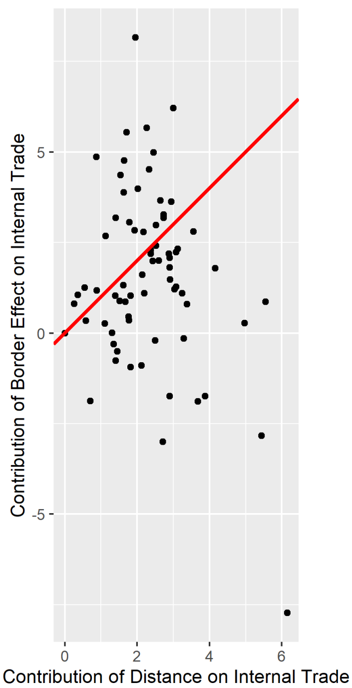
Figure 3: Internal Trade Cost Measure across Country Size



Note: The internal trade cost measure in the USA is normalized at 1. The year for population size is 2014.

The increasing trend of the fitted line between the internal trade cost measure and the size of the country is evident from above figure confirming the theory prediction in the previous section and this result is robust against various alternative country size measures which is shown in the appendix B.2. In fact, Ramondo et al. (2016) also show this relationship using the constructed internal trade cost measure from distance only. However, as Table 1 implies distance alone is not able to effectively capture all the trade cost components. The following Figure 4 further illustrates this idea by comparing the border effect estimate and estimate from distance.

Figure 4: The Impact on Internal Trade Flows from Border Effects versus Distance

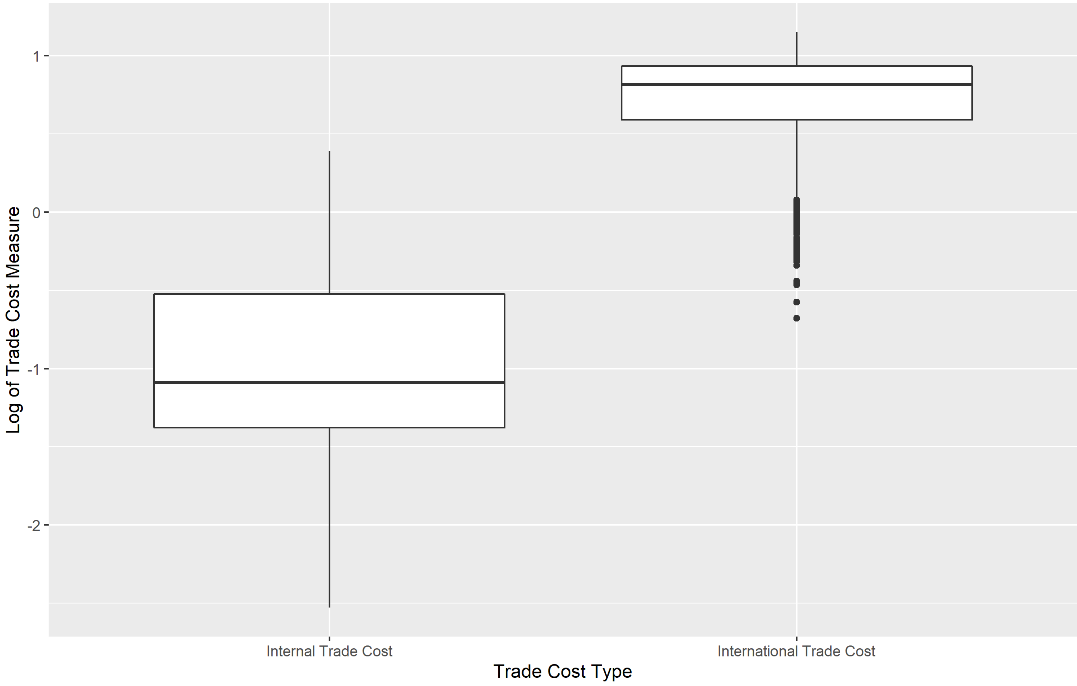


Note: The values on both x and y axis are normalized relatively to the ones in USA. This is appropriate because throughout this paper USA will always be the reference country and its internal trade cost measure always be normalized at 1, which effectively means the contribution from both border effects and distance for USA will be 0. The contribution of distance means this value is computed directly multiplying distance measure with its estimated coefficient in a log-linear regression without any transformation. And this value is a direct measure how much distance impact on internal trade flow. The contribution of border effect is similarly constructed.

The value of y axis spans from -7.7 to 8.2 whereas the value of x axis spans from 0 to 6.1 resulting a rectangular shape of this figure. Therefore, in terms of size, the impact from contribution of border effect is at least comparable to the impact from contribution of distance. And the red positive 45 degree line shows x value and y values do not have apparent correlation meaning there is no obvious pattern in contribution from border effect that can be represented by the contribution from distance. To summarize, using distance alone to represent internal trade cost is ineffective.

It is possible that internal trade cost measure exhibits large variations across countries whereas this variation is negligible comparing to international trade cost measure. To address this question, Figure 5 displays both distribution of two types of measure and their relative sizes, denying this possibility.

Figure 5: Internal Trade Cost Measure versus International Trade Cost Measure



Note: The internal trade cost measure in the USA is normalized at 1 and both types of trade cost measures are computed assuming $\theta = 4$.

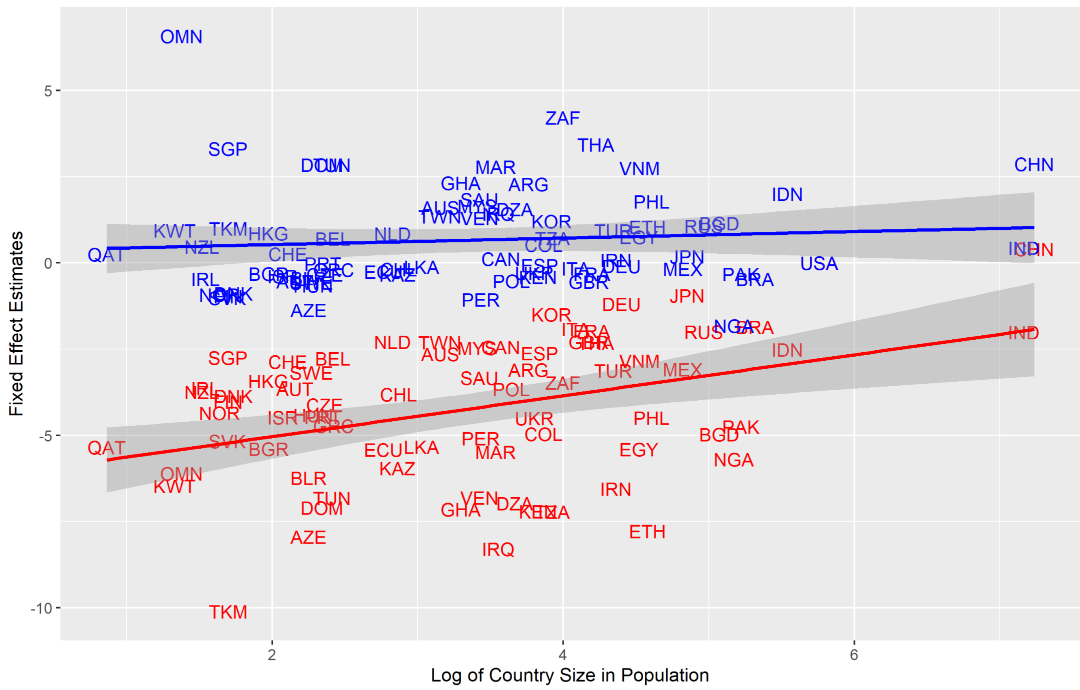
Figure 5 shows that in general international trade cost measure is higher than internal trade cost measure confirming the validity of trade cost measures. Also it is evident that internal trade cost measure has a bigger dispersion than international trade cost measure in log scale. In terms of size of trade cost, the top 25% of internal trade cost roughly overlap with the bottom 25% of international trade cost, dissipating the concern that internal trade cost can be safely viewed as homogeneous comparing to the international trade cost in size.

4.3 Ratio Type Gravity Estimation

As shown in the theory section, another striking difference between the standard gravity model and the model in this paper with heterogeneous internal trade cost locates in their prediction in ratio type gravity estimation. The standard gravity model predicts that both exporter side fixed effect and importer side fixed effect for the same country are measures of average competitiveness in production and should be roughly the same. However, in the estimation practice, it is rarely the case. The Figure 6 shows what typically happens in a cross-sectional ratio type gravity equation using trade data in year 2014.

The two types of fixed effects are by no means close to each other, even though

Figure 6: Exporter Fixed Effect versus Importer Fixed Effect in Ratio Gravity Equation



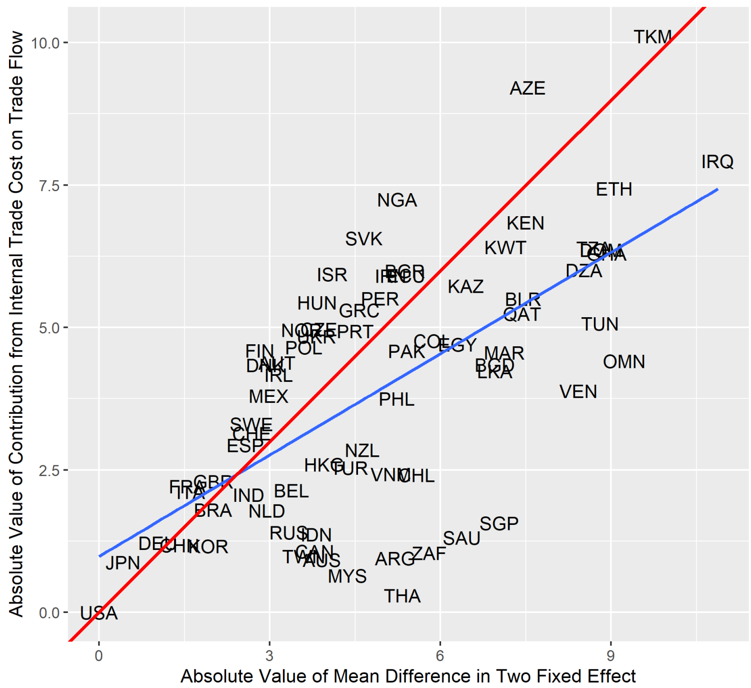
Note: Both exporter fixed effects and importer fixed effects are normalized with respect to USA. The red color represents values from exporter fixed effects. The blue color represents values from importer fixed effects.

the values for USA are set to be the same. More importantly, those two type fixed effects tell different stories about country's competitiveness—exporter fixed effects imply large countries tend to be more competitive than small countries whereas importer fixed effects imply similar competitiveness across countries. As competitiveness is a composite of country level technology parameter and unit cost of inputs, this disparity indicates R&D data should be used to distinguish these two.

Observing this type of disparity in these two types of fixed effects, Waugh (2010) introduces country specific trade cost and claims the difference between country's exporter specific fixed effect and importer specific fixed effect is the country specific trade cost with supporting evidence. His key message is that as the exporter fixed effect contains the effect of exporter specific trade cost, the importer fixed effect should be the correct measure of country's competitiveness and price index calculated using importer fixed effect demonstrates same pattern as the data, whereas price index generated using exporter fixed effect conflicts with data pattern. However, the quantification section of this paper shows that the model with internal trade cost also aligns well with price data and the final distinction has to be made based on R&D data.

The following Figure 7 shows that the correlation between the internal trade cost measure obtained through trade level gravity estimation against the difference between fixed effects using trade ratio gravity estimation.

Figure 7: Differences in Two Fixed Effects versus Internal Trade Cost



Note: The values on x axis are the absolute value of mean difference between two types of fixed effects of the same country across years in a panel ratio type gravity estimation. The values on y axis are the absolute value of contribution from internal trade cost on internal trade flow in a panel trade level gravity estimation. Both sets of values are computed relative to the USA level.

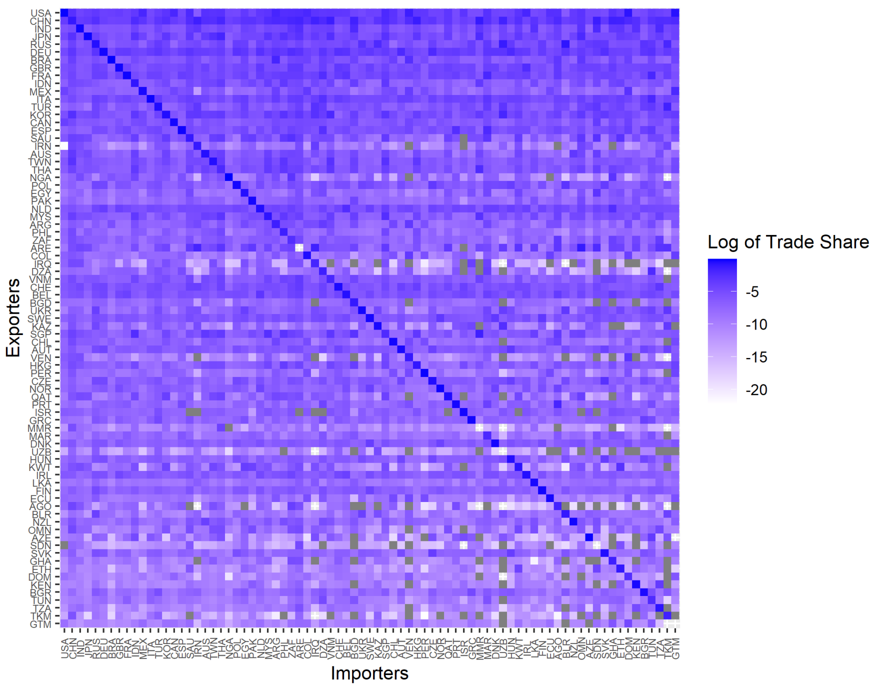
First thing to be noticed from Figure 7 is that those two sets of values are not statistically bound to be similar—the values on x axis are estimated using a panel ratio type gravity estimation and the values on y axis are estimated using a panel trade level gravity estimation. The identification of internal trade cost measure uses variations of internal trade flows across time and the identification of fixed effects use variations both across time and partners. The second salient feature is that those two sets of values are generally close to the 45 degree red line, though their best fitted line has a bit smaller slope. This is remarkable and generally a valid support to the second prediction made in the theory section, provided that the conventional trade theory states that there should not be clear pattern between them. The third noteworthy point is that, the top 10 GDP countries except Canada, actually align very closely to the red line around the left bottom corner of the figure. In fact, all the above features preserve when internal trade cost measure is compared with differences between fixed effects estimated in a cross sectional setting, which is delegated into appendix B.3.

5 Regularities Regarding Trade Shares

This section is a preparation for the quantification analysis. It reiterates several regularities found in Waugh (2010) and presents some new regularities regarding trade shares. Meanwhile this section constructs illustrative examples to show how the model with internal trade cost can potentially match with price data.

As a starting point for general equilibrium analysis, we represent the empirical regularities found in Waugh (2010). The following Figure 8 beautifully synthesizes the regularities about the trade shares.

Figure 8: Trade Shares across Countries



Note: The trade shares are the expenditure share a typical importer spends on each exporter's goods and they are calculated among the countries presented with trade flows in year 2014.

The counties are sorted in a descending order in country size from top to bottom and from left to right. The darker the color of a cell is, the higher the trade share for that country pair. Three patterns are evident. First, the diagonal line has the darkest color and roughly the same across countries. Second, for each importer, the upper half generally has denser color than the bottom half, meaning each country imports more from bigger country. Third, for each exporter, the left half generally has similar color density as the right half, meaning roughly all countries spend similar share on one country's goods. Together with the regularity found in tradable goods price index in Waugh (2010)—the price indices in tradable goods are roughly the same across countries, we can reformulate those regularities in terms of this paper's notation:

1. $\pi_{ii} = \pi_{jj} \quad \forall i, j;$

2. $\pi_{ij} > \pi_{kj}$ for any country i's size is bigger than country k;

3. $\pi_{ij} = \pi_{ik} \quad \forall j, k$;

4. $P_i = P_j \quad \forall i, j$.

The above four regularities have some implications on bilateral trade cost. To see this, we assume country i has bigger size than country j. The second and third regularities implies:

$$\pi_{ij} = \pi_{ik} > \pi_{jk} = \pi_{ji}. \tag{35}$$

Using the first regularity, the following expression holds:

$$\frac{\pi_{ij}}{\pi_{ii}} > \frac{\pi_{ji}}{\pi_{jj}}. \tag{36}$$

Because $\frac{\pi_{ij}}{\pi_{ii}} = \left(\frac{P_i}{P_j}\right)^{-\theta} \times \left(\frac{D_{ij}}{D_{ii}}\right)^{-\theta}$, the above expression can be transformed as:

$$\left(\frac{\pi_{ij}}{\pi_{ji}}\frac{\pi_{jj}}{\pi_{ii}}\right) = \left(\frac{P_i}{P_j}\right)^{-2\theta} \left(\frac{D_{ij}}{D_{ji}}\frac{D_{jj}}{D_{ii}}\right)^{-\theta} > 1. \tag{37}$$

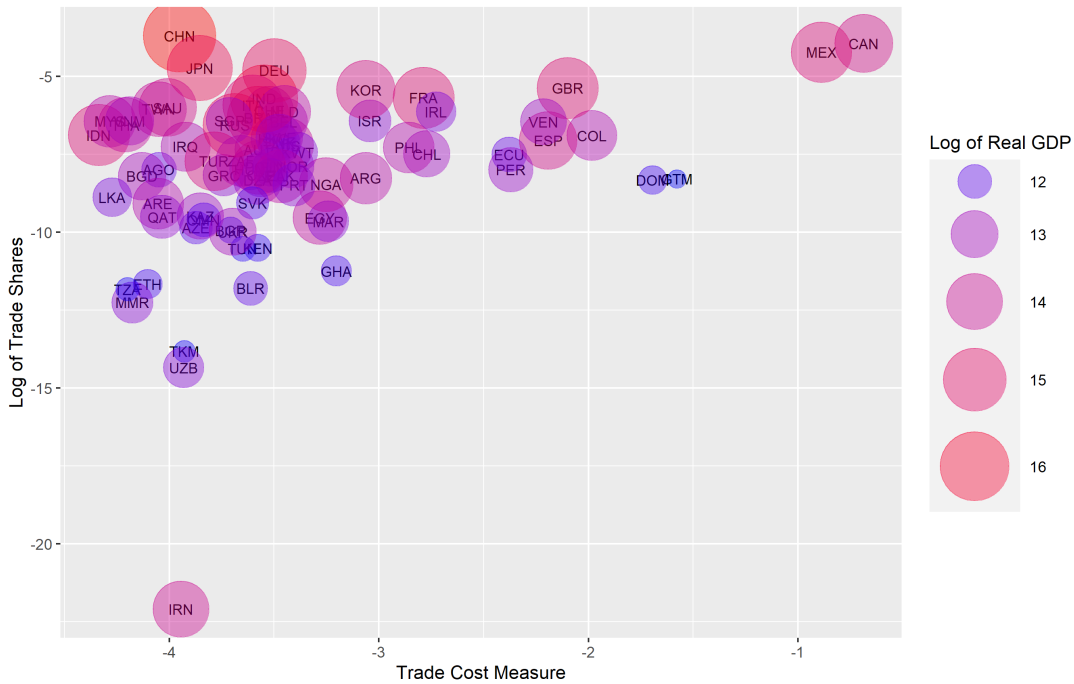
Because of the forth regularity, the trade share variations between country i and j should be traced back to the variations in either international trade cost or internal trade cost. In conventional trade models where internal trade cost is homogeneous, the variations in international trade cost becomes the only source of variations in trade share ratio. Therefore, Waugh (2010) points out that this finding hints that the variation in trade share ratio implies that international trade cost could possibly be asymmetric. However, within this paper's framework using the results established in previous sections—larger country has higher internal trade cost, namely $D_{ii} > D_{jj}$, even under the conventional assumption $D_{ij} = D_{ji}$, the variation in internal trade cost will imply variation in trade share ratio consistent with all the aforementioned empirical regularities. Therefore, only based on the aforementioned empirical regularities, the variation in internal trade cost has equal explanatory power as asymmetric international trade cost in explaining the variation in trade share ratio.

Additional regularities can be discovered, when zooming in and focusing on one particular country. The following Figure 9 shows the trade share dispersion across exporters for the US. It is apparent that USA imports more from countries with bigger size even those countries share similar trade cost when trading with the US. The similar pattern prevails for the most of the rest countries. The same graphs are shown in the appendix B.4 for China and Ethiopia.

This new regularity means:

$$\pi_{ik} > \pi_{jk} \quad when, \; D_{ik} = D_{jk}. \tag{38}$$

Figure 9: Trade Shares across Countries for USA as Importer



Note: The trade shares are the expenditure share USA spends on each exporter's goods and they are calculated among the countries displayed with trade flows in year 2014.

If internal trade cost is homogeneous across countries, the equalized international trade cost would imply the trade share ratio $\frac{\pi_{ij}}{\pi_{ji}}$ in equation (37) equals to 1 contradicting to the new empirical regularity found. However, a model with heterogeneous trade cost can naturally reconcile this contradiction, because larger country is shown to have higher internal trade cost.

Any discussions in this section before now relies on a crucial assumption that the model with internal trade cost advocated in this paper can replicate the forth empirical regularity—price index equivalence across countries, which will be shown in the quantification section. Here we provide an illustrate example to show why introduction of internal trade cost can achieve price index equivalence.

Suppose there are three countries indexed by 1 to 3. The country's size is decreasing in its index. We assume country 2 and country 3 are not trading with each other. Therefore the trade share matrix is given as follows:

$$\begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & 0 \\ \pi_{31} & 0 & \pi_{33} \end{bmatrix},$$

where π_{ij} is defined the same as in the theoretical section:

$$\pi_{ij} = \frac{T_i c_i^{-\theta} D_{ij}^{-\theta}}{P_j^{-\theta}}.$$

Then the above trade share matrix can be transformed as:

$$\begin{bmatrix} 1 & \frac{\pi_{12}}{\pi_{22}} & \frac{\pi_{13}}{\pi_{33}} \\ \frac{\pi_{21}}{\pi_{11}} & 1 & 0 \\ \frac{\pi_{31}}{\pi_{11}} & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{T_1 c_1^{-\theta} D_{12}^{-\theta}}{T_2 c_2^{-\theta} D_{22}^{-\theta}} & \frac{T_1 c_1^{-\theta} D_{13}^{-\theta}}{T_3 c_3^{-\theta} D_{33}^{-\theta}} \\ \frac{T_2 c_2^{-\theta} D_{21}^{-\theta}}{T_1 c_1^{-\theta} D_{11}^{-\theta}} & 1 & 0 \\ \frac{T_3 c_3^{-\theta} D_{31}^{-\theta}}{T_1 c_1^{-\theta} D_{11}^{-\theta}} & 0 & 1 \end{bmatrix}.$$

The regularities of trade share implies the following:

$$\frac{\pi_{21}}{\pi_{11}} > \frac{\pi_{31}}{\pi_{11}}, \frac{\pi_{12}}{\pi_{22}} = \frac{\pi_{13}}{\pi_{33}}.$$

In situation one, suppose there exists exporter side specific trade cost $D_{12} = D_{13} = d$ and internal trade costs are normalized to be 1 across countries, namely $D_{22} = D_{33} = 1$. This situation corresponds to the exporter specific trade cost model proposed in Waugh (2010). Because of $\frac{\pi_{12}}{\pi_{22}} = \frac{\pi_{13}}{\pi_{33}}$, we have the following:

$$T_2 c_2^{-\theta} = T_3 c_3^{-\theta}. \tag{39}$$

As for the price index in country 2,

$$P_2^{-\theta} = \left(T_1 c_1^{-\theta} d^{-\theta} + T_2 c_2^{-\theta}\right) = P_3^{-\theta} = \left(T_1 c_1^{-\theta} d^{-\theta} + T_3 c_3^{-\theta}\right).$$

This situation naturally yields the pattern that price index is invariant across countries of different size.

In situation two, suppose there exists importer side specific trade cost $D_{21} = D_{31} = d$ but with internal trade costs. First, consider the case where internal trade costs are normalized to be 1 across countries. Because of $\frac{\pi_{21}}{\pi_{11}} > \frac{\pi_{31}}{\pi_{11}}$, we have the following:

$$T_2 c_2^{-\theta} > T_3 c_3^{-\theta}, \tag{40}$$

meaning larger country is more competitive in production. Notice that equation (40) holds regardless of the assumption on internal trade cost. In the absence of heterogeneous internal trade cost, equation (40) necessarily implies

$$D_{12}^{-\theta} > D_{13}^{-\theta}, \tag{41}$$

which in turn indicates:

$$P_2^{-\theta} = \left(T_1 c_1^{-\theta} D_{12}^{-\theta} + T_2 c_2^{-\theta}\right) > P_3^{-\theta} = \left(T_1 c_1^{-\theta} D_{13}^{-\theta} + T_3 c_3^{-\theta}\right), \tag{42}$$

meaning the importer side specific trade cost in conventional trade model is inconsistent with price index data. However, this inconsistency can be easily resolved in a model with internal trade cost, because higher internal trade cost in larger country sabotages its higher competitiveness, which will not necessarily induce $D_{12} < D_{13}$. Then it becomes possible that price indices in country 2 and 3 are equalized. The different results shown in equation (39) and equation (40) suggests the further distinction between export specific trade cost model and internal trade cost model can reply on their fit to the R&D data.

6 Quantification Analysis

This section examines the fit of the model with internal trade cost in various dimensions and compares its performance with the model with exporter specific trade cost. Particularly, the model with internal trade cost achieves similar, if not better, fit to price index data as the model with exporter specific trade cost. And the internal trade cost model outperforms the model with exporter specific trade cost by a huge margin in fitting R&D data. However, they achieve similar fit to final goods price level and real income level. To facilitate understanding, we dub the model with internal trade cost model A and the model with exporter specific trade cost model B.

6.1 Calibration Procedure

The GTAP trade data is used for this quantification exercise. The parameters needed to recover include $\{\theta, \alpha_i, \beta, \gamma\}$. θ is the productivity dispersion parameter. Eaton and Kortum (2002) estimate it to be from 4 to 12 using wage data and price data. Additionally, Simonovska and Waugh (2014) show it to be around 4 to 8 using simulated method of moment and Ramondo et al. (2016) calibrate it to be 4 from growth literature. So here we prefer $\theta = 4$. As for α_i and β, in order to yield direct contrast to the results in Waugh (2010), we keep the labor share to be constant across countries and equal to $\alpha = 0.7$. The value added share in intermediate inputs β equals to 0.33. As the value added share in final goods should be somewhat similar to the one in intermediate goods, we follow Ma and Tang (2020) and pick 0.47 as the value added share for the final non-tradable goods production.

We follow the methodology of calibration used in Waugh (2010) in order to yield direct contrast to his model. The description of steps given here will be brief and the cookbook way of conducting it is delegated into appendix B.5. Matching the exact value of employment and trade share, we can recover wage rate up to a normalization. Using the capital labor ratio, rental rate can be backed out and throughout the quantification analysis, the estimated trade costs from gravity estimation are used. The most crucial difference from other calibration methods is that we use the value from the model implied fixed effects in a ratio type gravity equation as country's competitiveness measure to recover country specific technology parameter, which is in direct contrast with Ramondo et al. (2016). They match the technology parameter to the exact value of R&D data, which ex ante is a very risky move because, as mentioned before, the so called technology parameter is a composite containing not only the impact from technology level but also other factors which are not able to be explicitly explained by the structure of the model, although ex post this paper shows after netting out the internal trade cost, the technology parameter can be approximated fairly well using R&D data. Another key difference of this paper from Ramondo et al. (2016) is the formulation of trade costs. They approximate the trade costs only by distance which, as shown in previous section, will fail to capture the impact from aggregation. In fact, in an exercise in their paper, when they adjust internal trade cost measure to match trade share moment, their model fit to price index improves, meaning a precise trade cost measure is at the very core of the calibration and trade share is a critical moment a model should match to keep good fit to the data.

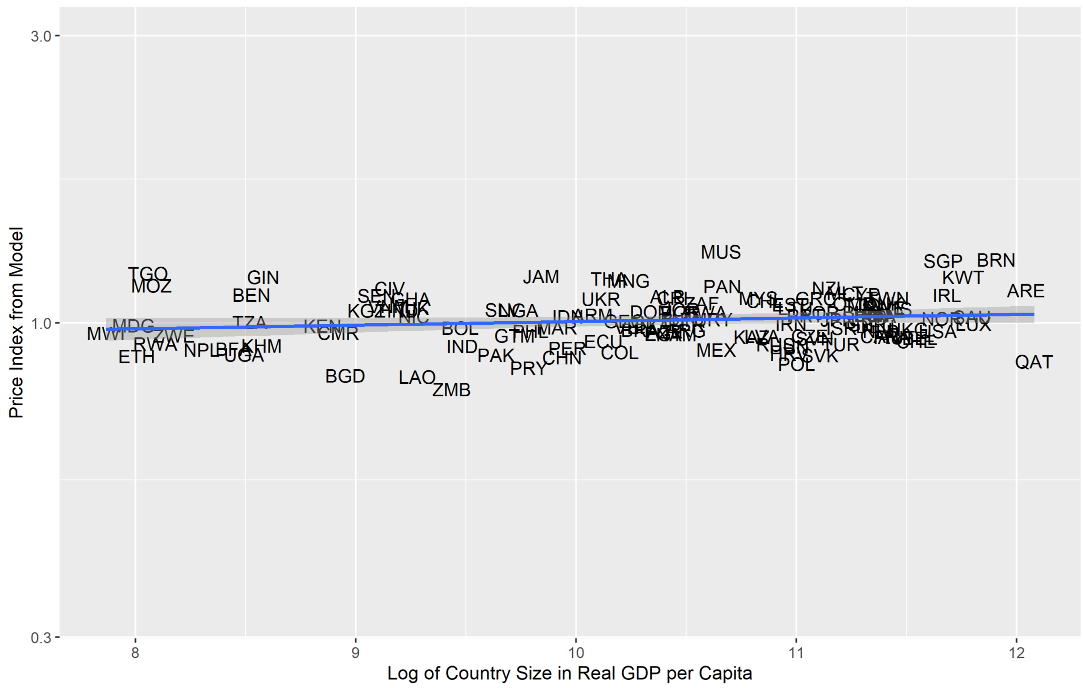
To summarize, our model matches the exact moments of trade share and factor of productions and uses backed out values of trade costs and country's competitiveness from gravity estimation to match target moments of price index, technology parameter and real income.

6.2 Tradable Price Indices

Waugh (2010) argues that the use of exporter side fixed effect estimated in a ratio type gravity equation as country's competitiveness is prone to be inconsistent with the price index data. However, the following Figure 10 shows that after introducing internal trade cost, this inconsistency disappears in model A and it achieves slightly better fit to the price index data than model B.

Although in the illustrative example in the previous section, we discuss the possibility of price index being the same across countries, because of its parsimonious setting, this equivalence is not guaranteed. However, the Figure 10 shows the clear pattern that price indices from model A are roughly same across countries which is consistent with the regularity discovered about price index data in Figure 1 of Waugh (2010). This better fit to data from model A is more pronounced when comparing to Figure 11 which shows the cross country differences of price index recovered from model B advocated by Waugh (2010). These price indices recovered from model B are roughly similar across countries but they have an apparent downward sloping trend. This mild downward sloping trend

Figure 10: Price Index Recovered from Model with Internal Trade Cost across Countries



Note: The price indices are normalized to the reference country USA and the trade cost measure and country's competitiveness measure are recovered from the gravity estimation.

of the fitted line also appears in Figure 4 in Waugh (2010) using 1996 year's data cross validating the validity of quantification method. Meanwhile the Figure 4 in Waugh (2010) shows the true data exhibits a slightly upward sloping trend which coincides with the slightly upward sloping trend of price indices in model A in Figure 10.

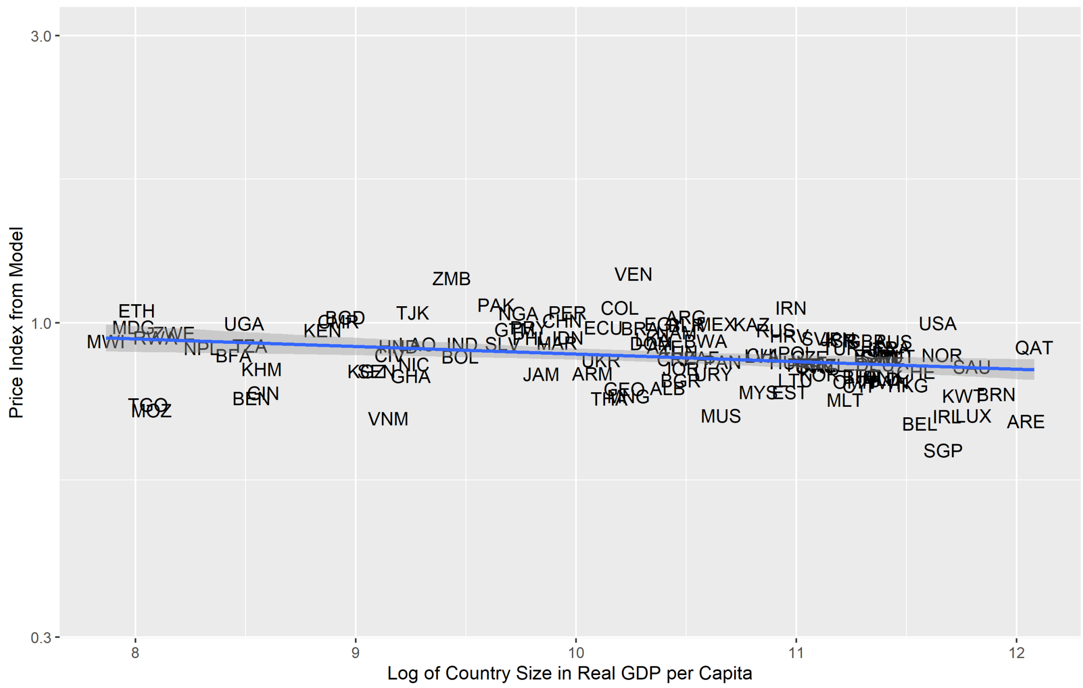
To summarize, in terms of fitting price index data, both model A and model B roughly replicates the equivalence of price indices across countries. However, only model A can generate a slightly upward sloping fitting trend which true data also display.

6.3 Innovation Intensity

According to the theory section $T_i = \phi_i L_i$, the technology parameter is assumed to be a linear function of total amount of labor and its coefficient ϕ_i is named innovation intensity. A natural question is how good it is to parameterize the technology parameter is such a parsimonious form? The answer to this question is demonstrated in Figure 12—the target moment the cross country variations of R&D counts of personnel per million people is well approximated by the innovation intensity parameter recovered from model A.

The correlation between data points and model A's predictions is as high as 0.857, meaning model A achieves generally good fit to data especially for top 20 GDP countries and most of Europe countries. Model A slightly over predicts the innovation intensity for extremely small country in size, *e.g.,* Luxembourg, Ireland, Belgium and Singapore.

Figure 11: Price Index from Model with Exporter Specific Trade Cost across Countries



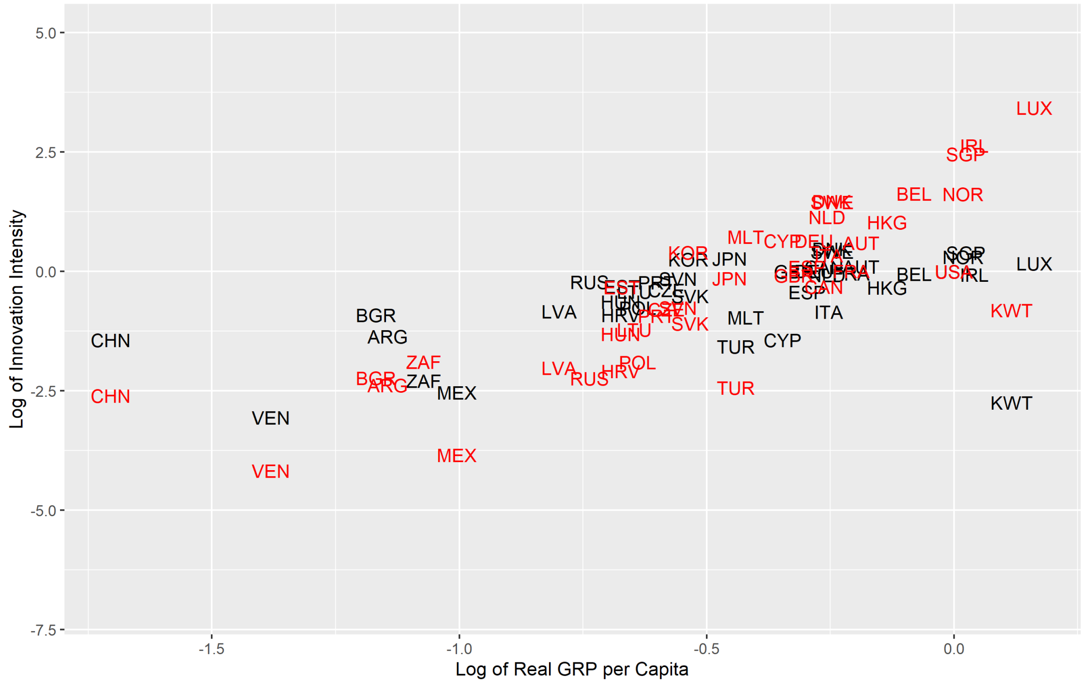
Note: The price indices are normalized to the reference country USA and the trade cost measure and country's competitiveness measure are recovered from the gravity estimation.

On the other hand the performance of model B in matching innovation intensity data is displayed in Figure 13. It is obvious that model B over predict for almost all countries except USA which serves as the reference country. The correlation between data points and its prediction is 0.59 much smaller than model A's score 0.857. However, it successfully replicates the upward trend exhibits in data but magnifies it by a huge margin. The significant difference in matching R&D data between two models has been anticipated in previous sections, because those two models are using different sets of estimates of country's competitiveness as shown in Figure 6. From the lens of model A, the importer fixed effect in a ratio type gravity estimation is contaminated with internal trade cost estimates. Completely negligence of internal trade cost in model B will erroneously attribute the effects from internal trade cost into innovation intensity and therefore distort the size of the innovation intensity prediction uniformly in model B. In a word, model A outperforms model B in matching R&D data.

6.4 Real Income and Final Price

It is instructive to keep in mind that both model A and B are matching exact value of factors of production and trade shares only. Therefore, their fit to real income data could be an unbiased criterion to access their performance. The following Figure 14 shows the fit of model A in matching real income data. It turns out almost all the countries are

Figure 12: Innovation Intensity from Model A versus Data



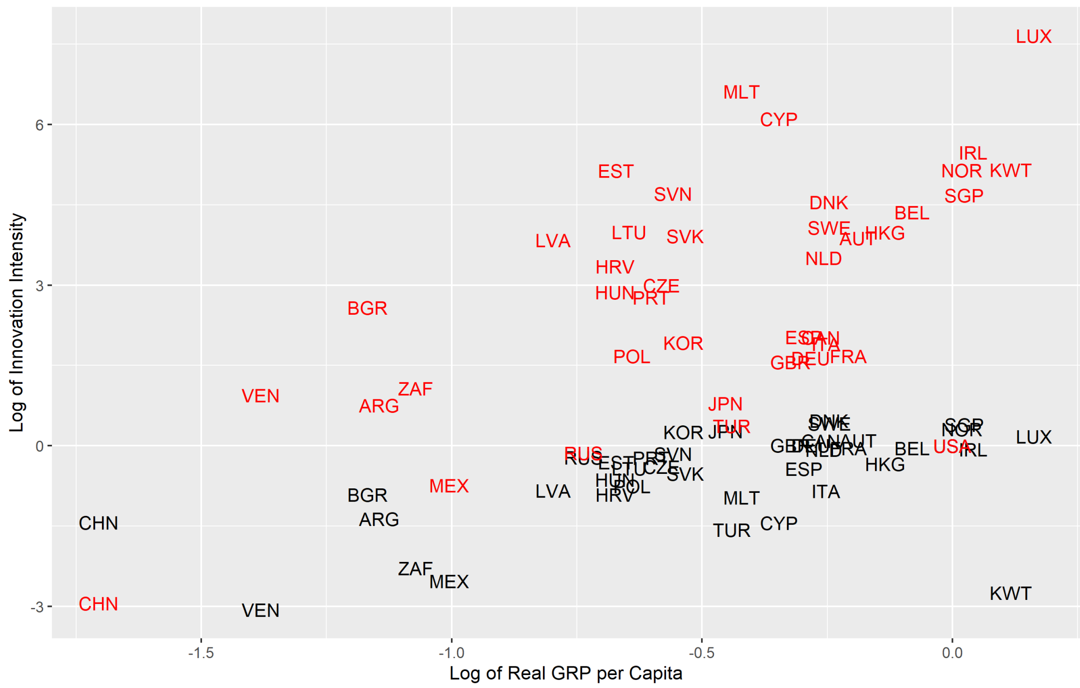
Note: The black font is for data values and the red font is for model counterparts. This figure only contains countries with records of innovation intensity longer than 15 years and each data point represents its average across years. Innovation intensity for both data and model prediction is normalized with respect to USA.

closely positioned around the 45 degree red line and the correlation between data and model prediction is 0.967. This high level fit of model A to data is surprising, given the parsimonious structure of the model with internal trade cost.

We also examine the fit of model A in aligning consumer price data which is given in Figure 15. Although not so close as the real income comparison, the points still locate around this 45 degree red line evenly and its best linear fit the blue line almost perfectly overlaps with the red line. The predictions from model A achieve 0.792 correlation with consumer price data. On the other hand, model B can achieve similar fit as model A in matching both real income and final price, whose graphs are delegated into appendix B.6. This is not surprising given the small difference between model A and B in generating price index as in Figure 10 and 11. Because both model A and B are matching exact values of factors of production and trade shares which will produce exactly same numbers for wage and rental rate in model A and B, the difference between price indices across models are washed out when compounded together with same wage and rental rate.

To summarize, the model with internal trade cost outperforms the model with exporter specific trade cost in matching price index data and R&D data and they are on par with each other when fitting real income and consumer price data. This distinction is important for two reasons. First, it establishes that after netting out the effect from the internal trade cost, the technology composite can be very well approximated by a

Figure 13: Innovation Intensity from Model B versus Data



Note: The black font is for data values and the red font is for model counterparts. This figure only contains countries with records of innovation intensity longer than 15 years and each data point represents its average across years. Innovation intensity for both data and model prediction is normalized with respect to USA.

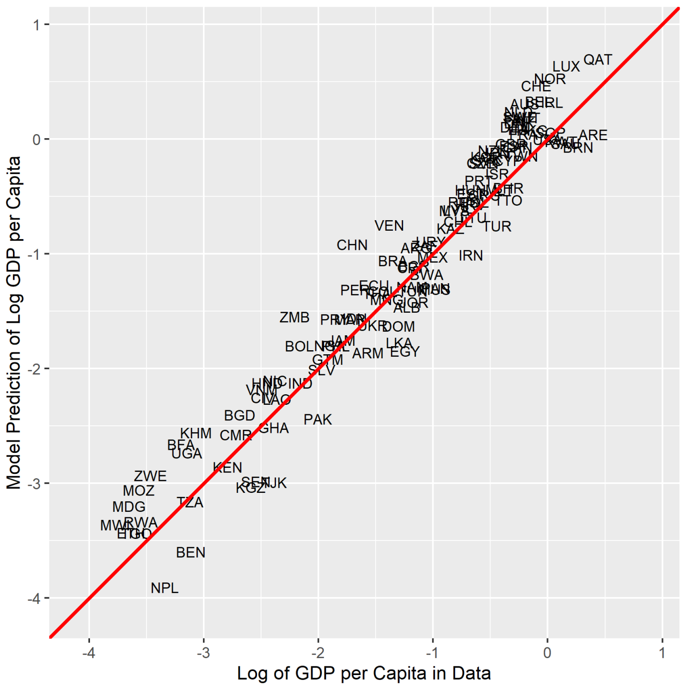
parsimonious function $T_i = \phi_i L_i$ and the value of ϕ_i is given by the R&D data. Second, as mentioned in previous section, the two models prefer different sets of fixed effect as country's competitiveness measure. Because those two fixed effects are completely different in their patterns across countries as shown in Figure 6, very different results could possibly emerge using different sets of fixed effect, such as using fixed effect estimates to back out country level productivity.

7 Conclusion

This paper casts doubt on the convenient but unwarranted assumption—internal trade costs across countries are homogeneous and documents a new stylized fact that size-adjusted internal trade decreases with country size. Motivated by this new stylized fact, this paper lays out a compelling theoretical framework of aggregation, identifying the condition which allows aggregation to any arbitrary level above basic geographical units and providing a theoretical foundation to researches applying gravity structure to any level of interests.

In addition, This theoretical framework makes two major predictions. First, country level equivalent internal trade cost should be heterogeneous and increasing in country size. Second, the difference between the two sets of fixed effect in ratio type gravity equation

Figure 14: Real Income Prediction from Model A versus Data



Note: The real GDP per capita is computed as real GDP divided by total employment in that country in year 2014. Both data and model predictions are normalized with respect to the USA.

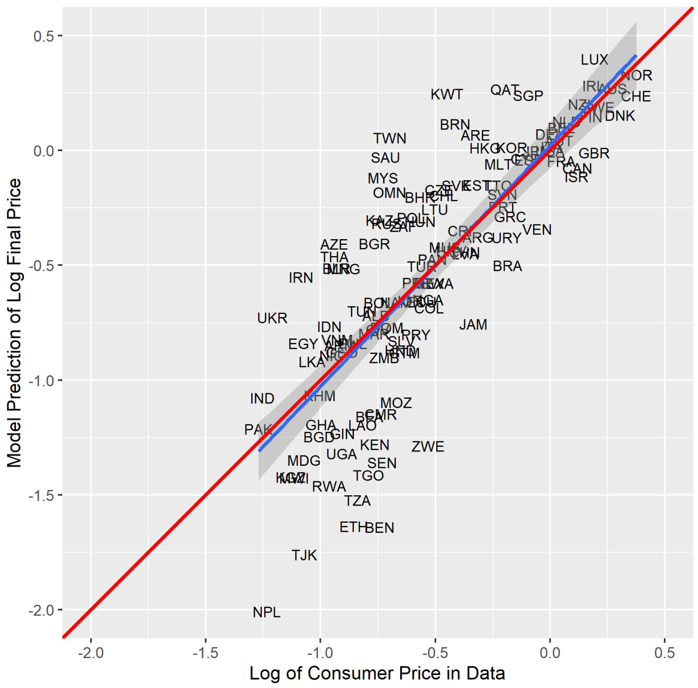
should equal to the internal trade cost. Both of predictions are tested to be true in the empirical analysis. Meanwhile, in the empirical section, it highlights the insufficiency of using distance only to capture internal trade cost and the unnegligible size of internal trade cost comparing to internal trade cost.

Then this paper provides new regularity regarding trade shares on top of the ones mentioned in Waugh (2010) and this new regularity can only be explained by a model with heterogeneous internal trade cost. To further distinguish the model with internal trade cost from alternatives, this paper conducts quantification analysis and shows the model with internal trade cost aligns well with data in various dimensions and outperforms its competitor—the model with exporter specific trade cost in matching price index data and R&D data. At the same time, this paper suggests exporter side fixed effect is more accurate a measure of country's competitiveness because importer side fixed effect is contaminated with internal trade cost effect in a ratio type gravity estimation.

On the other hand, this paper provides a novel explanation to the question that why small countries export less than big countries. Conditional on country size, in small countries, they trade more with themselves because of lower internal trade cost and big countries have to trade more with international partners because trade with themselves is costly.

A promising venue to extend this paper is to disentangle the components of the internal trade cost measure recovered in this paper and conduct counterfactual exercises

Figure 15: Consumer Price Prediction from Model A versus Data



Note: Both data and model predictions are normalized with respect to the USA. The data in year 2014 is used.

by twitching its components.

References

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, 129(3):1085–1139.

Allen, T. and Arkolakis, C. (2022). The Welfare Effects of Transportation Infrastructure Improvements. *Review of Economic Studies*.

Alvarez, F. and Lucas, R. E. (2007). General equilibrium analysis of the Eaton-Kortum model of international trade. *Journal of Monetary Economics*, 54(6):1726–1768.

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

Anderson, J. E. and Yotov, Y. V. (2010). The changing incidence of geography. *American Economic Review*, 100(5):2157–2186.

Antràs, P., Fort, T. C., and Tintelnot, F. (2017). The margins of global sourcing: Theory and evidence from US firms. *American Economic Review*, 107(9):2514–2564.

Arkolakis, C. (2010). Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy*, 118(6):1151–1199.

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., and Yeaple, S. R. (2018). Innovation and production in the global economy. *American Economic Review*, 108(8):2128–2173.

Atkin, D. and Donaldson, D. (2017). Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs. *SSRN Electronic Journal*.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–1721.

Chen, N. (2004). Intra-national versus international trade in the European Union: Why do national borders matter? *Journal of International Economics*, 63(1):93–118.

Clark, X., Dollar, D., and Micco, A. (2004). Port efficiency, maritime transport costs, and bilateral trade. *Journal of Development Economics*, 75(2 SPEC. ISS.):417–450.

Combes, P. P., Lafourcade, M., and Mayer, T. (2005). The trade-creating effects of business and social networks: Evidence from France. *Journal of International Economics*, 66(1):1–29.

Coughlin, C. C. and Novy, D. (2021). ESTIMATING BORDER EFFECTS: THE IMPACT OF SPATIAL AGGREGATION. *International Economic Review*.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics*, 131(2):799–858.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from French firms. *Econometrica*, 79(5):1453–1498.

Egger, P. H. and Nigai, S. (2015). Structural gravity with dummies only: Constrained ANOVA-type estimation of gravity models. *Journal of International Economics*, 97(1):86–99.

Garmendia, A., Llano, C., Minondo, A., and Requena, F. (2012). Networks and the disappearance of the intranational home bias. *Economics Letters*, 116(2):178–182.

Han, Z. and Li, H. (2022). Transportation Infrastructure and Trade.

Head, K. and Mayer, T. (2014). Gravity Equations: Workhorse,Toolkit, and Cookbook. In *Handbook of International Economics*, volume 4, pages 131–195. Elsevier B.V.

Heid, B., Larch, M., and Yotov, Y. V. (2021). Estimating the effects of non-discriminatory trade policies within structural gravity models. *Canadian Journal of Economics*, 54(1):376–409.

Hummels, D. and Klenow, P. J. (2005). The variety and quality of a nation's exports. *American Economic Review*, 95(3):704–723.

Kerem Coşar, A. and Fajgelbaum, P. D. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics*, 8(1):24–56.

Levchenko, A. A. and Zhang, J. (2016). The evolution of comparative advantage: Measurement and welfare implications. *Journal of Monetary Economics*, 78:96–111.

Ma, L. and Tang, Y. (2020). Geography, trade, and internal migration in China. *Journal of Urban Economics*, 115.

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

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Ramondo, N., Rodríguez-Clare, A., and Saborío-Rodríguez, M. (2016). Trade, domestic frictions, and scale effects. *American Economic Review*, 106(10):3159–3184.

Redding, S. J. and Weinstein, D. E. (2019). Aggregation and the Gravity Equation. *AEA Papers and Proceedings*, 109:450–455.

Santamaria, M. A., Ventura, J., and Yesilbayraktar, U. (2020). Borders Within Europe.

Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics*, 92(1):34–50.

Waugh, M. E. (2010). International trade and income differences. *American Economic Review*, 100(5):2093–2124.

Wrona, J. (2018). BORDER EFFECTS WITHOUT BORDERS: WHAT DIVIDES JAPAN'S INTERNAL TRADE? *International Economic Review*, 59(3):1209–1262.

Yotov, Y. V. (2012). A simple solution to the distance puzzle in international trade. *Economics Letters*, 117(3):794–798.

Appendices

A Theory Appendix

A.1 Heterogeneous Price Index at the Basic Geographical Unit Level

When the price indices at the basic geographical unit level are heterogeneous within a country, the following equation remains valid:

$$Y_i Y_i \frac{D_{ii}^{-\theta}}{P_i^{-\theta} P_i^{-\theta}} = \sum_{m \in \Xi_i} y_{i,m} y_{i,m} \frac{d_{mm}^{-\theta}}{p_{i,m}^{-\theta} p_{i,m}^{-\theta}} + \sum_{n \in \Xi_i} \sum_{m \in \Xi_i, m \neq n} y_{i,m} y_{i,n} \frac{d_{mn}^{-\theta}}{p_{i,m}^{-\theta} p_{i,n}^{-\theta}},$$

where the income $y_{i,m}$ and the price index $p_{i,m}$ at the basic graphical unit level could be heterogeneous. Moving the country level income and price indices to the right, the above equation could be transformed into:

$$D_{ii}^{-\theta} = P_i^{-2\theta} \left(\sum_{m \in \Xi_i} s_{i,m}^2 \frac{d_{mm}^{-\theta}}{p_{i,m}^{-\theta} p_{i,m}^{-\theta}} + \sum_{n \in \Xi_i} \sum_{m \in \Xi_i, m \neq n} s_{i,m} s_{i,n} \frac{d_{mn}^{-\theta}}{p_{i,m}^{-\theta} p_{i,n}^{-\theta}} \right), \tag{43}$$

where $s_{i,m} := y_{i,m}/Y_i$ the share of income of one basic geographical unit in country i. Under *ceteris paribus* assumption—meaning the distributions of $p_{i,m}$ and $d_{n,m}$ are the same across countries and the observation that P_i is roughly the same across countries according to Waugh (2010), the summation of the first term of within the parenthesis increases with the size of the country at a rate of N_i whereas its weight $s_{i,m}^2$ increases with the size of the county at a rate of $1/N_i^2$. Therefore, in the very general situation, D_{ii} the country level internal trade cost measure will be a function of its size N_i and varies across countries when the price indices at the basic geographical unit level are heterogeneous.

To acquire an intuitive sense of how the country level internal trade cost measure D_{ii} could vary with country N_i, under the definition of the basic geographical units that they endowed with same amount of labor, each basic geographical unit m has the same income share, meaning $s_{i,m} = 1/N_i$, then the above equation gives:

$$D_{ii}^{-\theta} = P_i^{-2\theta} \left(\frac{1}{N_i^2} \sum_{m \in \Xi_i} \frac{d_{mm}^{-\theta}}{p_{i,m}^{-\theta} p_{i,m}^{-\theta}} + \frac{1}{N_i^2} \sum_{n \in \Xi_i} \sum_{m \in \Xi_i, m \neq n} \frac{d_{mn}^{-\theta}}{p_{i,m}^{-\theta} p_{i,n}^{-\theta}} \right), \tag{44}$$

where both terms in the parenthesis contain N_i, meaning generally D_{ii} should be a function of N_i.

However, under the general situation where price indices at the basic geographical unit level are heterogeneous, it is difficult to have a stand on what direction D_{ii} would change when N_i changes because d_{mn} could possibly be a function of N_i as a result of the economy of scale in shipping or port service mentioned in Clark et al. (2004) and modeled in Han and Li (2022). Despite the uncertainty in the direction of change, the other main conclusions in the main text only rely on the result that country level internal trade cost is heterogeneous across countries therefore they remain valid under this general situation.

A.2 Aggregation of Country Level Price Index

From the main text, the country level internal trade cost is defined as:

$$D_{ii}^{-\theta} := \frac{1}{N_i} + \frac{1}{N_i} \sum_{m \in \Xi_i, m \neq n} d_{mn}^{-\theta},$$

and the country level international trade cost satisfy:

$$D_{ij}^{-\theta} = d_{mn}^{-\theta}, \quad \forall m \in \Xi_i, n \in \Xi_j.$$

Under the above two conditions, the basic geographical unit level trade cost can be transformed as:

$$
\begin{aligned}
p_i^{-\theta} &= \sum_{i\in\Omega}\sum_{m\in\Xi_i} t_m c_m^{-\theta} d_{mn}^{-\theta} \\
&= \sum_{m\in\Xi_i} t_m c_m^{-\theta} d_{mn}^{-\theta} + \sum_{j\neq i, j\in\Omega}\sum_{m\in\Xi_j} t_m c_m^{-\theta} d_{mn}^{-\theta} \\
&= \sum_{m\in\Xi_i} t_m c_m^{-\theta} d_{mn}^{-\theta} + \sum_{j\neq i, j\in\Omega} t_m c_m^{-\theta} \sum_{m\in\Xi_j} D_{ji}^{-\theta} \\
&= \sum_{m\in\Xi_i} t_i c_i^{-\theta} d_{mn}^{-\theta} + \sum_{j\neq i, j\in\Omega} N_j t_j c_j^{-\theta} D_{ji}^{-\theta} \\
&= t_i c_i^{-\theta}\left(1 + \sum_{m\in\Xi_i, m\neq n} d_{mn}^{-\theta}\right) + \sum_{j\neq i, j\in\Omega} N_j t_j c_j^{-\theta} D_{ji}^{-\theta} \\
&= N_i t_i c_i^{-\theta} D_{ii}^{-\theta} + \sum_{j\neq i, j\in\Omega} N_j t_j c_j^{-\theta} D_{ji}^{-\theta} \\
&= N_i l_i \phi_i c_i^{-\theta} D_{ii}^{-\theta} + \sum_{j\neq i, j\in\Omega} N_j l_j \phi_j c_j^{-\theta} D_{ji}^{-\theta} \\
&= L_i \phi_i c_i^{-\theta} D_{ii}^{-\theta} + \sum_{j\neq i, j\in\Omega} L_j \phi_j c_j^{-\theta} D_{ji}^{-\theta} \\
&= \sum_{j\in\Omega} L_j \phi_j c_j^{-\theta} D_{ji}^{-\theta}.
\end{aligned}
$$

The last step of the equation uses the fact that one country's total labor is the sum of the labor of each basic units $L_j = N_j l_j$. Recalling that $P_i^{-\theta} \coloneqq \sum_{j\in\Omega} T_j c_j^{-\theta} d_{ji}^{-\theta}$, to ensure $P_i^{-\theta} = p_i^{-\theta}$, T_j has to be defined as

$$
T_j = \phi_j L_j.
$$

A.3 Real Income Formula

To facilitate derivation, we define $VA \coloneqq w_i^{\alpha_i} r_i^{1-\alpha_i}$, then from final goods price equation we have:

$$
\begin{aligned}
p_F &= B\left(w_i^{\alpha_i} r_i^{1-\alpha_i}\right)^{\gamma} p_i^{1-\gamma} \\
p_F &= B\left(VA\right)^{\gamma} p_i^{1-\gamma} \\
\frac{VA}{p_F} &= \left(\frac{VA}{p_i}\right)^{1-\gamma} \frac{1}{B}.
\end{aligned}
\tag{45}
$$

From the intermediate input unit cost formula, we know that

$$c_i = A\,(VA)^{\beta}\,p_i^{1-\beta}$$
$$\frac{VA}{p_i} = \left(\frac{c_i}{p_i}\right)^{1/\beta}\frac{1}{A^{1/\beta}}. \tag{46}$$

From the trade share equation, we know that

$$\Pi_{ii} = \frac{T_i c_i^{-\theta} D_{ii}^{-\theta}}{\sum_{k\in\Omega} T_k c_k^{-\theta} D_{kj}^{-\theta}} = \frac{T_i c_i^{-\theta} D_{ii}^{-\theta}}{p_i^{-\theta}}$$
$$\frac{c_i}{p_i} = \left(\frac{T_i}{\Pi_{ii}}\right)^{1/\theta}\frac{1}{D_{ii}}. \tag{47}$$

Because of $\frac{w_i L_i}{\alpha_i} = \frac{r_i K_i}{1-\alpha_i}$, the value added expression can be transformed into:

$$VA = w_i^{\alpha_i} r_i^{1-\alpha_i}$$
$$VA = \left(\frac{1-\alpha_i}{\alpha_i}\right)^{1-\alpha_i}\left(\frac{L_i}{K_i}\right)^{1-\alpha_i} w_i$$
$$VA = \left(\frac{1-\alpha_i}{\alpha_i}\right)^{1-\alpha_i}\left(\frac{L_i}{K_i}\right)^{1-\alpha_i} \alpha_i \frac{I_i}{L_i}. \tag{48}$$

After substituting equation (47) into equation (46) and equation (46) into equation (45), replace the VA in equation (45) with the formula in equation (48), we can arrive at the following real income per capita formula:

$$\frac{I_i/L_i}{p_F} = C\left(\frac{T_i}{\Pi_{ii}}\right)^{\frac{1-\gamma}{\beta\theta}}\left(\frac{1}{D_{ii}}\right)^{\frac{1-\gamma}{\beta}}\left(\frac{K_i}{L_i}\right)^{1-\alpha_i}.$$

B Empirics Appendix

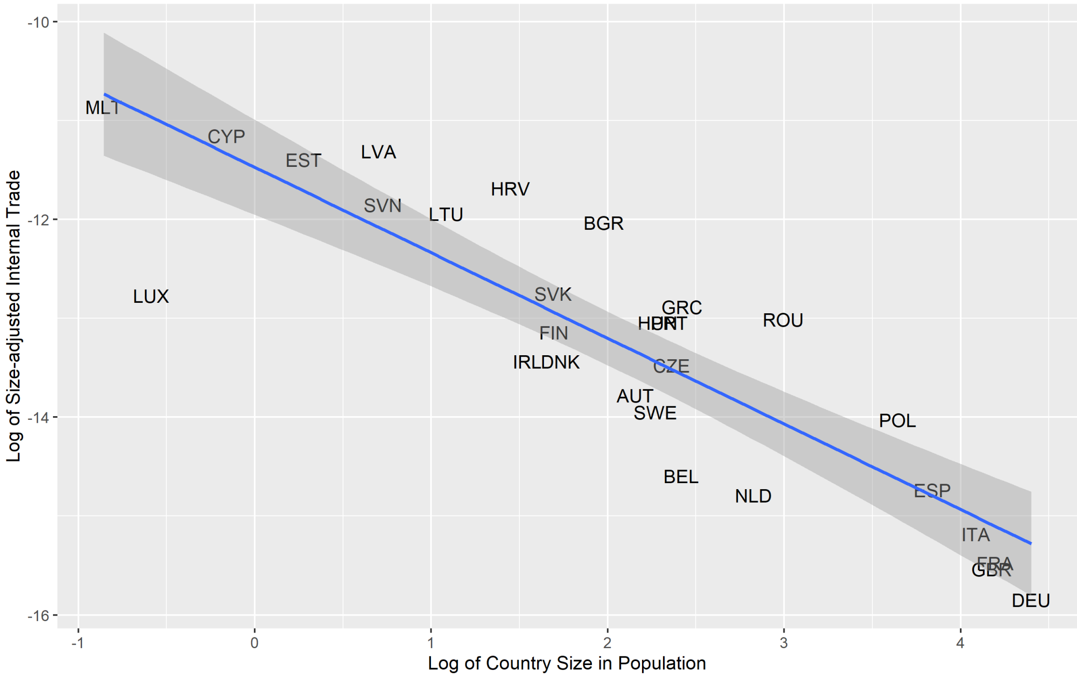
B.1 Robustness of Size-adjusted Trade against Country Size

The Figure 16 and Figure 17 show the stylized fact that size-adjusted internal trade decreases with country size is robust with respect to various country size measures for EU countries. The only difference is that those figures have more noise than the one in the main text.

B.2 Robustness of Internal Trade Cost Measure against Country Size

The Figure 18 shows the internal trade cost measure against real GDP. The Figure 19 shows the internal trade cost measure against real GDP per capita.

Figure 16: Size-adjusted Internal Trade across Country Population



Note: The year for this figure is 2014 when the Great Britain was still part of EU.

From Figures 3, 18, 19, although there might be difference in how tight the increasing relationship between internal trade cost measure and various country size measures, it is generally true internal trade cost measure increase in country size.

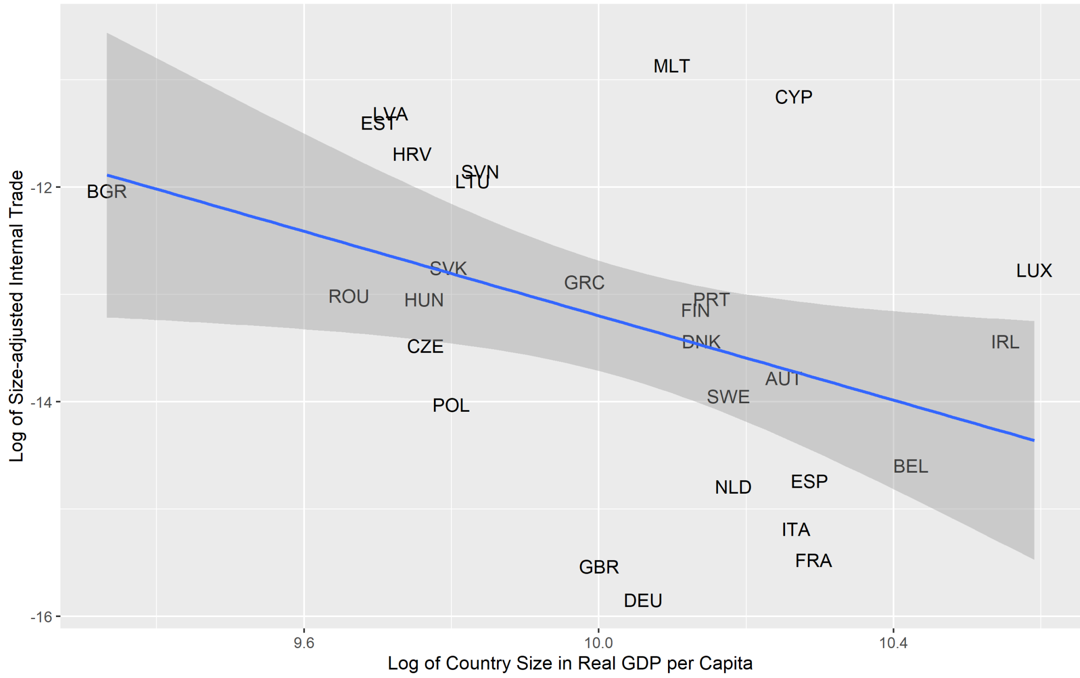
B.3 Comparison of Internal Trade Cost and Difference between Fixed Effects

The Figure 20 displays the relationship between contribution of internal trade cost measure on trade flows vis-à-vis difference between fixed effects estimated in a cross-sectional setting. The identification of fixed effect uses variations across partners and not across time. And this tight relationship especially for the top 10 GDP countries between the two sets of values validates the prediction in the theory section on the ratio type gravity equation. The main features observed in Figure 7 are preserved in Figure 20 except it's more noisy here because of the year specific shocks.

B.4 Robustness Check for Trade Share Regularities

Here we present the same graph as Figure 9 for country China and Ethiopia. The similar pattern emphasized in he main text persists for country China and Ethiopia.

Figure 17: Size-adjusted Internal Trade across Real GDP per Capita



Note: The year for this figure is 2014 when the Great Britain was still part of EU.

B.5 Cookbook Steps of Calibration

1. From ratio gravity estimation,

$$\frac{X_{ij}}{X_{jj}} = \frac{T_i c_i^{-\theta}}{T_j c_j^{-\theta} D_{jj}^{-\theta}} D_{ij}^{-\theta},$$

   obtain the exporter fixed effect $\hat{S}_i$ as estimates of $\log(T_i c_i^{-\theta})$ and $\hat{\tau}_{ij}$ and $\hat{\tau}_{jj}$as estimates of $\log(D_{ij}^{-\theta})$ and $\log(D_{jj}^{-\theta})$.

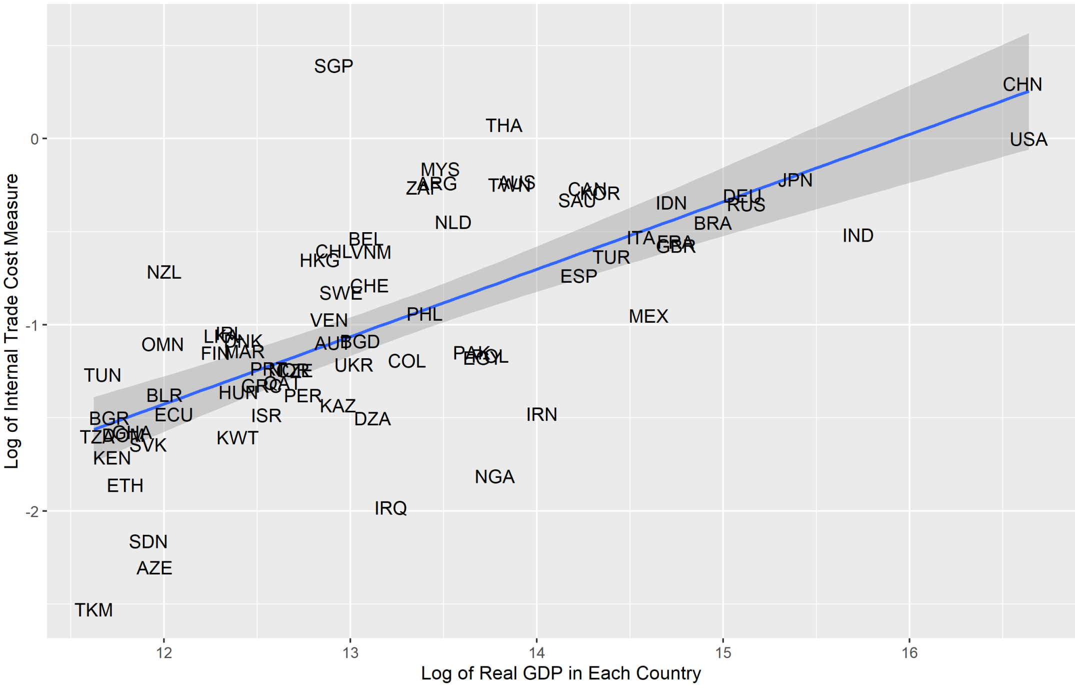
2. Given $\hat{S}_i$ and $\hat{\tau}_{ij}$ as estimates of country competitiveness and trade cost, the price index estimate can be computed as:

$$p_i = \left(\sum_j \exp(\hat{S}_j \hat{\tau_{ji}}) \right)^{-1/\theta}.$$

3. Given the value of L_i, α_i and Π_{ij}, the following equation can be used to obtain w_i up to a normalization:

$$w_i \frac{L_i}{\alpha_i} = \sum_j \Pi_{ij} w_j \frac{L_j}{\alpha_j}.$$

Figure 18: Internal Trade Cost Measure across Country Real GDP



Note: The internal trade cost measure in the USA is normalized at 1. The year for population size is 2014.

4. Using the Capital Labor ratio to obtain r_i:

$$\frac{w_i L_i}{\alpha_i} = \frac{r_i K_i}{1-\alpha_i}.$$

5. Given estimates of w_i, r_i, p_i, $\hat{S}_i$, the technology paramater can be recovered:

$$\exp(\hat{S}_i) = T_i c_i^{-\theta},$$
$$c_i = A\left(w_i^{\alpha} r_i^{1-\alpha}\right)^{\beta} p_i^{1-\beta}.$$

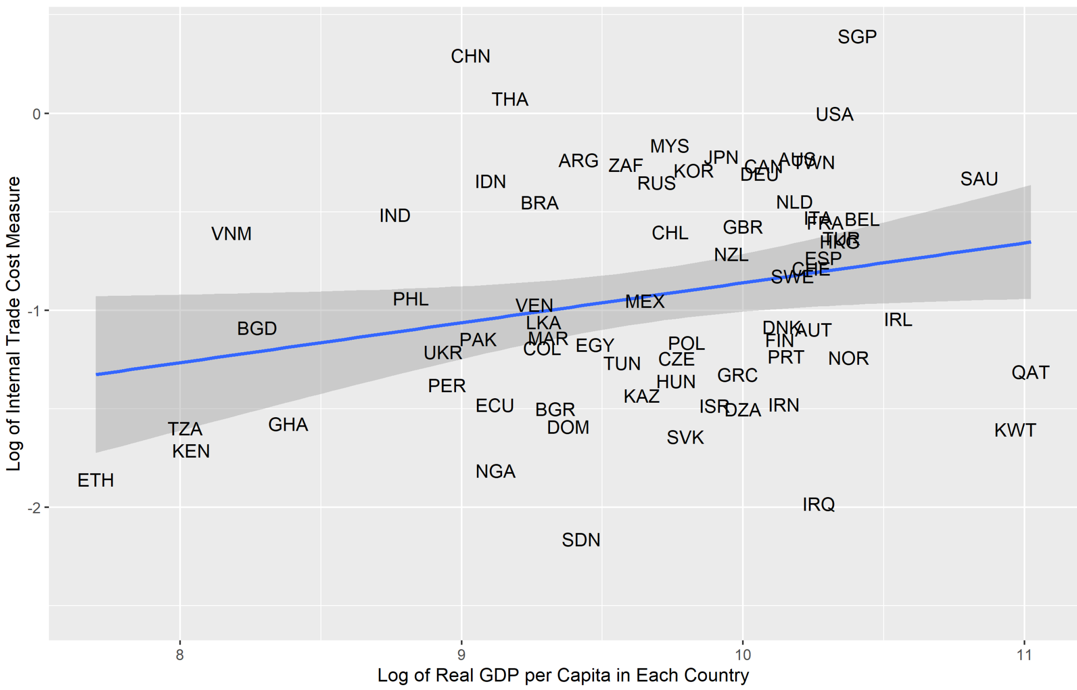
6. Given the value of L_i, the technology intensity parameter can be recovered:

$$\phi_i = T_i / L_i.$$

7. The final goods price and real income can be backed up as:

$$p_F = B\left(w_i^{\alpha_i} r_i^{1-\alpha_i}\right)^{\gamma} p_i^{1-\gamma},$$
$$W_i = \frac{w_i + r_i K_i / L_i}{p_F}.$$

Figure 19: Internal Trade Cost Measure across Country Real GDP per Capita

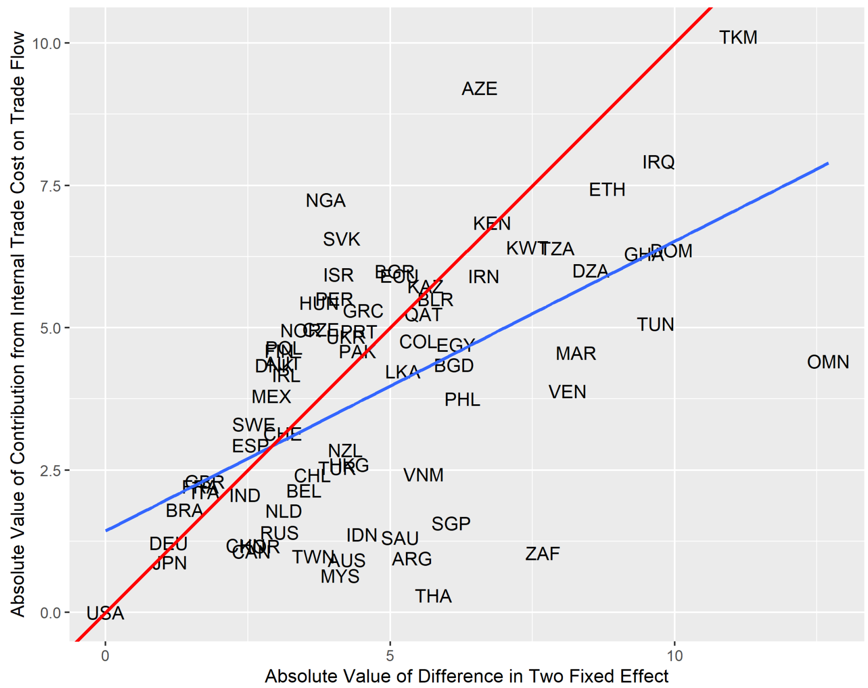


Note: The internal trade cost measure in the USA is normalized at 1. The year for population size is 2014. The real GDP per capita is computed as real GDP divided by the total number of employment.

B.6 Model B's Fit in Real Income and Final Price

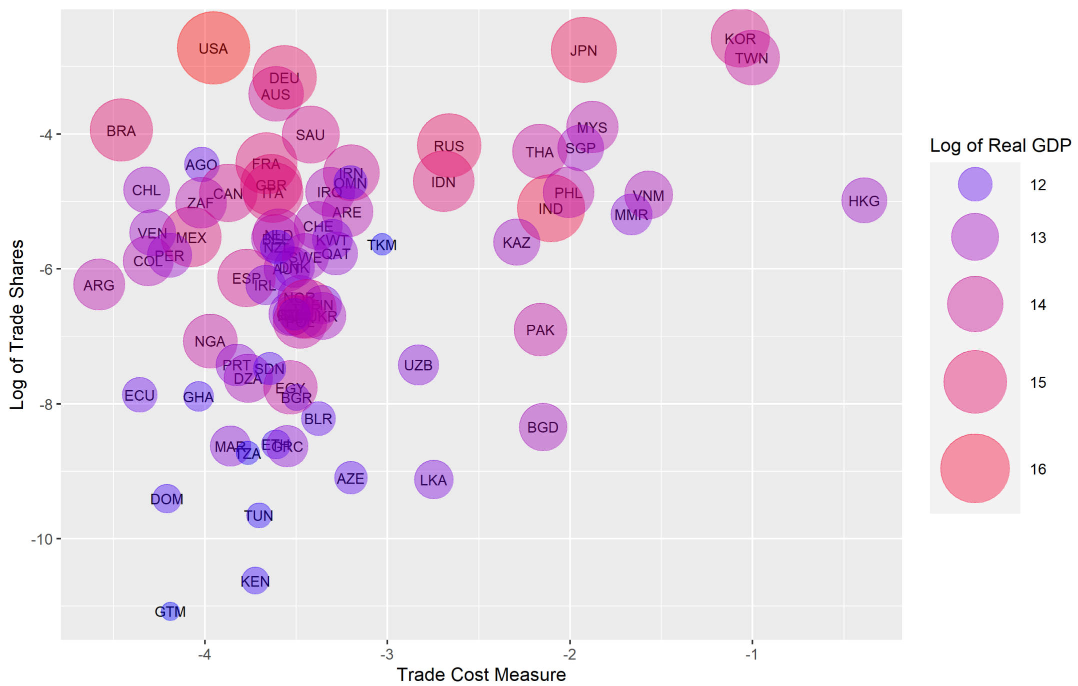
The model B's fit in matching real income data is 0.967 which is the same as model A and its fit in matching final price is 0.785 which is slightly lower than model A.

Figure 20: Differences in Fixed Effects versus Internal Trade Cost



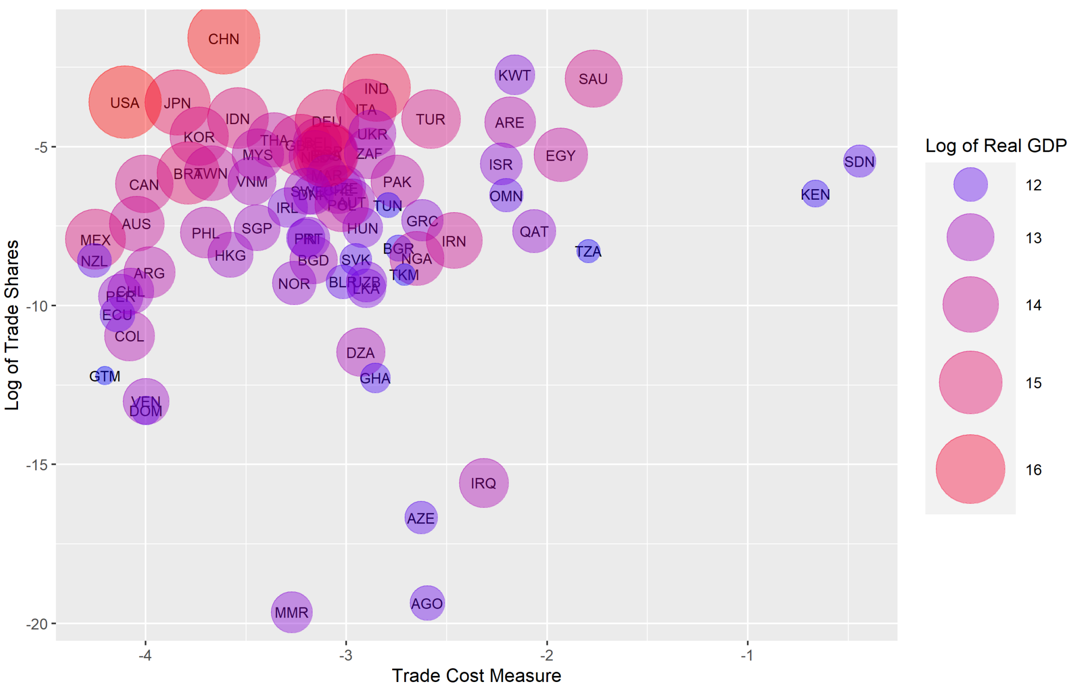
Note: The values on x axis are the absolute value of mean difference between two types of fixed effects of the same country in a ratio type gravity estimation for year 2014. The values on y axis are the absolute value of contribution from internal trade cost on internal trade flow in a panel trade level gravity estimation. Both sets of values are computed relative to the USA level.

Figure 21: Trade Shares across Countries for CHN as Importer



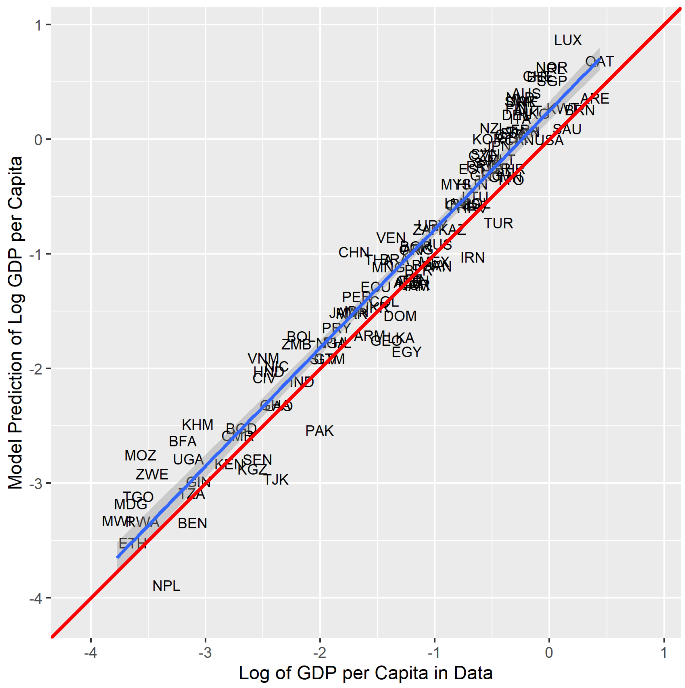
Note: The trade shares are the expenditure share CHN spends on each exporter's goods and they are calculated among the countries displayed with trade flows in year 2014.

Figure 22: Trade Shares across Countries for ETH as Importer



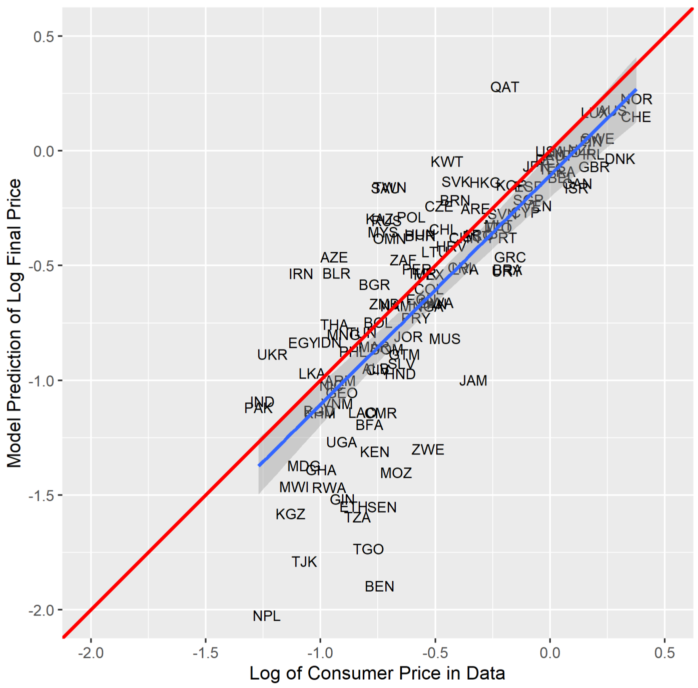
Note: The trade shares are the expenditure share ETH spends on each exporter's goods and they are calculated among the countries displayed with trade flows in year 2014.

Figure 23: Real Income Prediction from Model B versus Data



Note: The real GDP per capita is computed as real GDP divided by total employment in that country in year 2014. Both data and model predictions are normalized with respect to the USA.

Figure 24: Consumer Price Prediction from Model B versus Data



Note: Both data and model predictions are normalized with respect to the USA. The data in year 2014 is used.