Trade and Interdependence in a Spatially Complex World*

Michal Fabinger†

Harvard University

December 31, 2011

Abstract

This paper presents an analytic solution framework applicable to a wide variety of general equilibrium international trade models, including those of Krugman (1980), Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Melitz (2003), in multi-location cases. For asymptotically power-law trade costs and in the large-space limit, it is shown that there are parameter thresholds where the qualitative behavior of the model economy changes. In the case of the Krugman (1980) model, the relevant parameter is closely related to the elasticity of substitution between different varieties of goods. The geographic reach of economic shocks changes fundamentally when the elasticity crosses a critical threshold: below this point shocks are felt even at long distances, while above it they remain local. The value of the threshold depends on the approximate dimensionality of the spatial configuration. This paper bridges the gap between empirical work on international and intranational trade, which frequently uses data sets involving large numbers of locations, and the theoretical literature, which has analytically examined solutions to the relevant models with realistic trade costs only for the case of very few locations.

*I am grateful to Mark Aguiar, Manuel Amador, James Anderson, Pol Antràs, Dave Donaldson, Gita Gopinath, Elhanan Helpman, Oleg Itskhoki, Marc Melitz, Robert Townsend, and Glen Weyl for extremely helpful discussions, and to seminar participants at Harvard University for very useful comments.

†Email: fabinger@fas.harvard.edu

1 Introduction

Imagine that there are two large neighboring countries and that the costs of moving goods across their shared border changes. How far from the border is the economic impact going to be felt? Do such changes mostly affect regions close to the border, or do they significantly affect even very distant locations? What if productivity increases or decreases in one of these countries, due to an economic boom or due to a crisis? How is the productivity change going to influence the level of welfare at various places in the other country?

To address these questions, it is natural to employ standard models of international trade, such as Krugman (1980). The solutions to these models have been theoretically analyzed for some cases. If there are just two or three locations where economic activity takes place, the analysis is very straightforward.[1] To gain insight into situations with many locations, the theoretical literature has used certain analytically convenient specifications of trade costs. Apart from zero trade costs, the most popular assumption corresponds to 'symmetric trade costs', in which case the cost of trade between any pair of distinct locations is the same.[2] For example, the multilateral trade policy analysis in Baldwin et al. (2005) builds on this assumption.

For the present purposes, however, it is necessary to work in a multi-location setting with more realistic trade costs. Clearly, the transportation costs should grow with distance. At the same time, they should reflect economies associated with shipping goods over long distances: the per-unit-distance transportation cost should be a decreasing function of distance.[3]

The empirical literature has been working with trade models at this level of realism for a long time. In recent years, the multi-location aspect has become prominent in empirical work. Due to falling costs of information technology, highly spatially disaggregated data

[1]Matsuyama (1999) solves interesting cases with as many as eight locations in the context of the model introduced in Section 10.4 of Helpman and Krugman (1987), which adds a costlessly tradable homogeneous good to Krugman (1980).

[2]In the context of economic geography models (see Fujita et al. (1999)), trade costs exponential in distance proved to be a convenient choice. In that specification, the per-unit-distance trade cost is an increasing function of distance.

[3]See Anderson and van Wincoop (2004) and Hummels (2001) for empirical evidence on trade costs.

sets are becoming available for empirical analysis. For example, Hillberry and Hummels (2008) study manufacturers' shipments within the United States with 5-digit zip code precision. Compared to previous studies this is a remarkable improvement in spatial resolution.

The aim of the present paper[4] is to bridge the gap between the context in which international trade models are used for empirical purposes and the context in which they are studied theoretically. The article introduces a mathematical framework[5] that allows one to solve and analyze such trade models in basic cases involving many locations. The model discussed extensively is that of Krugman (1980), but this choice is made primarily for expositional purposes. The models of Anderson and van Wincoop (2003)/Armington (1969)[6] or Melitz (2003) have a very similar mathematical structure,[7] and only minor modifications are needed to write down their solutions once the solutions to the Krugman model are known. The same is true[8] for the Ricardian model of Eaton and Kortum (2002). This method may also be applied to many other types of trade models, such as those in Baldwin et al. (2005), where some factors of production are frequently assumed to be mobile.

What are the practical lessons coming from the analysis? Take the Krugman model as a representative example. Let the transportation costs be of the 'iceberg' type and asymptotically power-law[9] in distance, as commonly assumed in the empirical literature. Suppose also that the spatial geometry is very large and homogeneously populated. In this case, it turns out, the way general equilibrium effects spread through the economy depends very strongly on the elasticity of substitution between different varieties of goods. When the elasticity is above a certain threshold, disturbances spread through the economy by short-distance interactions. With the elasticity below the threshold, interactions between

[4] The final version of this paper will include additional discussion of the intuition behind certain technical step and results, omitted in the current version due to temporal constraints. The paper will be continuously updated at http://www.people.fas.harvard.edu/~fabinger/papers.html

[5] The framework makes extensive use of standard tools of functional analysis. In the concrete examples considered, these are Fourier series expansion and spherical harmonic expansion.

[6] The model of Anderson and van Wincoop (2003) is an extension of Armington (1969) and Anderson (1979).

[7] See Arkolakis et al. (forthcoming) for a detailed analysis of the similarities between the models.

[8] Also the portfolio choice model of Okawa and van Wincoop (2010) has the same property.

[9] For a clarification of the term 'asymptotically power-law,' see Subsection 4.3.

economic agents separated by long distances play a crucial role. This fact has important consequences for various quantities of interest.

Consider the case of two large neighboring countries mentioned earlier, and suppose that the cost associated with moving goods across the border increases slightly. If the elasticity is above the threshold, only locations close to the border will be affected. On the other side of the threshold, the change in the border cost significantly affects all locations. In the case of a productivity change in one of the countries, the situation is similar. With the elasticity above the threshold, the effects on the other country will be restricted to a small region close to the border. For the elasticity below the threshold, the consequences of the productivity change will be felt throughout the other country.

At the empirical level, these observations imply that when fitting a similar trade model to the data, the usual practice of assuming that all differentiated goods have the same elasticity of substitution can lead to unexpectedly strong biases. The properties of the model are highly non-linear in the elasticity. Under such circumstances, replacing heterogeneous goods with a single type of good having the average elasticity is misleading. A related kind of bias arises when the elementary regions in the data set do not have the same size. The range of goods contributing to the observed trade flows strongly depends on the size of each elementary region, leading to a spatial version of selection bias.

The existence of the threshold arises from the interplay between the economic structure of the model and its spatial properties. It is not something that two-, three-, or four-location cases would reveal. The value of the threshold is closely tied to the dimensionality[10] of the spatial configuration. If the spatial geometry is roughly one-dimensional, meaning that economic agents are arranged along a line or a circle, the threshold lies at one particular value for the elasticity of substitution. If economic agents are spread through a two-dimensional geometry, the value of the threshold is significantly higher.

The solution method used here is easy to generalize to more complex situations. For example, even though the focus of this paper is on static models, dynamic models can be

[10]The value of the threshold is a linear function of the dimension of space. It is meaningful to consider zero-dimensional cases as well. This corresponds to spatial configurations with just a few (point-like) locations. Here the threshold condition translates into the requirement that the elasticity of substitution be equal to 1. In this case the utility function becomes Cobb-Douglas, which is known to exhibit behavior qualitatively different from the cases with elasticity of substitution greater than 1.

solved[11] in a similar fashion. Adding uncertainty does not represent an obstacle, nor does the addition of differentiated goods with different elasticities of substitution.

The present paper is related to two overlapping strands of economic[12] research. The first one is concerned with various aspects of empirical data on trade flows (which are generally consistent with the 'gravity model of trade'). The analysis here is most closely connected to the four models of Krugman (1980), Eaton and Kortum (2002), Melitz (2003), and Anderson and van Wincoop (2003)/Armington (1969), each associated with empirical literature[13] too rich to explicitly cite here.

The other strand of related research studies the influence of international borders on trade flows (McCallum (1995), Anderson and van Wincoop (2003), Behrens et al. (2007), Rossi-Hansberg (2005)) and on price fluctuations (Engel and Rogers (1996), Gorodnichenko and Tesar (2009), Gopinath et al. (forthcoming)).

The rest of the paper is organized as follows. The next section justifies the use of functional analysis in later parts of the paper by discussing various pitfalls associated with oversimplified approaches to multi-location economies. Section 3 reviews the basics of the representative example of choice, namely the Krugman (1980) model. It also introduces certain concepts needed to characterize the comparative statics of the model. Section 4 provides a formal (first-order) solution to the model in the form of an infinite series. Section 5 uses Fourier series expansion to derive an explicit general solution to the model in the case of a circular geometry. The resulting formula is then used to analyze two special cases: the impact of changes in border costs in Section 6, and changes in productivity in Section 7. Spherical geometry is discussed in Section 8, with spherical harmonic expansion playing the role of the Fourier series expansion. Section 9 considers the structure of higher-order terms. In addition to the appendices included in the paper, there is an online appendix[14] providing detailed derivations of certain results, as well as

[11] The solutions will appear in Fabinger (2011).

[12] There is also a close link to the physics literature; see Section 9 and Appendix L.

[13] Recent examples include Helpman et al. (2004) and Helpman et al. (2008). In the context of the present paper, it is worth noting that Alvarez and Lucas (2007) establish important properties of the Eaton and Kortum (2002) model and provide a basis for solving the model numerically. In addition, they solve the model analytically under the assumption of zero trade costs and under the assumption of 'symmetric trade costs' mentioned earlier.

[14] The current (incomplete) version is available at http://www.people.fas.harvard.edu/~fabinger/papers.html

a discussion of additional examples of interest.

2 Challenges of multi-location models

Differentiated goods models, as well as a certain type of Ricardian models, typically lead to large non-linear systems of equations.[15] The number of equations as well as the number of unknowns is proportional to the number of locations considered. It is clearly desirable to be able to theoretically analyze the solutions to these models even when there is a large number of locations. However, with realistic[16] trade costs this represents a technical challenge. Even after (log)-linearization the behavior of the system is far from obvious. The equations become linear,[17] which certainly is a simplification, but the number of equations and unknowns is not reduced. To solve the system, one needs to invert a large matrix, which is an obstacle[18] for the analytic approach.

The present paper uses methods of functional analysis to overcome this difficulty. The reader may ask whether it is really necessary to go through all the calculations in order to get a correct picture of the economic phenomena. Could it be that certain shortcuts lead to qualitatively correct results? The rest of the section is devoted to two such possibilities: working with a few locations only (Subsection 2.1) and neglecting indirect general equilibrium interdependencies (Subsection 2.2).

2.1 Working with only a few locations

Let us look at a very simple situation in which economic activity takes place at many different locations. In this example, the physical space is a continuous circle parametrized by the angle $\theta \in (-\pi, \pi]$. At every point, there are profit-maximizing firms, each produc-

[15] An example may be found in Section 3, eq. (3), where each equation links the GDP at a particular location to the GDP elsewhere in the economy. This particular case corresponds to Krugman (1980), but analogous equations for other models have a very similar structure.

[16] The term 'realistic trade costs' here refers to trade costs that increase with distance, but not as fast as to make the per-unit-distance cost also increasing in distance, as discussed in the introduction.

[17] For trade models where already the exact equations are linear, see Baldwin et al. (2005). An example is the 'footlose capital' model of Martin and Rogers (1995).

[18] Cramer's rule, which expresses the solution to a linear system of equations in terms of a ratio of determinants, is of little help here. The determinants are so complicated that they provide little insight into the nature of the solution.

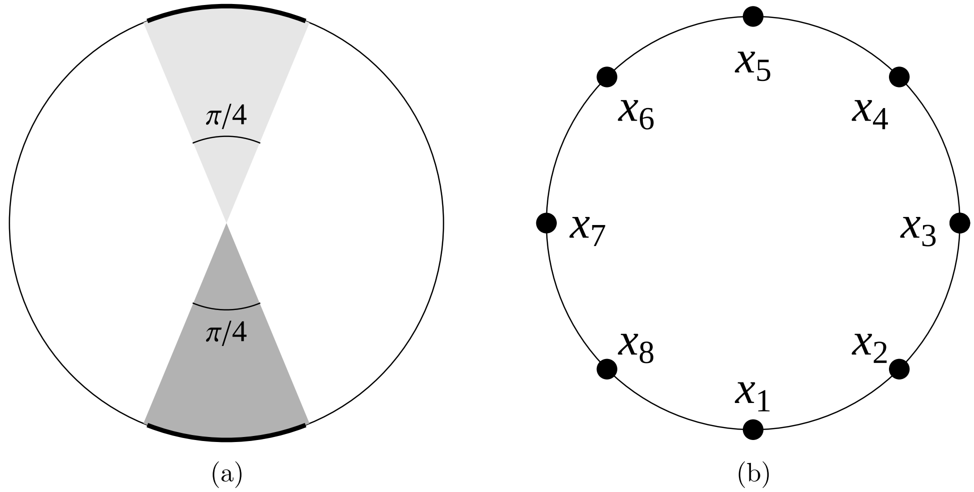


Figure 1: (a) The continuous spatial configuration and (b) its discrete approximation.

ing a different variety of differentiated goods. Only local inputs are used in production. Consumer preferences for the varieties correspond to a constant elasticity of substitution $\sigma \in (1, \infty)$. Apart from the monopoly power of the firms, all markets are free and competitive. Both the setup of the model and the equilibrium involve a complete symmetry between different locations on the circle. To have a concrete model in mind, one can consider, for example, the model of Krugman (1980) or Anderson and van Wincoop (2003) /Armington (1969).

Trade costs are of the 'iceberg' type and are characterized by the function[19] $\tau(d) = (1 + \alpha d)^{\rho}$. When any good is transported over a distance d, a fraction $(\tau(d) - 1)/\tau(d)$ will be lost. Distance is measured along the circle, and is proportional to the angle between the two locations. The parameters α and ρ are positive exogenously-given constants.

These assumptions are enough to determine the share of expenditures a consumer at location θ spends on products from any given region. For concreteness, consider the consumer located in the middle of the lower shaded angle in Fig. 1a and calculate the share s of expenditures on goods from the upper shaded angle in the figure. A short calculation reveals that

$$s = \frac{\pi\alpha R - 1}{1 + \pi\alpha R - (1 + \pi\alpha R)^{\rho(\sigma-1)}} \left(1 - \frac{7}{8}\left(\frac{1 + \pi\alpha R}{1 + \frac{7}{8}\pi\alpha R}\right)^{\rho(\sigma-1)}\right),$$

[19]The qualitative conclusions of this subsection apply to any trade cost function $\tau(d)$ that is 'asymptotically power-law,' in the sense of Subsection 4.3.

where R is the radius of the circle. In the large-radius limit, the expression for s simplifies.

$$\lim_{R\to\infty} s = \begin{cases} 1 - \left(\frac{7}{8}\right)^{1-\rho(\sigma-1)} & \text{for } \rho\,(\sigma-1) < 1, \\ 0 & \text{for } \rho\,(\sigma-1) > 1. \end{cases} \tag{1}$$

Now suppose that we approximate the circle with a small and fixed number of locations, say eight, as in Fig. 1b. If the radius of the circle is very large, consumers at x_1 find varieties produced at other locations very expensive relative to those from x_1. They will spend almost all of their income on local products. As a result, the counterpart[20] of s approaches zero as $R \to \infty$ even when $\rho\,(\sigma-1) < 1$.

This line of reasoning leads to the conclusion that it is impossible to qualitatively reproduce the correct result (1) with a finite and fixed number of locations.[21] It is worth emphasizing that the word 'fixed' is important in the last sentence. The behavior of the continuous model may be reproduced with a discrete one. To do that, one has to increase the number of locations properly with the radius of the circle when taking the large-space limit. In other words, there is nothing special about working with a continuum of locations from the beginning. What is responsible for the failure of the few-location model is not the discreteness of space, but the fact that additional locations are not added when the radius of the circle is increased.

2.2 Neglecting changes in general equilibrium effects

We have seen that one simple way of avoiding algebraic complications, namely working with only a few locations, leads to an impasse. Another way to circumvent the difficulty

[20] In the discrete approximation, there is just one location, namely x_5, at the position of the upper shaded angle of the continuous case. For this reason, the discrete counterpart of s is the share of expenditures of consumers at x_1 on products from x_5.

[21] The reader may ask whether it is possible to make the few-location model correctly reproduce the qualitative behavior of the continuous model by a simple modification of its assumptions. What if we assume that even goods produced and consumed at the same location have to travel a certain distance, say one-half of the spacing between neighboring locations? It turns out that such assumption does not lead to the desired outcome. It is true that for $\rho\,(\sigma-1) < 1$ the counterpart of s will be non-zero in the large-space limit. However, under the same assumption the limit of the counterpart of s remains large even in the case $\rho\,(\sigma-1) > 1$. Moreover, the magnitude of the deviation from (1) depends strongly on the arbitrary choice of the number of locations in the discrete model. The departure from the correct value is attenuated only if the number of locations is chosen to be large, contradicting the purpose of the approximation.

is to neglect general equilibrium feedback effects when performing comparative statics exercises. In principle, such approach could yield qualitatively correct results. It turns out, however, that even the signs of the resulting quantities may be incorrect, as discussed in Appendix B.

To answer the questions raised in the introduction, it is necessary to work with a model involving many locations and to incorporate all general equilibrium effects.

3 The Krugman model

3.1 Production and transportation

Consider the static model[22] of trade described in Krugman (1980). The spatial geometry consists of N locations x_i with $i = 1, 2, .., N$. There is a single factor of production, referred to as labor. Labor markets are competitive, and labor is inelastically supplied. Its endowment at location x_i will be denoted $L_{(x_i)}$. There is a continuum of varieties of goods, each produced by a different monopolistically competitive firm at a single location. Individual varieties are labelled by $\omega \in \Omega$, where Ω is the variety space. To produce an amount q of all varieties between ω and $\omega + d\omega$, for some infinitesimal measure $d\omega$ of varieties, the firms need $F\ d\omega$ units of labor to cover their fixed overhead costs, and additional $q\ d\omega$ units of labor to cover their variable costs. Note that this choice corresponds to a particular normalization of the measure of quantity of the goods.

The model uses the 'iceberg' specification of trade costs. The goods can be transported from any location x_i to any location x_j, but a fraction $\left(\tau_{(x_i,x_j)} - 1\right)/\tau_{(x_i,x_j)}$ will be lost on the way, making the total marginal cost $\tau_{(x_i,x_j)}$ times higher than the manufacturing marginal cost. For obvious reasons, $\tau_{(x_i,x_j)} \geq 1$.

Entry into the industry is free. Consequently, the firms earn zero profits. Given this assumption, the reader can easily verify that if the elasticity of substitution between any two varieties is σ, the firm will find it optimal to spend $\sigma - 1$ times more on variable costs than on fixed costs. As a result, the total measure of varieties produced at x_i is

[22] The introductory exposition closely follows that of Eaton and Kortum (in progress). The reader may consult this reference for more detail on the derivation of the main equations of the model.

$H_{(x_i)} = \frac{1}{\sigma F} L_{(x_i)}$ in this case.

3.2 Consumption

The per-capita consumer utility at a particular location is given by

$$u = \left(\int q^{\frac{\sigma-1}{\sigma}} (\omega) \, d\omega \right)^{\frac{\sigma}{\sigma-1}},$$

where $q(\omega)$ represents the per capita consumption of variety ω, $\sigma > 1$ is the elasticity of substitution, and the integral is over all varieties available. The per capita spending $p(\omega) q(\omega)$ on variety ω is given by

$$p(\omega) q(\omega) = \left(\frac{p(\omega)}{P} \right)^{1-\sigma} c.$$

Here $p(\omega)$ denotes the price of variety ω, the per capita consumption expenditure is $c = \int p(\omega) q(\omega) \, d\omega$, and the local price index P is defined as

$$P = \left(\int p^{1-\sigma} (\omega) \, d\omega \right)^{\frac{1}{1-\sigma}}.$$

To avoid terminological complications, each person is endowed with one unit of labor, and per capita and per unit labor quantities coincide. GDP per capita will be denoted y, to be consistent with the notation for consumption per capita.

3.3 Closing the model

The GDP[23] $y_{(x_i)} L_{(x_i)}$ at location x_i is equal to the revenue its firms collect from the measure $\frac{1}{\sigma F} L_{(x_i)}$ of varieties they produce,

$$y_{(x_i)} = \frac{1}{\sigma F} \sum_{j=1}^{N} \left(\frac{p_{(x_i, x_j)}}{P_{(x_j)}} \right)^{1-\sigma} c_{(x_j)} L_{(x_j)}.$$

[23] Note that local wages are equal to the local GDP per capita, because labor is the only factor of production and firms earn zero profits.

Here $p_{(x_i,x_j)}$ is the price firms from x_i charge at x_j. Setting the markup $p_{(x_i,x_j)}/\left(\tau_{(x_i,x_j)}y_{(x_j)}\right)$ to its optimal value of $\sigma/(\sigma-1)$ and imposing budget constraints $y_{(x_j)}=c_{(x_j)}$, the equation becomes

$$y_{(x_i)}=\frac{1}{\sigma F}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\sum_{j=1}^{N}\left(\frac{y_{(x_i)}\tau_{(x_i,x_j)}}{P_{(x_j)}}\right)^{1-\sigma}y_{(x_j)}L_{(x_j)},$$

with the price index given as

$$P_{(x_j)}=\frac{\sigma}{\sigma-1}\left(\frac{1}{\sigma F}\sum_{k}\tau_{(x_k,x_j)}^{1-\sigma}y_{(x_k)}^{1-\sigma}L_{(x_k)}\right)^{\frac{1}{1-\sigma}}. \tag{2}$$

Combining the last two equations yields

$$y_{(x_i)}^{\sigma}=\sum_{j=1}^{N}\frac{\tau_{(x_i,x_j)}^{1-\sigma}y_{(x_j)}L_{(x_j)}}{\sum_{k=1}^{N}\tau_{(x_k,x_j)}^{1-\sigma}y_{(x_k)}^{1-\sigma}L_{(x_k)}}. \tag{3}$$

This is a set of N equations that must hold in equilibrium, and together they determine the economic outcome. The choice of units in which y is measured is arbitrary.[24] We are free to pick a numéraire good and normalize its price to 1. (In the subsequent discussion, a different, more abstract condition will be imposed, in order to keep the calculations simple.)

3.4 Comparative statics - part 1

The rest of the section discusses the comparative statics of the Krugman model, motivates the definition of the GDP propagator, and establishes its basic properties. Readers interested primarily in the concrete results of the paper, not in their detailed derivation, may proceed to Section 4.

Consider a small change in trade costs,[25] with the goal of evaluating the induced change in GDP at different places. For ease of notation, denote $T_{(x_i,x_j)}\equiv\tau_{(x_i,x_j)}^{1-\sigma}$. This

[24]The set of equations (3) is homogeneous in y.

[25]The general method employed in this paper is elucidated using simple examples in Appendix C.

quantity is sometimes referred to as freeness of trade. The GDP equations are

$$y_{(x_i)} = \left(\sum_{j=1}^{N} \frac{T_{(x_i,x_j)} y_{(x_j)} L_{(x_j)}}{\sum_{k=1}^{N} T_{(x_k,x_j)} y_{(x_k)}^{1-\sigma} L_{(x_k)}} \right)^{\frac{1}{\sigma}} . \tag{4}$$

Suppose we know y corresponding to some particular T. We are interested in the change $y \to y + dy$ caused by a change $T \to T + dT$. Here $y \equiv \left(y_{(x_1)}, ..., y_{(x_N)}\right)'$ and T is a collection of $T_{(x_i,x_j)}$. The standard prescription for deriving first-order comparative statics is to differentiate both sides of the equation, leading to

$$dy_{(x_i)} = \sum_{i=1}^{N} G_{(x_i,x_j)} L_{(x_j)} dy_{(x_j)} + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} H_{(x_i,x_j,x_k)} dT_{(x_j,x_k)}. \tag{5}$$

Here $L_{(x_j)} G_{(x_i,x_j)}$ is the derivative of the right-hand side of the ith equation (4) with respect to $y_{(x_j)}$, and $H_{(x_i,x_j,x_k)}$ is its derivative with respect to $T_{(x_j,x_k)}$. In matrix notation, the set of equations above becomes

$$(1 - GL_N)\, dy = \sum_{j=1}^{N} \sum_{k=1}^{N} H_{(x_j,x_k)} dT_{(x_j,x_k)}, \tag{6}$$

with the $N \times N$ matrix G containing elements $G_{(x_i,x_j)}$, and with the N-dimensional vectors $H_{(x_j,x_k)} \equiv \left(H_{(x_1,x_j,x_k)}, ..., H_{(x_N,x_j,x_k)}\right)'$. The diagonal $N \times N$ matrix $L_N \equiv \text{diag}\left(L_{(x_1)}, ..., L_{(x_N)}\right)$ contains the labor endowments of individual locations on the diagonal. The elements of all of these objects can be computed explicitly if y is known.

The next standard step is to use these equations to express dy in terms of $dT_{(x_j,x_k)}$. To achieve that, one may be tempted to multiply both sides of (6) by $(1 - GL_N)^{-1}$, but the situation requires more caution because such matrix is not well-defined. The homogeneity of eq. (3) implies[26] that GL_N has one eigenvalue equal to 1, associated with the eigenvector y: $GL_N y = y$. Consequently, $1 - GL_N$ has a vanishing eigenvalue and cannot be inverted. For this reason, let us pause here to discuss other properties of the matrix G, which will enable us to complete the calculation.

[26]If eq. (4) is satisfied for some vector y, it must also be satisfied for γy, where γ is a positive number. Replacing y by γy in (4), differentiating with respect to γ, and setting $\gamma = 1$ leads to the conclusion that $GL_N y = y$.

3.5 The GDP propagator

Performing the differentiation of the right-hand side of (4), $G_{(x_i,x_j)}$ can be written as a sum of two parts,

$$G_{(x_i,x_j)} = G_{c,(x_i,x_j)} + G_{p,(x_i,x_j)}, \tag{7}$$

with

$$G_{c,(x_i,x_j)} \equiv \frac{1}{\sigma} y_{(x_i)}^{1-\sigma} \frac{T_{(x_i,x_j)}}{\sum_{k=1}^{N} T_{(x_k,x_j)} y_{(x_k)}^{1-\sigma} L_{(x_k)}},$$

$$\begin{aligned} G_{p,(x_i,x_j)} &\equiv \frac{\sigma-1}{\sigma} y_{(x_i)}^{1-\sigma} y_{(x_j)}^{-\sigma} \sum_{l=1}^{N} \frac{T_{(x_i,x_l)} y_{(x_l)} T_{(x_j,x_l)} L_{(x_l)}}{\left(\sum_{k=1}^{N} T_{(x_k,x_l)} y_{(x_k)}^{1-\sigma} L_{(x_k)}\right)^2} \\ &= \sigma\left(\sigma-1\right) \frac{1}{y_{(x_j)}} \sum_{l=1}^{N} G_{c,(x_i,x_l)} G_{c,(x_j,x_l)} y_{(x_l)} L_{(x_l)}. \end{aligned}$$

The matrix G will be referred to as the GDP propagator,[27] and G_c and G_p are its 'consumption part' and 'production part', respectively. These objects capture the strength of GDP spillovers from one location to another.

The intuition behind these expressions is simple. The GDP at location x_j will affect the GDP at x_i through two different channels. The first channel relates to the consumption at x_j, and corresponds to $G_{c,(x_i,x_j)}$. Location x_i is influenced by the consumption at x_j since firms from x_i have customers there. If GDP increases at x_j, the firms will receive more revenue. This is the reason why $G_{c,(x_i,x_j)}$ is positive. The second channel is more closely related to the production at x_j, and is captured by $G_{p,(x_i,x_j)}$. Firms from x_i compete with firms from x_j for customers elsewhere. Higher $y_{(x_j)}$ means more expensive products from x_j, raising the revenue that firms from x_i receive at x_l. This is again a positive effect, translating into a positive $G_{p,(x_i,x_j)}$. The first effect is direct, so $G_{c,(x_i,x_j)}$ contains $T_{(x_i,x_j)}$. The second effect is indirect, mediated through a third location x_l. For this reason $G_{p,(x_i,x_j)}$ contains $T_{(x_i,x_l)}T_{(x_j,x_l)}$ with l being summed over. The presence of the Ts in the denominators is related to the 'multilateral resistance' terms in the corresponding 'gravity

[27] The algebraic framework used in this paper is an adaptation of the technique of Feynman diagrams, which has become ubiquitous in physics. The term 'propagator' is borrowed from that literature.

equation', whose importance has been emphasized by Anderson and van Wincoop (2003).

Notice that if trade costs are not symmetric (in the sense that $T_{(x_i,x_j)} \neq T_{(x_j,x_i)}$), then the matrix $G_{(x_i,x_j)}$ will not in general be symmetric. (Even if $y_{(x_i)}$ is the same everywhere and $G_{p,(x_i,x_j)}$ is symmetric as a consequence, the consumption part $G_{c,(x_j,x_i)}$ of the propagator can still be asymmetric.)

The N-dimensional vector space to which y belongs can be thought of as a one-dimensional space spanned by y times an $(N-1)$-dimensional vector space $\hat{Y}_{N-1}$ whose elements $\hat{y}$ satisfy $y^T L_N \hat{y} = 0$. We already know that the action of GL_N preserves the one dimensional space: $GL_N y = y$. But it is also true[28] that the $(N-1)$-dimensional space $\hat{Y}_{N-1}$ is preserved by the action of this matrix. (In other words, if $y^T L_N \hat{y} = 0$, then also $y^T L_N (GL_N \hat{y}) = 0$.)

Because both the space $\hat{Y}_{N-1}$ and the space spanned by the vector y are preserved by the action of GL_N, the matrix GL_N may be written as

$$GL_N = \mathbf{P}_{span\{y\}} GL_N \mathbf{P}_{span\{y\}} + \mathbf{P}_{\hat{Y}_{N-1}} GL_N \mathbf{P}_{\hat{Y}_{N-1}} = \mathbf{P}_{span\{y\}} + \mathbf{P}_{\hat{Y}_{N-1}} GL_N \mathbf{P}_{\hat{Y}_{N-1}}, \quad (8)$$

where $\mathbf{P}_{span\{y\}}$ is the projector onto the one-dimensional space generated by the vector y, and $\mathbf{P}_{\hat{Y}_{N-1}}$ is the projector onto $\hat{Y}_{N-1}$.

3.6 Comparative statics - part 2

Now let us go back to the discussion of (6). We have not imposed any normalization condition on $y+dy$ yet. The international trade literature typically chooses a definite good to serve as numéraire, and normalizes its price to 1. Such choice would be inconvenient in the present context. To take advantage of the decomposition (8), we need to impose the more abstract[29] condition $y^T L_N dy = 0$, i.e. $dy \in \hat{Y}_{N-1}$. It follows that $GL_N dy \in \hat{Y}_{N-1}$,

[28] To verify this property, it is sufficient to show that $G^T L_N y = ay$ for some constant a. Direct evaluation using the expressions for $G_{c,(x_i,x_j)}$ and $G_{p,(x_i,x_j)}$ above confirms that this is indeed the case with $a = 1$, i.e. that $G^T L_N y = y$.

[29] If all elements of the vector y have the same magnitude, this condition translates into the requirement that the total (nominal) GDP remain fixed as the trade costs change. More generally, the quantity kept fixed is a weighted average of the GDP at individual locations. The same condition may be interpretted in terms of wages, since these are identically equal to the GDP per capita in this model.

and as a result of (6), also that $\sum_{j=1}^{N}\sum_{k=1}^{N} H_{(x_j,x_k)} dT_{(x_j,x_k)} \in \hat{Y}_{N-1}$. Thanks to these properties, the equation (6) can be written as

$$\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}} dy = \sum_{j=1}^{N}\sum_{k=1}^{N} H_{(x_j,x_k)} dT_{(x_j,x_k)}.$$

Since dy and the right-hand side of this equation belong to $\hat{Y}_{N-1}$, and $\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}}$ is an operator on $\hat{Y}_{N-1}$, we can restrict attention to that space and conclude that[30]

$$dy = \left(\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}}\right)^{-1} \sum_{j=1}^{N}\sum_{k=1}^{N} H_{(x_j,x_k)} dT_{(x_j,x_k)}. \tag{9}$$

Here, of course, the inversion is performed in $\hat{Y}_{N-1}$, not in the full N-dimensional space. As discussed in Subsection 3.4, GL_N has one eigenvalue equal to 1 and associated with the eigenvector y. Stability of the system implies that all other eigenvalues are smaller than 1 in absolute value. For this reason $\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}}$ is invertible in $\hat{Y}_{N-1}$, and the final expression for dy is well-defined.

4 The Krugman model in continuous space

While introductory exposition is simpler with a finite number of locations, the examples discussed below will involve continuous space. Retaining a fine discrete grid in the model would not lead to any additional economic insights, and the continuous-space examples provide greater algebraic convenience. The equations of the model may easily be translated into continuum notation.

Let the spatial geometry be a continuous space with points parametrized by a vector of coordinates x. In general, the space can be curved. The coordinates are chosen arbitrarily.

[30] The continuous-space analog of this equation is the relation (21) in Section 4.

Denote the labor element[31] at location x by $dL(x)$. The equation (3) for GDP becomes

$$y^{\sigma}(x) = \int \frac{T(x,x')\, y(x')}{\int T(x'',x')\, y^{1-\sigma}(x'')\, dL(x'')} dL(x'), \tag{10}$$

where $T(x,x') \equiv \tau^{1-\sigma}(x,x')$. The degree of interdependence between different locations is captured by the GDP propagator[32] defined[33] as

$$G(x,x') = G_c(x,x') + G_p(x,x'), \tag{11}$$

with the 'consumption part'

$$G_c(x,x') = \frac{1}{\sigma} \frac{y^{1-\sigma}(x)\, T(x,x')}{\int y^{1-\sigma}(x'')\, T(x'',x')\, dL(x'')}, \tag{12}$$

and the 'production part'

$$G_p(x,x') = \sigma(\sigma-1) \frac{1}{y(x')} \int G_c(x,x'')\, G_c(x',x'')\, y(x'')\, dL(x''). \tag{13}$$

Intuitively, the GDP propagator $G(x,x')$ measures how strongly an infinitesimal change in GDP at x' influences the GDP at x. The consumption part reflects the fact that if consumption at x' increases, this will directly increase the sales of firms from x. The production part arises from the fact that increased GDP (wages) at x' make it easier for firms from x to compete in other markets.[34]

The GDP propagator satisfies the conditions[35]

$$y(x) = \int G(x,x')\, y(x')\, dL(x'), \;\; y(x) = \int G(x',x)\, y(x')\, dL(x'). \tag{14}$$

[31] To follow the discussion, the reader does not have to be familiar with various concepts of differential geometry. Nevertheless, they are useful for expressing $dL(x)$ in more explicit terms. The distances in the physical space are captured by a definite metric tensor whose values depend on x. Denoting its determinant $g(x)$, the endowment of labor $dL(x)$ in a particular coordinate element dx equals $\sqrt{g(x)}dx$ times the labor density.

[32] As mentioned in Subsection 3.5, the term 'propagator' comes from related physics literature.

[33] For the discrete analog of this definition, see eq. (7).

[34] This intuition is discussed in more detail in Subsection 3.5.

[35] These are analogous to the conditions $y = GL_N y$ and $y = G^T L_N y$ of Subsection 3.5.

The expression for the price index analogous to (2) is now

$$P(x) = \frac{\sigma - 1}{\sigma} \left(\frac{1}{\sigma F} \int T(x', x) \, y^{1-\sigma}(x') \, dL(x') \right)^{\frac{1}{1-\sigma}}. \tag{15}$$

4.1 Change in the solution in response to a small change in trade costs

Now suppose that the trade costs change[36] so that

$$T(x, x') \to (1 - \kappa b(x, x')) \, T(x, x'). \tag{16}$$

The small but finite parameter κ sets the size of the change, while $b(x, x')$ captures the geometric aspects of the change. For example, if the change under consideration was the introduction of a (proportional) cost of crossing a border, then $b(x, x')$ could be set to one whenever x and x' were separated by the border, and set to zero otherwise. The GDP equation (10) will now take the form

$$y^{\sigma}(x) = \int \frac{(1 - \kappa b(x, x')) \, T(x, x') \, y(x')}{\int (1 - \kappa b(x'', x')) \, T(x'', x') \, y^{1-\sigma}(x'') \, dL(x'')} dL(x'). \tag{17}$$

Let us expand the new GDP values in a Taylor series

$$y(x) = y_0(x) + \kappa y_1(x) + \kappa^2 y_2(x) + \ldots$$

Here $y_0(x)$ represents the GDP before the change. The functions $y_1, y_2, y_3, \ldots$ are required to be orthogonal[37] to y_0, in the sense that $\int y_n(x) \, y_0(x) \, dL(x) = 0$ for $n > 0$. These conditions are imposed (instead of fixing the price of a numéraire good) in order to keep

[36] The change in trade costs corresponding to (16) is analogous to the change $T \to T + dT$ considered in the discrete-space case of Subsection 3.4. Besides working in continuous space, the difference here is that the change does not have to be infinitesimal.

[37] The discrete-space analog of these conditions would be $y_0^T L_N y_n = 0$ for $n > 0$. The space of functions considered here is the space of real square-integrable functions with measure $dL(x)$, i.e. the space of functions f for which $\int f^2(x) \, dL(x)$ is finite. This space is usually denoted L_2; see, for example, Section 15.1 of Stokey et al. (1989) for its formal definition. The inner product of functions f and g is defined as $\int f(x) \, g(x) \, dL(x)$.

the calculations simple. The rationale behind this choice is explained in Subsection 3.6.

The main focus of this paper is on the first-order change $y_1(x)$. The higher-order terms y_n, $n \geq 2$, may be computed in an analogous way. They are the subject of Section 9. An equation for the first-order term $y_1(x)$ can be obtained by plugging the Taylor expansion into the GDP equation and comparing terms of the first-order in κ. The details of the calculation can be found in Appendix D. The result is

$$y_1(x) = \int G(x, x') \, y_1(x') \, dL(x') + \int B(x, x') \, y_0(x') \, dL(x'), \tag{18}$$

with the 'primary impact function' $B(x, x')$ defined as

$$B(x, x') = -b(x, x') \, G_c(x, x') + \sigma G_c(x, x') \int b(x'', x') \, G_c(x'', x') \, dL(x''). \tag{19}$$

Alternatively, using an operator notation, this is

$$y_1(x) = (Gy_1)(x) + (By_0)(x).$$

In general, for a given function $F(x, x')$ the action of the corresponding operator F on a function f will be defined[38] as

$$(Ff)(x) = \int F(x, x') \, f(x') \, dL(x'). \tag{20}$$

Since y_1 is orthogonal to y_0, and, due to (14), so is Gy_1, it must be that By_0 is orthogonal to y_0 as well. The equation for $y_1(x)$ can be iterated indefinitely, giving[39]

$$y_1(x) = \sum_{n=0}^{\infty} (G^n B y_0)(x). \tag{21}$$

[38] Note that the measure $dL(x')$ used here corresponds to the labor endowment. The discrete-space analog would be multiplication by the matrix FL_N.

[39] The discrete-space counterpart of this equation is the relation (9). When interpreting the result (21) for y_1, it is useful to compare it to the expression (60) in Appendix C, which applies to the case of two endogenous variables. Obviously, the function By_0 plays the role of the vector v. It is an initial effect of the change in κ. Just like in (60), this effect has an infinite number of echoes, described by the terms $G^n B y_0$ with positive n.

Here we used the identity[40] $\lim_{n\to\infty} G^n y_1 = 0$. For later convenience, let us define also the 'general equilibrium GDP propagator'[41] $G_g(x, x')$ as the integral kernel of the operator

$$G_g = -\sum_{n=0}^{\infty} G^{n+1}. \tag{22}$$

In terms of G_g, the result (21) becomes

$$y_1(x) = ((1 + G_g) B y_0)(x).$$

Another useful expression for y_1 may obtained using the identity $B = -(1 - \sigma G_c)\tilde{G}_c$, which follows from the definition (19) of B. Here $\tilde{G}_c$ is the integral operator corresponding to

$$\tilde{G}_c(x, x') \equiv G_c(x, x')\, b(x, x'). \tag{23}$$

Denoting also

$$\tilde{g}_c(x) \equiv \int \tilde{G}_c(x, x')\, dL(x'), \tag{24}$$

we have

$$\frac{y_1}{y_0} = -(1 + G_g)(1 - \sigma G_c)\, \tilde{g}_c. \tag{25}$$

For future convenience, let us also introduce the notation

$$\hat{g}_c(x) \equiv \int \tilde{G}_c(x', x)\, dL(x'). \tag{26}$$

Of course, if $G_c(x, x')\, b(x, x') = G_c(x', x)\, b(x', x)$ in general, then $\hat{g}_c(x) = \tilde{g}_c(x)$.

The intuition behind the expression (19) for the primary impact function $B(x, x')$ is as follows. If a new trade barrier, say a border, is introduced between x and x', such change is captured by positive $b(x, x')$. There will be two immediate effects on x. First, with the new barrier, firms from x will lose some part of their revenues from x'. This lowers $y(x)$ and is consistent with the first term in (19) being negative. Second, it will be easier for

[40] This follows from the fact that any $G^n y_1$ is orthogonal to y_0, thanks to (14), and from the fact that all eigenvalues of G are smaller than 1 in absolute value, except for the one corresponding to the eigenfunction y_0.

[41] This object captures not only the direct interdependencies, but also all the general equilibrium feedback effects.

these firms to compete with firms from x'' in the market at x', as long as $b\left(x'', x'\right)$ is also positive. This effect increases $y\left(x\right)$. For this reason, the second term in (19) is positive.

4.2 Welfare

The welfare of individual agents is characterized by the local GDP per capita adjusted for the local price index, $y^{(P)}\left(x\right) \equiv y\left(x\right)/P\left(x\right)$, where the price index $P\left(x\right)$ is given by (15) with the replacement $T\left(x', x\right) \rightarrow \left(1 - \kappa b\left(x', x\right)\right) T\left(x', x\right)$. Appendix E shows that the price-index-adjusted analog of $y_1\left(x\right)$, namely $y_1^{(P)}\left(x\right) \equiv \lim_{\kappa \to 0} \left(y^{(P)}\left(x\right) - y_0^{(P)}\left(x\right)\right)/\kappa$, is given by

$$\frac{y_1^{(P)}\left(x\right)}{y_0^{(P)}\left(x\right)} = \frac{y_1\left(x\right)}{y_0\left(x\right)} - \sigma \int G_c\left(x', x\right) \frac{y_1\left(x'\right)}{y_0\left(x'\right)} dL\left(x'\right) - \frac{\sigma}{\sigma - 1} \hat{g}_c\left(x\right), \tag{27}$$

or, in operator notation,

$$\frac{y_1^{(P)}}{y_0^{(P)}} = \left(1 - \sigma G_c\right) \frac{y_1}{y_0} - \frac{\sigma}{\sigma - 1} \hat{g}_c. \tag{28}$$

Here $\hat{g}_c$ is the function defined in (26).

4.3 Asymptotically power-law transportation costs

Before specializing to concrete economic situations, let us pause here to clarify the choice of trade cost functions that will be used in the rest of the paper. The Krugman model uses the 'iceberg' form of trade costs, characterized by the quantity $\tau\left(x, x'\right)$. In principle, the trade costs can depend on many characteristics of location pairs. For example, they are likely to be lower when the two locations share a common language. The present work will abstract from many such possibilities. Instead, the trade costs will take the simple form

$$\tau\left(x, x'\right) = \tilde{\tau}\left(d\right) \tilde{b}\left(x, x'\right).$$

The first factor $\tilde{\tau}\left(d\right)$ corresponds to transportation costs, and depends only on the distance d between x and x'. The second factor $\tilde{b}\left(x, x'\right)$ represents additional costs, such as the cost of crossing international borders. In baseline cases without any additional trade costs, $\tilde{b}(x, x')$ will be set to 1.

It is common in the empirical literature[42] to assume that for large d, $\tilde{\tau}(d)$ is well approximated by a power law:

$$\tilde{\tau}(d) \approx (\alpha d)^{\rho},$$

with $\rho > 0$ and $\alpha > 0$. Of course, $\tilde{\tau}(d)$ cannot be exactly equal to $(\alpha d)^{\rho}$ at short distances. Otherwise the obvious restriction $\tau(d) \geq 1$ would be violated[43] for small enough d. There are several convenient functional forms that ensure the $\tau \geq 1$ restriction is satisfied while preserving the power-law behavior at long distances, for example $(1 + \alpha^2 d^2)^{\frac{\rho}{2}}$, $(1 + \alpha d)^{\rho}$, or $1 + (\alpha d)^{\rho}$. The present article works with finite geometries, such as a circle or a sphere of radius R. In these cases, closely related functional forms $\left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$, $\left(1 + 2\alpha R \sin \frac{d}{2R}\right)^{\rho}$, and $1 + \left(\alpha R \sin \frac{d}{R}\right)^{\rho}$ provide a greater algebraic convenience. At short distances, these coincide with the previous three, while at long distances they still have the same order of magnitude.

For future purposes, let us mention one important property of the six functional forms above. Define the function $\hat{\tau}(d)$ as

$$\hat{\tau}(d) \equiv \begin{cases} 1 \text{ for } d \leq \frac{1}{\alpha}, \\ (\alpha d)^{\rho} \text{ for } d \geq \frac{1}{\alpha}. \end{cases}$$

It is true that for each of the six functional forms $\tilde{\tau}(d)$ considered above, there exist[44] positive constants a_l and a_h independent of R such that

$$a_l \hat{\tau}(d) \leq \tilde{\tau}(d) \leq a_h \hat{\tau}(d) \tag{29}$$

for all $d \in [0, \pi R]$. Loosely speaking, this means that these $\tilde{\tau}(d)$ are similar to the simple function $\hat{\tau}(d)$. In general, monotonic functions satisfying this condition will be referred

[42] See, for example, Anderson and van Wincoop (2003).

[43] Unless $\tilde{b}(x, x')$ is chosen to precisely compensate for the small magnitude of $\tilde{\tau}(d)$ whenever x and x' are close to each other.

[44] Concrete values of these coefficients $\{a_l, a_h\}$ that can be used for the functional forms $\left(1 + \alpha^2 d^2\right)^{\frac{\rho}{2}}$, $(1 + \alpha d)^{\rho}$, $1 + (\alpha d)^{\rho}$, $\left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$, $\left(1 + 2\alpha R \sin \frac{d}{2R}\right)^{\rho}$, and $1 + \left(\alpha R \sin \frac{d}{R}\right)^{\rho}$ are $\left\{1, 2^{\rho/2}\right\}$, $\{1, 2^{\rho}\}$, $\{1, 2\}$, $\left\{(2/\pi)^{\rho}, 2^{\rho/2}\right\}$, $\{(2/\pi)^{\rho}, 2^{\rho}\}$, and $\{(2/\pi)^{\rho}, 2\}$, respectively.

to as 'asymptotically power-law', despite the fact that the geometries under consideration have finite R. In the large R limit, the term 'asymptotically' regains its conventional meaning.

To simplify notation in the rest of the paper, the following combination of ρ and σ will be denoted δ:

$$\delta \equiv \frac{1}{2}\rho\left(\sigma - 1\right).$$

5 The Krugman model on the circle

5.1 Basic setup

Consider the case where the spatial geometry is a circle[45] of radius R with points parametrized by $\theta \in (-\pi, \pi]$, and where the labor density is constant. Identify the coordinate x with θ. The labor element is now $dL(\theta) = \rho_L d\theta$ with $\rho_L = L/(2\pi)$. The endowment of labor per unit of physical length is ρ_L/R. A baseline solution to the Krugman model corresponding to $\tau(\theta, \theta') = \tilde{\tau}(d)$ is easy to obtain. Due to rotational symmetry, the GDP equation is solved by setting the GDP density to a constant, $y_0(\theta) = y_0$. The GDP propagator $G(\theta, \theta')$ associated with this solution depends only on the distance $d(\theta, \theta')$ between its arguments, defined as the smaller of $|\theta - \theta'|$ and $2\pi - |\theta - \theta'|$. For this reason, all the information in $G(\theta, \theta')$ can be captured by a function G with only one argument defined by $G(d(\theta, \theta')) = G(\theta, \theta')$. This specifies the single-argument $G(\theta)$ only for $\theta \in [0, \pi]$. For notational convenience, extend it symmetrically to negative arguments, $G(-\theta) = G(\theta)$, and then periodically over the entire real line, $G(\theta + 2\pi n) = G(\theta)$, $n \in \mathbb{Z}$. The action (20) of the operator G on any (periodically extended) function $f(\theta)$ on the circle can be written as

$$(Gf)(\theta) = \rho_L (G * f)(\theta) = \rho_L \int_{-\pi}^{\pi} G(\theta - \theta') f(\theta') d\theta'. \tag{30}$$

Define also the single argument functions $G_c(\theta)$, $G_p(\theta)$, and $G_g(\theta)$ in a similar way. The symbol $*$ here stands for a 2π-periodic convolution. For any two 2π-periodic functions f

[45] The case of a finite number of locations symmetrically arranged on a circle can be solved in a similar fashion, employing discrete Fourier tranform instead of Fourier series expansion.

and g their 2π-periodic convolution is defined as

$$(f * g)(\theta) = \int_{-\pi}^{\pi} f(\theta - \theta') \, g(\theta') \, d\theta'.$$

In the context of the circular geometry, the term 'convolution' will always refer to the 2π-periodic convolution.

5.2 Expansion in terms of convolution powers of $G_c(\theta)$

We will see that the numerical values of the solutions y_1 can have very different orders of magnitude depending on the values of the parameters of the model, such as ρ or σ. It is desirable to have an intuitive way of finding the correct order of magnitude without performing explicit calculations. For this purpose, let us take a closer look at the mathematical objects the solution contains. Readers interested primarily in the final results for y_1, not in the properties of individual contributions to it, may proceed to the next subsection.

The formal solution (21) can be written as

$$y_1 = \sum_{n=0}^{\infty} \rho_L^n G^{*n} * (By_0), \tag{31}$$

where the nth convolution power $G^{*n}(\theta)$ of $G(\theta)$ is the n-fold (2π-periodic) convolution of the function $G(\theta)$ with itself. Because equations (11) and (13) imply $G(\theta) = G_c(\theta) + \sigma(\sigma - 1)\rho_L G_c^{*2}(\theta)$, the expression for y_1 can be written as

$$y_1 = \sum_{n=0}^{\infty} \rho_L^n \left(G_c + \sigma(\sigma - 1)\rho_L G_c^{*2}\right)^{*n} * (By_0). \tag{32}$$

We see that the right-hand side is a linear combination of various convolution powers G_c^{*m} of the function G_c, convoluted with the function By_0. In order to gain some intuition about the behavior of y_1 for large R, it is necessary understand what the functions G_c^{*m} look like in that case.

5.2.1 The large R limit of $G_c^{*2}(\theta)$ and $G_c^{*m}(\theta)$ with $m \geq 3$

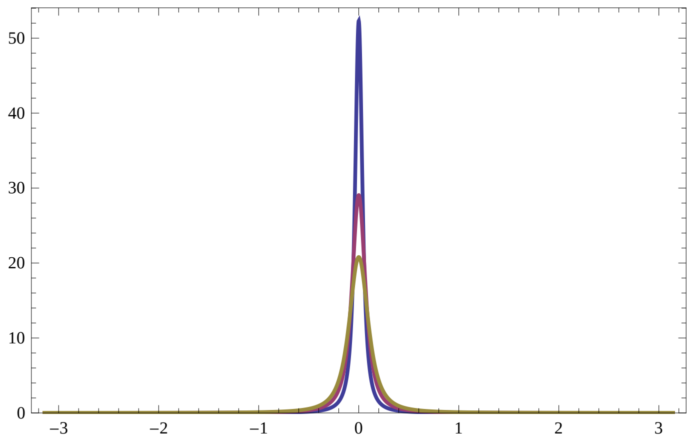
Suppose that R is very large, much larger than $1/\alpha$. The assumption of asymptotically power-law trade costs (29) has implications for the behavior of the function $G_c^{*2}(\theta) \equiv \int_{-\pi}^{\pi} G_c(\theta') G_c(\theta-\theta') d\theta'$. A few of its properties are immediately clear. We know that $G_c(\theta)$ is a positive decreasing function of $|\theta| \in [0,\pi]$. As a consequence, the same must be true for $G_c^{*2}(\theta)$. Also, decreasing $\delta \equiv \rho(\sigma-1)/2$ increases the importance of the tails of the function $G_c(\theta)$, and makes it more spread out. This means that relative to $G_c(\theta)$, any features of the function $G_c^{*2}(\theta)$ will be even more smoothed out. (Note that these observations, as well as those that follow, are consistent with the plots in Fig. 2.)

In order to gain a more detailed intuitive understanding of the properties of $G_c^{*2}(\theta)$, it is important to know which regions of the integration domain dominate the integral. This issue is technical, and for this reason the derivations are left for Appendix F, but the results follow. For $\delta \in \left(0, \frac{1}{4}\right)$, the main contribution to the integral comes from $|\theta'|$ and $|\theta-\theta'|$ being both of order one. For $\delta \in \left(\frac{1}{4}, \frac{1}{2}\right)$ it comes from $|\theta'|$ of order $|\theta|$. When $\delta \in \left(\frac{1}{2}, \infty\right)$, the integral is dominated by the region where $|\theta'|$ is of order $1/(\alpha R)$ and the region where $|\theta'-\theta|$ is of order $1/(\alpha R)$.

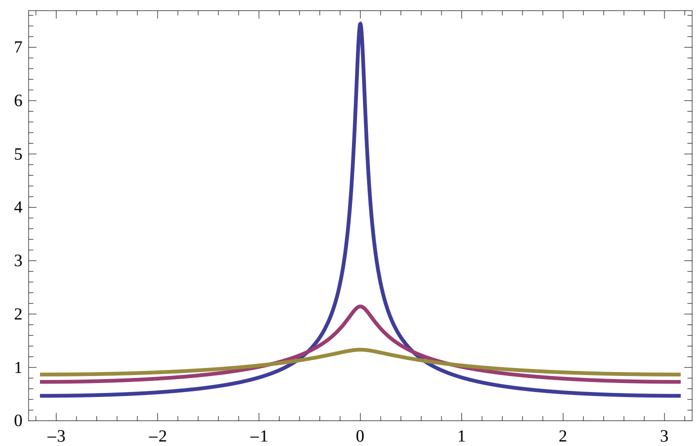
With this knowledge one can make an informed guess about the shape of $G_c^{*2}(\theta)$. With $\delta \in \left(0, \frac{1}{4}\right)$, the integral is insensitive to what happens at short distances of order $1/(\alpha R)$. For this reason, even though $G_c(\theta)$ has a relatively sharp peak, this feature will be smoothed out in the case of $G_c^{*2}(\theta)$. One can expect the maximum $G_c^{*2}(0)$ to be of the same order of magnitude as the minimum $G_c^{*2}(\pi)$. Moreover, $G_c^{*2}(\pi)$ should have a finite positive limit as $R \to \infty$.

For $\delta \in \left(\frac{1}{4}, \frac{1}{2}\right)$, the situation is a little more subtle. For $|\theta|$ of order one, the integral is still dominated by long distances, i.e. by $|\theta'|$ and $|\theta-\theta'|$ of order one. One would therefore expect the minimum $G_c^{*2}(\pi)$ to take similar values as in the previous case. It should stay finite and positive as $R \to \infty$. By contrast, for small $|\theta|$, say of order $1/(\alpha R)$, the integral is dominated by short distances, i.e. by $|\theta'|$ and $|\theta-\theta'|$ of order $1/(\alpha R)$. This contribution is larger than the contribution of long distances, and as a consequence there should be a substantial peak at $\theta = 0$. In other words, $G_c^{*2}(0) \gg G_c^{*2}(\pi)$.

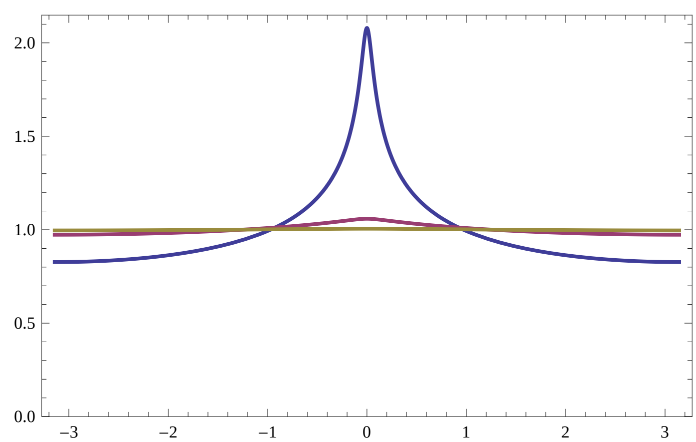
When $\delta \in \left(\frac{1}{2}, \infty\right)$, the story is again relatively simple. Irrespective of the value of θ, the dominant contribution to the integral comes from the short-distance regions, where



(a) δ=1.25



(b) δ=0.375



(c) δ=0.125

Figure 2: Plots of $G_c(\theta)$ (highest peak), $G_c^{*2}(\theta)$, and $G_c^{*3}(\theta)$ (lowest peak) for different values of δ. Trade costs are $(1+4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$ with $\alpha R = 20$. These functions characterize individual contributions to the spreading of economic shocks.

either $|\theta'|$ or $|\theta' - \theta|$ is of order $1/(\alpha R)$. This means that the shape of the function $G_c^{*2}(\theta)$ should be quite similar to the shape of $G_c(\theta)$, with a large peak and quickly decreasing tails. As $R \to \infty$, the minimum $G_c^{*2}(\pi)$ should approach zero. Given the normalization of $G_c(\theta)$, the maximum $G_c^{*2}(0)$ should be of order $G_c(0)/\rho_L$.

Appendix F also contains the derivation of various bounds on the values of $G_c^{*2}(\theta)$. These bounds provide a clearer quantitative picture of the behavior of $G_c^{*2}(\theta)$. As the reader can verify, they are consistent with the intuition just discussed. For the maximum of $G_c^{*2}(\theta)$, which is attained at $\theta = 0$, we have the following bounds

$$\begin{array}{ll} \frac{1}{2\pi\sigma^2}\frac{(1-2\delta)^2}{1-4\delta}\tilde{a}_l^2 \lesssim \rho_L^2 G_c^{*2}(0) \lesssim \frac{1}{2\pi\sigma^2}\frac{(1-2\delta)^2}{1-4\delta}\tilde{a}_h^2 & \text{for } \delta \in \left(0, \frac{1}{4}\right), \\ \frac{(1-2\delta)^2}{2\pi\sigma^2}(\pi\alpha R)^{4\delta-1}\tilde{a}_l^2 \lesssim \rho_L^2 G_c^{*2}(0) \lesssim \frac{(1-2\delta)^2}{2\pi\sigma^2}(\pi\alpha R)^{4\delta-1}\tilde{a}_h^2 & \text{for } \delta \in \left(\frac{1}{4}, \frac{1}{2}\right), \\ \frac{1}{2\sigma^2}\frac{(2\delta-1)^2}{\delta(4\delta-1)}\alpha R\tilde{a}_l^2 \lesssim \rho_L^2 G_c^{*2}(0) \lesssim \frac{1}{2\sigma^2}\frac{(2\delta-1)^2}{\delta(4\delta-1)}\alpha R\tilde{a}_h^2 & \text{for } \delta \in \left(\frac{1}{2}, \infty\right). \end{array} \tag{33}$$

They are written in terms of the quantity $\rho_L^2 G_c^{*2}(0)$, which does not depend on the choice of units in which labor is measured. The constants $\tilde{a}_l \equiv a_l^{\sigma-1}/a_h^{\sigma-1}$ and $\tilde{a}_h \equiv a_h^{\sigma-1}/a_l^{\sigma-1}$ are defined in terms of the constants appearing in (29). The important message these bounds convey is the dependence of $G_c^{*2}(0)$ on the radius R. If $\delta \in \left(0, \frac{1}{4}\right)$, the peak of $G_c^{*2}(0)$ is relatively small and independent of R. When $\delta \in \left(\frac{1}{4}, \frac{1}{2}\right)$, the maximum increases as $R^{4\delta-1}$. For $\delta \in \left(\frac{1}{2}, \infty\right)$, it increases is even faster; it is linearly proportional to R itself.

Now let us look at $G_c^{*2}(0)$ relative to $G_c(0)$.

$$\begin{array}{ll} \frac{1}{\sigma}\frac{1-2\delta}{1-4\delta}\frac{1}{(\pi\alpha R)^{2\delta}}\tilde{a}_l^2 \lesssim \frac{\rho_L G_c^{*2}(0)}{G_c(0)} \lesssim \frac{1}{\sigma}\frac{1-2\delta}{1-4\delta}\frac{1}{(\pi\alpha R)^{2\delta}}\tilde{a}_h^2 & \text{for } \delta \in \left(0, \frac{1}{4}\right), \\ \frac{1}{\sigma}\frac{1-2\delta}{(\pi\alpha R)^{1-2\delta}}\tilde{a}_l^2 \lesssim \frac{\rho_L G_c^{*2}(0)}{G_c(0)} \lesssim \frac{1}{\sigma}\frac{1-2\delta}{(\pi\alpha R)^{1-2\delta}}\tilde{a}_h^2 & \text{for } \delta \in \left(\frac{1}{4}, \frac{1}{2}\right), \\ \frac{1}{\sigma}\frac{4\delta-2}{4\delta-1}\tilde{a}_l^2 \lesssim \frac{\rho_L G_c^{*2}(0)}{G_c(0)} \lesssim \frac{1}{\sigma}\frac{4\delta-2}{4\delta-1}\tilde{a}_h^2 & \text{for } \delta \in \left(\frac{1}{2}, \infty\right). \end{array} \tag{34}$$

When $\delta \in \left(0, \frac{1}{2}\right)$, the ratio $G_c^{*2}(0)/G_c(0)$ decreases with R. This means that $G_c^{*2}(0)$ is quite small relative to $G_c(0)$, which is consistent with significant smoothing out. If $\delta \in \left(\frac{1}{2}, \infty\right)$, the ratio is independent of R. The peak of $G_c(0)$ is preserved to a large extent by the convolution.

For the minimum of $G_c^{*2}(\theta)$ at $\theta = \pi$, we have

$$\begin{aligned} \tfrac{\pi^{4\delta-2}}{4\sigma^2}(1-2\delta)^2 I(\pi)\,\tilde{a}_l^2 \lesssim \rho_L^2 G_c^{*2}(\pi) \lesssim \tfrac{\pi^{4\delta-2}}{4\sigma^2}(1-2\delta)^2 I(\pi)\,\tilde{a}_h^2 \quad & \text{for } \delta \in \left(0, \tfrac{1}{2}\right), \\ \tfrac{1}{2\pi\sigma^2}\tfrac{2\delta-1}{2\delta}\tfrac{1}{(\pi\alpha R)^{2\delta-1}}\tilde{a}_l^2 \lesssim \rho_L^2 G_c^{*2}(\pi) \lesssim \tfrac{1}{2\pi\sigma^2}\tfrac{2\delta-1}{2\delta}\tfrac{1}{(\pi\alpha R)^{2\delta-1}}\tilde{a}_h^2 \quad & \text{for } \delta \in \left(\tfrac{1}{2}, \infty\right). \end{aligned} \tag{35}$$

The function I is defined in (66). Its value $I(\pi)$ is independent of R and is roughly of order one when other parameters do not take extreme values. We see that the minimum is independent of R when $\delta \in \left(0, \frac{1}{2}\right)$, and decreases with R when $\delta \in \left(\frac{1}{2}, \infty\right)$.

The last set of inequalities presented here is

$$\begin{aligned} \tfrac{\pi^{4\delta-2}}{4\sigma^2}(1-2\delta)^2 I(\theta)\,\tilde{a}_l^2 \lesssim \rho_L^2 G_c^{*2}(\theta) \lesssim \tfrac{\pi^{4\delta-2}}{4\sigma^2}(1-2\delta)^2 I(\theta)\,\tilde{a}_h^2 \quad & \text{for } \delta \in \left(0, \tfrac{1}{2}\right), \\ \tfrac{1}{\sigma}\rho_L G_c(\theta)\,\tilde{a}_l \lesssim \rho_L^2 G_c^{*2}(\theta) \lesssim \tfrac{1}{\sigma}\rho_L G_c(\theta)\,\tilde{a}_h \quad & \text{for } \delta \in \left(\tfrac{1}{2}, \infty\right). \end{aligned} \tag{36}$$

These hold for $|\theta|$ much greater than $1/(\alpha R)$. When $\delta \in \left(0, \frac{1}{4}\right)$, the function $I(\theta)$ is roughly of order one for any θ. For $\delta \in \left(\frac{1}{4}, \frac{1}{2}\right)$, it is roughly of order one when $|\theta|$ is of order one. As $|\theta|$ decreases, the function increases indefinitely. But remember that the bound itself is valid only if $|\theta| \gg 1/(\alpha R)$. (A more careful analysis reveals that in this case the peak of $G_c^{*2}(\theta)$ is not very important, it does not contribute much when $G_c^{*2}(\theta)$ is integrated over θ.) When $\delta \in \left(\frac{1}{2}, \infty\right)$, the bound implies that $G_c^{*2}(\theta)$ has tails that look similar to those of $G_c(\theta)$.

Here we discussed only $G_c^{*2}(\theta)$, but $G_c^{*m}(\theta)$ with a low $m > 2$ behave in a similar fashion, as the reader can confirm by the same methods. The only qualitative difference is that for $\delta \in \left(\frac{1}{4}, \frac{1}{2}\right)$ and a high enough m, it ceases to be true that $G_c^{*m}(0) \to \infty$ as $R \to \infty$.

5.3 General solution for y_1 and $y_1^{(P)}$

The evaluation of $y_1(\theta)$ and $y_1^{(P)}(\theta)$ can be performed using Fourier series. A square integrable function $f(\theta)$ on the circle may be decomposed as

$$f(\theta) = \sum_{n=-\infty}^{\infty} f_n e^{in\theta}, \tag{37}$$

where $i = \sqrt{-1}$ is the imaginary unit and the Fourier coefficients f_n are given by

$$f_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta. \tag{38}$$

In general, the notation used here for the nth Fourier coefficients will be to add subscript n to the symbol of the corresponding function. The convolution theorem for Fourier series states that for two functions f and g the Fourier coefficients $(f * g)_n$ of their (2π-periodic) convolution $f*g$ may be computed by multiplying the Fourier coefficients of the individual functions,

$$(f * g)_n = 2\pi f_n g_n.$$

The operator G acts according to (30) as a convolution with $\rho_L G(\theta)$, so

$$(Gf)_n = LG_n f_n.$$

Identical relations hold also for G_c, G_p, and G_g. In the last case, one should remember that G_g is only well defined when acting on functions orthogonal to the constant function y_0, i.e. on functions f whose zeroth Fourier coefficient f_0 vanishes.

We can now find an expression for the Fourier coefficients $y_{1,n}$ of the function $y_1(\theta)$. The zeroth coefficient $y_{1,0}$ vanishes since y_1 is chosen to be orthogonal to the constant function y_0. For nonzero n, applying the convolution theorem to the general expression (25) gives

$$\frac{y_{1,n}}{y_0} = -(1 + LG_{g,n})(1 - \sigma LG_{c,n})\tilde{g}_{c,n}.$$

This can be further simplified by two identities. The first identity, $LG_{g,n} = 1/(1 - LG_n) - 1$, comes from the definition (22), and the standard formula for the sum of a geometric series. The second identity, $LG_n = (1 + \sigma(\sigma - 1) LG_{c,n}) LG_{c,n}$, is a consequence of (11) and (13). Together they imply that $-(1 + LG_{g,n})(1 - \sigma LG_{c,n})\tilde{g}_{c,n} = -\frac{\tilde{g}_{c,n}}{1+(\sigma-1)LG_{c,n}}$. The conclusion is that

$$\frac{y_{1,n}}{y_0} = \begin{cases} 0 & \text{for } n = 0, \\ -\frac{\tilde{g}_{c,n}}{1+(\sigma-1)LG_{c,n}} & \text{for } n \neq 0. \end{cases} \tag{39}$$

For the local-price-index-adjusted GDP $y_1^{(P)}$ (28) leads to

$$\frac{y_{1,n}^{(P)}}{y_0^{(P)}} = \begin{cases} -\frac{\sigma}{\sigma-1}\hat{g}_{c,0} & \text{for } n = 0, \\ -\frac{1-\sigma L G_{c,n}}{1+(\sigma-1)LG_{c,n}}\tilde{g}_{c,n} - \frac{\sigma}{\sigma-1}\hat{g}_{c,n} & \text{for } n \neq 0. \end{cases} \tag{40}$$

If $b(\theta, \theta') = b(\theta', \theta)$, the Fourier coefficients are real and $\tilde{g}_{c,n} = \hat{g}_{c,n}$. In that case (40) simplifies to

$$\frac{y_{1,n}^{(P)}}{y_0^{(P)}} = \begin{cases} -\frac{\sigma}{\sigma-1}\tilde{g}_{c,0} & \text{for } n = 0, \\ -\frac{2\sigma-1}{\sigma-1}\frac{\tilde{g}_{c,n}}{1+(\sigma-1)LG_{c,n}} & \text{for } n \neq 0. \end{cases} \tag{41}$$

5.4 Fourier coefficients of $G_c(\theta)$ for specific functional forms of trade costs

The general formula (12) for $G_c(x, x')$ reduces in the case under consideration to

$$G_c(\theta, \theta') = G_c(\theta - \theta') = \frac{1}{\sigma\rho_L}\frac{T(\theta - \theta')}{\int_{-\pi}^{\pi} T(\theta'' - \theta')\, d\theta''} = \frac{1}{\sigma\rho_L}\frac{T(\theta - \theta')}{\int_{-\pi}^{\pi} T(\theta'')\, d\theta''}. \tag{42}$$

Here, of course, the $T(\theta - \theta')$ corresponds to the trade costs before the introduction of border costs, $T(\theta - \theta') = \tilde{\tau}^{1-\sigma}(\theta - \theta')$. The Fourier coefficients of $G_c(\theta)$ are

$$G_{c,n} = \frac{1}{\sigma L T_0} T_n.$$

Note that this implies that $G_{c,0} = 1/(\sigma L)$, and via (11) and (13) also that $G_{c,0} = 1/L$, as expected from (14).

Subsection (4.3) mentioned several convenient functional forms for transportation costs. They all have similar properties. For the purpose of finding analytic solutions to the Krugman model, we will focus mostly on one of them, namely $\tilde{\tau}(d) = \left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$, but the other ones can be treated similarly.[46] For the functional form of choice, $T(\theta)$ can be written as

$$T(\theta) = \left(\frac{1}{1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}}\right)^{\delta},$$

[46] A discussion of other asymptotically power-law trade costs will be included in a future version of the online appendix at http://www.people.fas.harvard.edu/~fabinger/papers.html

where the important parameter δ is defined as

$$\delta \equiv \frac{1}{2}\rho\left(\sigma - 1\right).$$

An alternative expression for $T\left(\theta\right)$ is

$$T\left(\theta\right) = Z^{2\delta}\left(\frac{1}{Z^2\cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}}\right)^{\delta} \tag{43}$$

with

$$Z^2 \equiv \frac{1}{1 + 4\alpha^2 R^2}$$

As shown in Appendix I, the Fourier coefficients of $T\left(\theta\right)$ are

$$T_n = \frac{Z^{\delta}\left(-1\right)^n}{\left(1-\delta\right)_n} P_{\delta-1}^{n}\left(\frac{1+Z^2}{2Z}\right).$$

$P_a^b\left(z\right)$ is the associated Legendre function.[47] The Pochhammer symbol $\left(a\right)_n$ is defined in terms of the gamma function as $\Gamma\left(a+n\right)/\Gamma\left(a\right)$, and should not be confused with the notation for Fourier coefficients. For positive integer n, this definition reduces to the nth order polynomial $\left(a\right)_n = a\left(a+1\right)\left(a+2\right)...\left(a+n-1\right)$. The resulting expression for $G_{c,n}$ is

$$G_{c,n} = \frac{1}{\sigma L}\frac{\left(-1\right)^n}{\left(1-\delta\right)_n}\frac{P_{\delta-1}^{n}\left(\frac{1+Z^2}{2Z}\right)}{P_{\delta-1}\left(\frac{1+Z^2}{2Z}\right)}. \tag{44}$$

6 The impact of border costs

6.1 General solution for GDP in the presence of border costs

Now consider the introduction of a small border cost. Let us split the circle into two 'countries', country A characterized by $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and country B by $\theta \in \left(-\pi, \frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right]$, separated by a border consisting of two points, $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. This assumption is made for

[47] Mathematica introduces three distinct definitions of associated Legendre functions. The function used here corresponds to the third definition, i.e. to LegendreP[ν,μ,3,z]. See Appendix A for a list of special functions and other mathematical notation.

simplicity, and generalization to different situations is straightforward. The trade costs are now $\tau(\theta,\theta') = \tilde{\tau}(d)\,\tilde{b}(\theta,\theta')$, with

$$\tilde{b}(\theta,\theta') \equiv 1 + \tilde{\kappa} 1_{C_A}(\theta)\, 1_{C_B}(\theta') + \tilde{\kappa} 1_{C_B}(\theta)\, 1_{C_A}(\theta'),$$

where 1_{C_A} and 1_{C_B} are the country indicator functions. The small positive parameter $\tilde{\kappa}$ is related to the parameter κ considered in the general discussion by $\kappa \equiv 1-(1+\tilde{\kappa})^{1-\sigma}$. For small $\tilde{\kappa}$, this is roughly $\kappa \approx (\sigma-1)\,\tilde{\kappa}$. In terms of $T(\theta,\theta')$ the change associated with the introduction of the border cost is $T(\theta,\theta') \to (1-\kappa b(\theta,\theta'))\, T(\theta,\theta')$ with

$$b(\theta,\theta') \equiv 1_{C_A}(\theta)\, 1_{C_B}(\theta') + 1_{C_B}(\theta)\, 1_{C_A}(\theta').$$

The Fourier coefficients $\tilde{g}_{c,n}$ of the function $\tilde{g}_c(\theta)$ are given in Appendix H,

$$\tilde{g}_{c,n} = \begin{cases} 0 & \text{for } n \text{ odd,} \\ \frac{1}{2\sigma}\delta_{0n} - \frac{4(-1)^{\frac{n}{2}}}{\pi^2}\sum_{m=0}^{\infty}\frac{LG_{c,2m+1}}{(2m+1)^2-n^2} & \text{for } n \text{ even.} \end{cases} \tag{45}$$

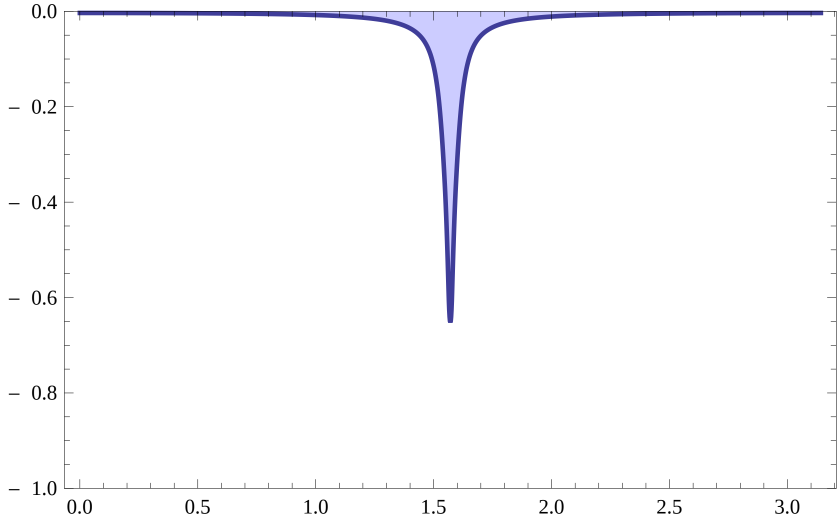
The result (39) then becomes

$$\frac{y_{1,n}}{y_0} = \begin{cases} 0 & \text{for } n \text{ odd or zero,} \\ \frac{4(-1)^{\frac{n}{2}}}{\pi^2}\frac{1}{1+(\sigma-1)LG_{c,n}}\sum_{m=0}^{\infty}\frac{LG_{c,2m+1}}{(2m+1)^2-n^2} & \text{for } n \text{ even and nonzero,} \end{cases}$$

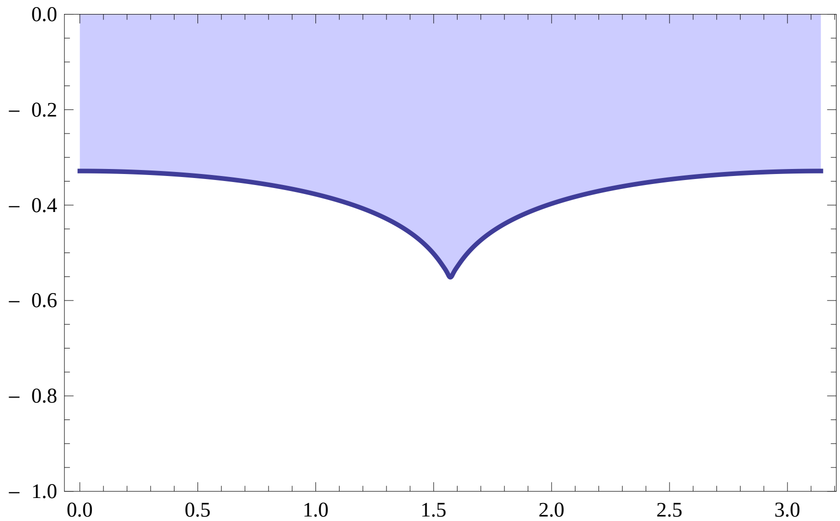
while (41) gives

$$\frac{y_{1,n}^{(P)}}{y_0} = \begin{cases} -\frac{1}{2}\frac{1}{\sigma-1} + \frac{4}{\pi^2}\frac{\sigma}{\sigma-1}\sum_{m=0}^{\infty}\frac{LG_{c,2m+1}}{(2m+1)^2} & \text{for } n \text{ zero,} \\ \frac{4(-1)^{\frac{n}{2}}}{\pi^2}\frac{2\sigma-1}{\sigma-1}\frac{1}{1+(\sigma-1)LG_{c,n}}\sum_{m=0}^{\infty}\frac{LG_{c,2m+1}}{(2m+1)^2-n^2} & \text{for } n \text{ even nonzero,} \\ 0 & \text{for } n \text{ odd.} \end{cases} \tag{46}$$

The resulting function $y_1^{(P)}(\theta)$ is plotted in Fig. 3 for different values of the parameter δ.



(a) $\delta = 1.25$



(b) δ=0.25

Figure 3: Plots of a measure of welfare changes, $(\sigma - 1)y_1^{(P)}(\theta)/y_0^{(P)}$, induced by an increase in the border costs between two semi-circular countries. Only half of each country is shown. Locations on the circle are parameterized by $\theta \in (-\pi, \pi]$. The border is located at $\theta = \pm\pi/2$, and $\theta = 0$ and $\theta = \pi$ correspond to the midpoints of the two countries. In part (a) δ is above the threshold of 1/2 while in part (b) it is below the threshold. Trade costs are $(1 + 4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$ with $\rho = 0.5$ and $\alpha R = 20$. In part (a) $\sigma = 6$, and in part (b) $\sigma = 2$.

6.2 Bounds on $y_1^{(P)}(0)/y_1^{(P)}\left(\frac{\pi}{2}\right)$

Using the explicit solution (45) for functional forms of the trade costs discussed in Subsection 4.3, one can derive[48] simple bounds on the values of y_1. In particular, (44) can be

[48] See the online appendix at http://www.people.fas.harvard.edu/fabinger/papers.html

used to show that for $\delta < 1/2$,

$$\lim_{R\to\infty} \frac{y_1^{(P)}(0)}{y_1^{(P)}\left(\frac{\pi}{2}\right)} \geq \frac{\sigma-1}{2\sigma-1}(1-2\delta), \tag{47}$$

while for $\delta > 1/2$,

$$\lim_{R\to\infty} \frac{y_1^{(P)}(0)}{y_1^{(P)}\left(\frac{\pi}{2}\right)} = 0. \tag{48}$$

In other words, there is a sharp change of behavior at $\delta = 1/2$ in the large-space limit. Above 1/2, locations in the middle of the country will not be impacted by the presence of the border at all. Below 1/2, the impact on the middle of the country will be comparable to that on the border region.

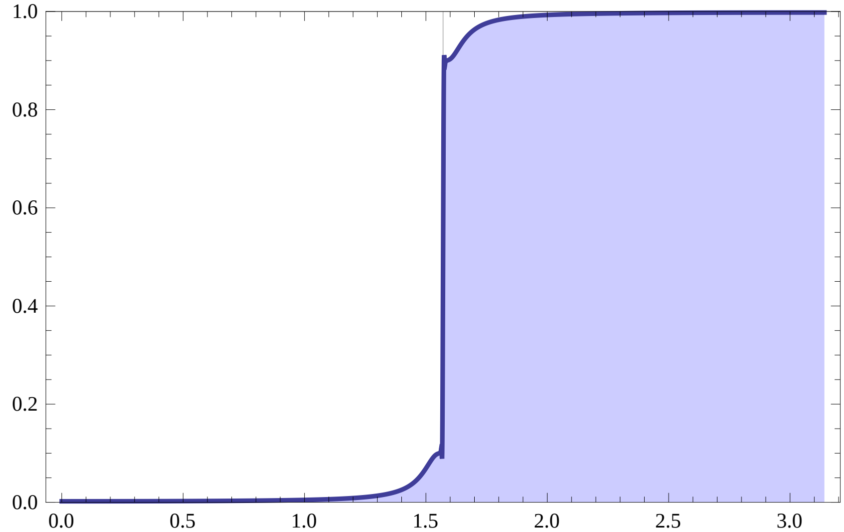
7 The impact of changes in productivity

Suppose that the productivity in a particular country changes. How are individual locations inside and outside of this country going to be affected? This question can be answered in a way very similar to the case of the border cost. If the country in question is large, one can consider the same spatial setup as for the border cost. There are two countries, A and B. Suppose that country B, represented by the 'southern' semicircle experiences a productivity increase. If productivity in B increases by a factor of $1+\tilde{\kappa}$, then this is equivalent to decreasing $\tau(x, x')$ from any location x in country B by the same factor. In terms of the function T, this corresponds to the change $T(x, x') \to (1 - \kappa b(x, x'))\, T(x, x')$ with $b(x, x') = -1_{C_B}(x)$ and $\kappa = 1 - (1+\tilde{\kappa})^{1-\sigma}$. Again, for small $\tilde{\kappa}$, $\kappa \approx (\sigma - 1)\tilde{\kappa}$. Now we can express the main quantities of interest in terms of $G_{c,n}$.

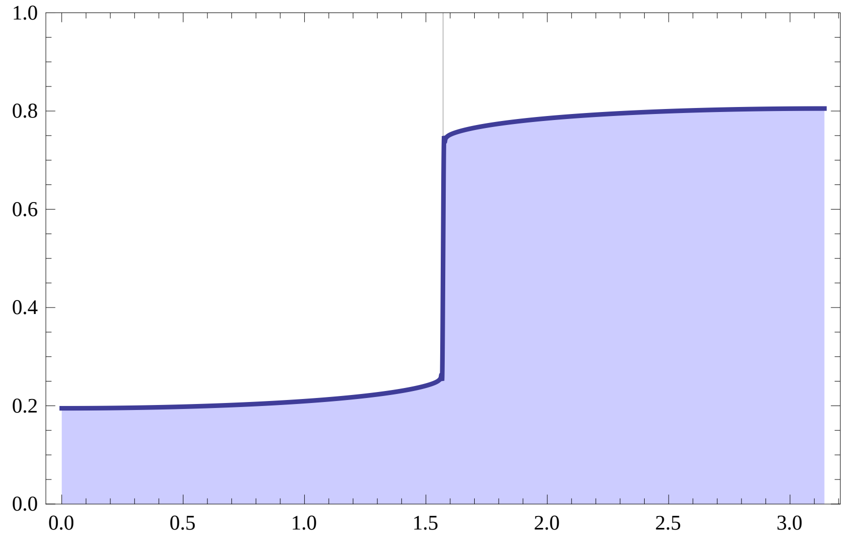
Evaluation of the Fourier coefficients of $\tilde{g}_c(\theta)$ is simple. Since

$$-\tilde{g}_c(\theta) = \rho_L 1_{C_B}(\theta) \int_{-\pi}^{\pi} G_c(\theta - \theta')\, d\theta' = \frac{1}{\sigma} 1_{C_B}(\theta),$$

they are proportional to the Fourier coefficients (70) of the indicator function of country



(a) $\delta = 1.25$



(b) δ=0.25

Figure 4: This is a productivity change counterpart of Fig. 3. It shows a measure of welfare changes, $(\sigma-1)y_1^{(P)}(\theta)/y_0^{(P)}$, induced by an increase of productivity in country B. Only half of each country is shown. The border is located at $\theta = \pm\pi/2$. As in Fig. 3, in part (a) δ is above the threshold of 1/2 while in part (b) it is below the threshold. Trade costs are $(1+4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$ with $\rho = 0.5$ and $\alpha R = 20$. In part (a) $\sigma = 6$, and in part (b) $\sigma = 2$.

B:

$$\tilde{g}_{c,n} = -\frac{1}{\sigma} 1_{C_B,n} = \begin{cases} -\frac{1}{2\sigma} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even and nonzero}, \\ -\frac{(-1)^{\frac{n+1}{2}}}{\pi\sigma n} & \text{for } n \text{ odd.} \end{cases}$$

Substituting these expressions into (39) gives

$$\frac{y_{1,n}}{y_0} = \begin{cases} 0 & \text{for } n \text{ even }, \\ \frac{(-1)^{\frac{n+1}{2}}}{\pi n} \frac{1}{1+(\sigma-1)LG_{c,n}} & \text{for } n \text{ odd.} \end{cases}$$

For the local-price-index-adjusted GDP, we should use the general formula (40) instead of (41), since $b(\theta, \theta')$ is not identically equal to $b(\theta', \theta)$. Remembering that $G_{c,0} = 1/(\sigma L)$,

$$\begin{aligned} -\hat{g}_c(\theta) &= \rho_L \int_{-\pi}^{-\frac{\pi}{2}} G_c(\theta - \theta')\, d\theta' + \rho_L \int_{\frac{\pi}{2}}^{\pi} G_c(\theta - \theta')\, d\theta' \\ &= \rho_L \sum_{n=-\infty}^{\infty} G_{c,n} e^{in\theta} \left(\int_{-\pi}^{-\frac{\pi}{2}} e^{-in\theta'} d\theta' + \int_{\frac{\pi}{2}}^{\pi} e^{-in\theta'} d\theta' \right) \\ &= \frac{1}{2\sigma} + \frac{1}{\pi} \sum_{n=-\infty,\ n \text{ odd}}^{\infty} \frac{(-1)^{\frac{n+1}{2}}}{n} LG_{c,n} e^{in\theta}. \end{aligned}$$

For the individual Fourier coefficients $\hat{g}_{c,n}$ this implies

$$\hat{g}_{c,n} = \begin{cases} -\frac{1}{2\sigma} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even nonzero}, \\ -\frac{(-1)^{\frac{n+1}{2}}}{\pi n} LG_{c,n} & \text{for } n \text{ odd.} \end{cases}$$

The formula (40) then yields

$$\frac{y_{1,n}^{(P)}}{y_0^{(P)}} = \begin{cases} -\frac{\sigma}{\sigma-1} \hat{g}_{c,0} & \text{for } n = 0, \\ -\frac{1-\sigma LG_{c,n}}{1+(\sigma-1)LG_{c,n}} \tilde{g}_{c,n} - \frac{\sigma}{\sigma-1} \hat{g}_{c,n} & \text{for } n \neq 0. \end{cases}$$

$$\frac{y_{1,n}^{(P)}}{y_0^{(P)}} = \begin{cases} \frac{1}{2} \frac{1}{\sigma-1} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even nonzero}, \\ \frac{(-1)^{\frac{n+1}{2}}}{\pi \sigma n} \left(\frac{1-\sigma LG_{c,n}}{1+(\sigma-1)LG_{c,n}} + \frac{\sigma^2}{\sigma-1} LG_{c,n} \right) & \text{for } n \text{ odd.} \end{cases}$$

See Fig. 4 for plots of $y_1^{(P)}$ for different values of the parameter δ. Again, there is a threshold behavior at $\delta = 1/2$. Bounds analogous to (47) and (48) will be included a future version of the online appendix.

8 The Krugman model on the sphere

8.1 The role of dimensionality

The previous section established that in the large-space limit, the qualitative properties of the Krugman model on the circle with asymptotically power-law trade costs change as $\delta \equiv \rho(\sigma - 1)/2$ crosses the threshold of 1/2. This value is not universal, however. For spaces of different dimensionality, the value of the threshold is different. In general, for a d_s-dimensional space, the threshold condition is[49]

$$\delta = \frac{d_s}{2}.$$

Clearly, it is of little economic interest to study cases with $d_s \geq 3$. The choice $d_s = 2$, however, is more appropriate for real-world economies than $d_s = 1$.

For this reason, the present section is devoted to the Krugman model on a two-dimensional spatial geometry, the sphere. It turns out that its properties closely resemble the case of the circle, apart form the fact that the role of δ is now played by $\delta/2$.

8.2 Basic setup

Let the spatial geometry be a sphere S of radius R parameterized by colatitude $\theta \in [0, \pi]$ and longitude $\varphi \in [0, 2\pi)$. Identify these coordinates with x introduced previously, $x = (\theta, \varphi)$. As in the case of the circle, labor density is chosen to be constant. The labor element is $dL(\theta, \varphi) = \rho_L \sin\theta d\theta d\varphi$ with $\rho_L = L/(4\pi)$. The endowment of labor per unit physical area equals $\rho_L R^2$. Again, the baseline solution corresponds to constant GDP density: $y_0(\theta, \varphi) = y_0$. The GDP propagator $G(x, x')$ depends only on the (rescaled) distance $\tilde{d}(x, x')$ between x and x' given by

$$\cos\tilde{d}(x, x') = \sin\theta\sin\theta' + \cos\theta\cos\theta'\sin(\varphi - \varphi').$$

[49] A hint that this may be the case comes from repeating the calculations that led to eq. (1). A careful analysis of geometries of arbitrary dimension provides a confirmation.

The information contained in $G(x, x')$ can be captured by a single-argument function G, defined by the relation $G\left(\tilde{d}(x, x')\right) = G(x, x')$. The action (30) of the operator G can be thought of as a spherical convolution with $\rho_L G\left(\tilde{d}(x, x')\right)$,

$$(Gf)(x) = \rho_L (G * f)(x) = \rho_L \int_S G\left(\tilde{d}(x, x')\right) f(x')\, dA(x').$$

Here $dA(x')$ is the (rescaled) area element at point $x' = (\theta', \varphi')$ and may be written as $dA(x') = \sin\theta' d\theta' d\varphi'$. A similar statement holds for G_c, G_p, and G_g and analogously defined functions $G_c(\tilde{d})$, $G_p(\tilde{d})$, and $G_g(\tilde{d})$. Again, it is worth remembering that the action of G_g is defined only on functions orthogonal to the constant function y_0.

Convolutions on the sphere are a little more subtle than convolutions on the circle. In the case of the circle there is a natural definition of convolution for arbitrary functions as long as the corresponding integral is convergent. On the sphere a natural definition of convolution exists only if at least one of the convolution factors is required to be rotationally symmetric, in the sense that it depends only on θ but not on φ. The functions $G(\tilde{d}), G_c(\tilde{d}), G_p(\tilde{d})$ and $G_g(\tilde{d})$ all satisfy this requirement, so this is not a source of any difficulty here. The general definition of spherical convolution is

$$(F * f)(x) = \int_S F\left(\tilde{d}(x, x')\right) f(x')\, dA(x'). \tag{49}$$

Here F is the function that only depends on θ, identified with $\tilde{d}$, and f may depend on both spherical coordinates of the point x'.

The spherical analogs of (31) and (32) take the same form,

$$y_1 = \sum_{n=0}^{\infty} \rho_L^n G^{*n} * (By_0) = \sum_{n=0}^{\infty} \rho_L^n \left(G_c + \sigma(\sigma - 1)\rho_L G_c^{*2}\right)^{*n} * (By_0).$$

The large R results for $G_c^{*2}(\theta)$ (and higher $G_c^{*m}(\theta)$) for the case of the circle have a direct analog here. To avoid repetition, detailed discussion is left for Appendix G. As mentioned earlier, the main lesson is that the role of δ (defined as $\rho(\sigma - 1)/2$) in the case of the circle is now played by $\delta/2$. Otherwise the qualitative behavior remains the same.

8.3 General solution for y_1 and $y_1^{(P)}$

A square integrable function $f(\theta,\varphi)$ on the sphere can be written as

$$f(\theta,\varphi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_l^m Y_l^m(\theta,\varphi), \tag{50}$$

for some coefficients f_l^m. These coefficients may be computed as

$$f_l^m=\int_S f(\theta,\varphi)\,Y_l^{m*}(\theta,\varphi)\sqrt{g(x)}dx=\int_0^{\pi}\int_0^{2\pi} f(\theta,\varphi)\,Y_l^{m*}(\theta,\varphi)\,d\varphi\sin\theta d\theta, \tag{51}$$

where the star denotes complex conjugation. The spherical harmonic function $Y_l^m(\theta,\varphi)$ of degree l and order m is defined as

$$Y_l^m(\theta,\varphi)=N_l^{|m|}P_l^{|m|}(\cos\theta)\,e^{im\varphi}.$$

$P_l^{|m|}$ is the associated Legendre polynomial of degree l and order $|m|$, $i=\sqrt{-1}$ is the imaginary unit, and $N_l^{|m|}$ is a positive normalization factor needed to make the system orthonormal (without the Condon-Shortley phase). The general convention for spherical harmonic coefficients of a function on the sphere is to add the indices l and m to the corresponding symbol of the function. When the index m is zero, it may be omitted. In other words, $f_l\equiv f_l^0$. All spherical harmonics needed here will be of order zero.[50] They are given more explicitly as

$$Y_l^0(\theta,\varphi)\equiv\frac{\sqrt{2l+1}}{\sqrt{4\pi}}P_l(\cos\theta), \tag{52}$$

where P_l is the Legendre polynomial of degree l. According to the convolution theorem on the sphere, spherical harmonic coefficients of the convolution (49) are equal to properly normalized products of the spherical harmonic coefficients of the individual convolution factors:

$$(F*f)_l^m=\frac{\sqrt{4\pi}}{\sqrt{2l+1}}F_l f_l^m, \tag{53}$$

[50] In other applications of the same framework, working with spherical harmonics of nonzero order may be necessary.

with $F_l \equiv F_l^0$. For the GDP propagator G this implies

$$(Gf)_l^m = \frac{1}{\sqrt{4\pi}} \frac{1}{\sqrt{2l+1}} L G_l^0 f_l^m,$$

and similarly for G_c, G_p and G_g. Following the same steps as in the case of the circle, we obtain

$$\frac{(y_1)_l^m}{y_0} = \begin{cases} 0 & \text{for } l = 0 \text{ or } m \neq 0, \\ -\frac{\tilde{g}_{c,l}}{1+(\sigma-1)\frac{1}{\sqrt{4\pi}}\frac{1}{\sqrt{2l+1}}LG_{c,l}} & \text{for } l > 0 \text{ and } m = 0, \end{cases} \tag{54}$$

$$\frac{\left(y_1^{(P)}\right)_l^m}{y_0^{(P)}} = \begin{cases} 0 & \text{for } m \neq 0, \\ -\frac{\sigma}{\sigma-1}\hat{g}_{c,0} & \text{for } l = 0 \text{ and } m = 0, \\ -\frac{\sqrt{4\pi}\sqrt{2l+1}-\sigma LG_{c,l}}{\sqrt{4\pi}\sqrt{2l+1}+(\sigma-1)LG_{c,l}}\tilde{g}_{c,l} - \frac{\sigma}{\sigma-1}\hat{g}_{c,l} & \text{for } l > 0 \text{ and } m = 0. \end{cases} \tag{55}$$

8.4 Solutions for specific functional forms of trade costs

Proceeding in analogy with the case of the circle,

$$G_c(\theta,\varphi,\theta',\varphi') = G_c\left(\tilde{d}(\theta,\varphi,\theta',\varphi')\right) = \frac{1}{\sigma\rho_L} \frac{T\left(\tilde{d}(\theta,\varphi,\theta',\varphi')\right)}{\int_0^\pi \int_0^{2\pi} T\left(\tilde{d}(\theta'',\varphi'',\theta',\varphi')\right)\sin\theta'' d\varphi'' d\theta''},$$

$$G_c(\theta) = \frac{\sqrt{4\pi}}{\sigma L T_0^0} T(\theta).$$

The spherical harmonic coefficients of $G_c(\theta)$ are

$$(G_c)_l^m = \frac{\sqrt{4\pi}}{\sigma L T_0^0} T_l^m.$$

Note that $G_0^0 = \sqrt{4\pi}/L$, which is consistent with (14). For transportation costs of the form $\tilde{\tau}(d) = \left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$,

$$T(\theta) = Z^{2\delta}\left(\frac{1}{Z^2\cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}}\right)^\delta,$$

with $\delta \equiv \rho\,(\sigma-1)\,/21$ and $Z^2 \equiv \left(1+4\alpha^2 R^2\right)^{-1}$. Because of rotational symmetry $T_l^m = 0$ and $(G_c)_l^m = 0$ for $m \neq 0$. As shown in Appendix K, the remaining coefficients are

$$T_l = 2\pi\sqrt{2l+1}\,(\delta)_l\,\frac{Z^{\frac{1}{2}+\delta}}{\sqrt{1-Z^2}}P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right),$$

$$G_{c,l} = \frac{\sqrt{4\pi}\,(\delta)_l}{\sigma L}\frac{P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right)}{P_{\delta-\frac{3}{2}}^{-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right)}.$$

8.5 The impact of changes in border costs

The spherical counterpart of the two semicircular 'countries' are the northern and the southern hemisphere, $C_A = \{(\theta,\varphi)\,|\theta \in [0,\frac{\pi}{2})\}$, and $C_B = \{(\theta,\varphi)\,|\theta \in (\frac{\pi}{2},\pi]\}$. The corresponding spherical harmonic coefficients of $\tilde{g}_c$ (and $\hat{g}_c$) are given by equation (75) in Appendix J. The values of the coefficients can be used in the expressions (54) and (75) for the change in GDP. The resulting solutions exhibit the threshold behavior at $\delta = 1$. Bounds analogous to (47) and (48) will be included in a future version of the online appendix.

8.6 The impact of changes in productivity

Similarly to the case of the circle, $\tilde{g}_c\,(\theta,\varphi)$ equals $-1_{C_B}\,(\theta,\varphi)\,/\sigma$. As a result, its spherical harmonic coefficients with non-zero m vanish and the others are proportional to (74):

$$\tilde{g}_{c,l} = \begin{cases} -\frac{1}{\sigma}\sqrt{\pi} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero}, \\ -\frac{1}{\sigma}\sqrt{\pi}\sqrt{2l+1}\frac{(-1)^{\frac{-l-1}{2}}}{2^l}\frac{(l-1)!}{\frac{l-1}{2}!\frac{l+1}{2}!} & \text{for } l \text{ odd}. \end{cases}$$

The formula (54) then gives

$$\frac{(y_1)_l^m}{y_0} = \begin{cases} 0 & \text{for } l \text{ even or } m \neq 0, \\ \frac{1}{\sigma}\frac{(-1)^{\frac{-l-1}{2}}}{2^l}\frac{(l-1)!}{\frac{l-1}{2}!\frac{l+1}{2}!}\frac{2\pi(2l+1)}{\sqrt{4\pi}\sqrt{2l+1}+(\sigma-1)LG_{c,l}} & \text{for } l \text{ odd and } m = 0. \end{cases}$$

Recognizing that $\hat{g}_c(\theta) = -\rho_L (G_c * 1_{C_B})(\theta)$ and applying the spherical convolution theorem (53) leads to

$$\hat{g}_{c,l} = \begin{cases} -\frac{1}{2\sigma} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero}, \\ -\frac{(-1)^{\frac{-l-1}{2}}}{2^{l+1}} \frac{(l-1)!}{\frac{l-1}{2}!\frac{l+1}{2}!} LG_{c,l} & \text{for } l \text{ odd}. \end{cases}$$

Of course, $(\hat{g}_c)_l^m$ with nonzero m vanish. The price-index-adjusted change in GDP (55) becomes

$$\frac{\left(y_1^{(P)}\right)_l^m}{y_0^{(P)}} = \begin{cases} \frac{1}{2}\frac{1}{\sigma-1} \text{ for } l = 0 \text{ and } m = 0, \\ 0 \text{ for } l \text{ even nonzero or } m \neq 0, \\ \frac{1}{\sigma}\frac{(-1)^{\frac{-l-1}{2}}}{2^l}\frac{(l-1)!}{\frac{l-1}{2}!\frac{l+1}{2}!}\left(\sqrt{\pi}\sqrt{2l+1}\frac{\sqrt{4\pi}\sqrt{2l+1}-\sigma LG_{c,l}}{\sqrt{4\pi}\sqrt{2l+1}+(\sigma-1)LG_{c,l}} + \frac{1}{2}\frac{\sigma^2}{\sigma-1}LG_{c,l}\right) \\ \text{for } l \text{ odd and } m = 0. \end{cases}$$

As in the case of border cost, the qualitative behavior changes at $\delta = 1$. For bounds analogous to (47) and (48), see a future version of the online appendix.

9 Higher-order terms

The first-order changes in GDP $y_1(x)$ capture the full impact of changes in trade costs when κ is very small. There is an analytic way to evaluate this impact even for larger κ. The goal of this section is to provide basic insight into how this can be achieved. A detailed discussion is left for a future version of the online appendix, because this issue lies outside of the main focus of the paper.

As in Section 4, suppose that there is an initial equilibrium with GDP equal to $y_0(x)$, and consider a change in trade costs. If the trade costs were characterized initially by $T(x, x')$ and after the change by $(1 - \kappa b(x, x'))T(x, x')$, the new GDP equation reads

$$y^\sigma(x) = \int \frac{(1 - \kappa b(x, x'))T(x, x')y(x')}{\int (1 - \kappa b(x'', x'))T(x'', x')y^{1-\sigma}(x'')dL(x'')} dL(x'). \tag{56}$$

The Taylor series of the new solution $y(x)$ is

$$y(x) = y_0(x) + \kappa y_1(x) + \kappa^2 y_2(x) + ...$$

The functions $y_n(x)$ are required to satisfy $\int y_n(x)\, y_0(x)\, dL(x) = 0,\ n > 0$. In Section 4 we saw that inserting the Taylor expansion into (56) and comparing terms linear in κ provides an equation that determines y_1 in terms of y_0. An analogous statement can be made for the higher-order terms as well. Denote

$$\Delta y(x) \equiv y(x) - y_0(x) = \kappa y_1(x) + \kappa^2 y_2(x) + ...$$

The GDP equation (56) can be expanded as[51]

$$\begin{aligned}
&\Delta y(x) - (G\Delta y)(x) \\
&= -(G\Delta y)(x) + y_0(x) \textstyle\sum_{m=1}^{\infty} \frac{1}{m!} \left(\frac{1}{\sigma}\right)_m y_0^{-m\sigma}(x) \\
&\Big\{ \textstyle\int T(x,x')\, y_0(x') \sum_{k=1}^{\infty} \frac{(-1)^k}{(N(x'))^{k+1}} dL(x') \times \\
&\left(\textstyle\int (1-\kappa b(x'',x'))\, T(x'',x')\, y_0^{1-\sigma}(x'') \sum_{j=0}^{\infty} \frac{(1-\sigma)_j}{j!} y_0^{-j}(x'')\, (\Delta y(x''))^j\, dL(x'') - N(x') \right)^k \\
&+ \textstyle\int T(x,x')\, (\Delta y(x') - \kappa b(x,x')\, y_0(x') - \kappa b(x,x')\, \Delta y(x')) \sum_{k=0}^{\infty} \frac{(-1)^k}{(N(x'))^{k+1}} dL(x') \times \\
&\left(\textstyle\int (1-\kappa b(x'',x'))\, T(x'',x')\, y_0^{1-\sigma}(x'') \sum_{j=0}^{\infty} \frac{(1-\sigma)_j}{j!} y_0^{-j}(x'')\, (\Delta y(x''))^j\, dL(x'') - N(x') \right)^k \Big\}^m,
\end{aligned} \tag{57}$$

where $N(x) \equiv \int y_0^{1-\sigma}(x')\, T(x',x)\, dL(x')$ and $(a)_n$ is the Pochhammer symbol. The terms $-(G\Delta y)(x)$ which are present on both sides would cancel, of course. However, when the equation is written in this way, it has an important property. On the left-hand side the term proportional to $\kappa^n, n > 0$ is simply $\kappa^n y_n(x) - \kappa^n (G y_n)(x)$. On the right-hand side, the term proportional to κ^n contains functions $y_0, y_1, ..., y_{n-1}$, but does not contain $y_n, y_{n+1}, y_{n+2}, ...$ This makes the equation useful: one can first solve for y_1 (as in the previous sections), then for y_2, y_3, etc. To be more precise, at each step the equation allows one to compute only the function $y_n - G y_n$ directly. To recover y_n itself, one can use the identity

[51] Note that the two different summations over k have different starting points.

$$y_n(x) = ((1 + G_g)(y_n - Gy_n))(x).$$

The method[52] just described provides a way to express any y_n in terms of y_0, T, and b. The resulting expressions may seem very complicated, but they are not. The value of y_n can be written an a sum of a finite number of expressions. These can be evaluated explicitly by the same techniques that were used to compute the first-order terms. Derivation of individual equations from (57), as well as the process of solving them, is greatly simplified by a diagrammatic technique[53] analogous to the method of Feynman diagrams in physics. This will be discussed in a future version of the online appendix.

10 Conclusion

Traditional models of international and intranational trade, as well as models introduced in the last decade, have some unexpected spatial properties. As we have seen, under standard assumptions used in the empirical literature, their behavior is highly sensitive to the precise values of their parameters. Naturally, such high sensitivity can lead to strong biases in various estimation procedures. This raises the question: to what extent are existing empirical results affected by such biases?

To address this issue, future empirical work can employ trade models based on familiar principles, but rich enough to include economic sectors with heterogeneous characteristics. The present paper provides a convenient way to study the properties of these models

[52]The mathematical insights underlying the calculation framework introduced here are most closely associated with Richard Feynman. He observed in 1940s – just like Ernst Stückelberg years earlier – that solutions to certain complicated physics problems can be obtained by evaluating series of terms, and that each one of these terms may be represented by a simple cartoon. These "Feynman diagrams" play two different roles. They ensure that one does not get lost in the algebra, and they provide an intuitive way of thinking about the mechanism the model in question represents.

Interestingly, another line of Feynman's thinking has already influenced other parts of economics; the Feynman-Kac formula is often used in financial economics and related fields. Although this is a mathematically related topic, the typical series of Feynman diagrams with multivalent vertices are not present there. This ultimately follows from the fact that the variables representing physical space here represent state space in the financial application of the Feynman-Kac formula.

[53]There are two versions of the diagramatic technique. The first one has important consequences primarily for y_n with $n > 1$. The second one is slightly more complicated, but provides insight into the structure of y_1.

analytically, without having to rely on individual numerical solutions, each generated for a single point in a large parameter space. The results of such analytic inquiry will lead to more appropriate model selection for empirical estimation.

Appendices

A Mathematical notation

$\Gamma(x)$	Gamma function
$\mathrm{B}(x,y) \equiv \Gamma(x)\Gamma(y)/\Gamma(x+y)$	Beta function
$\mathrm{B}_x(p,q)$	Incomplete beta function
$(a)_n \equiv \Gamma(a+n)/\Gamma(a)$	Pochhammer symbol
${}_pF_q(\alpha_1,...,\alpha_p;\beta_1,...,\beta_q;x)$	Generalized hypergeometric function
$F(\alpha,\beta;\gamma;x) \equiv {}_2F_1(\alpha,\beta;\gamma;z)$	Gauss hypergeometric function
${}_2\tilde{F}_1(\alpha,\beta;\gamma;x) \equiv {}_2F_1(\alpha,\beta;\gamma;x)/\Gamma(\gamma)$	Regularized hypergeometric function
$U(a,b,z)$	Confluent hypergeometric function of the second kind
$P_\nu(x)$	Legendre function; Legendre polynomial for $\nu \in \mathbb{N}_0$
$P_\nu^\mu(x)$	Associated Legendre function of the first kind, LegendreP$[\nu,\mu,3,x]$ in Mathematica notation
$Y_l^m(\theta,\varphi)$	Spherical harmonic function
$K_\nu(x)$	Modified Bessel function of the second kind
$G_{p,q}^{m,n}\left(x \,\middle\vert\, \begin{matrix} a_1,...a_p \\ b_1,...,b_q \end{matrix}\right)$	Meijer G-function
$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$	Wigner 3j-symbol
p.v.$\int$	Cauchy principal value integral
δ_{nm}	Kronecker delta, equal to 1 if $n=m$, and 0 otherwise
f_n	Fourier coefficient of function $f(\theta)$
f_l^m	Spherical harmonic coefficient of function $f(\theta,\varphi)$

B Neglecting changes in general equilibrium effects

Consider the Krugman (1980) model in the case of a completely symmetric circle, as in Fig. 1a. Solving for the equilibrium is simple because the GDP density will be the same everywhere. Now suppose we would like to know the response to changes in trade costs. To be concrete, let us split the circle into two semicircular 'countries', and introduce additional 'iceberg' type border costs, as in Anderson and van Wincoop (2003), i.e. as goods cross the border a certain fixed percentage is lost.

The consequences of this change in trade costs are illustrated in Fig. 5b, which captures all general equilibrium effects. At no location will the GDP increase when the border costs are introduced. If we decided to neglect general equilibrium feedback, in the sense of neglecting the first term on the right-hand-side of (5) or (18), the calculations would be much simpler, and we would get Fig. 5a. The results are quite different. In certain regions we would not get even the sign of the overall effect right.

C Remarks on methodology: reverse engineering equilibria from comparative statics

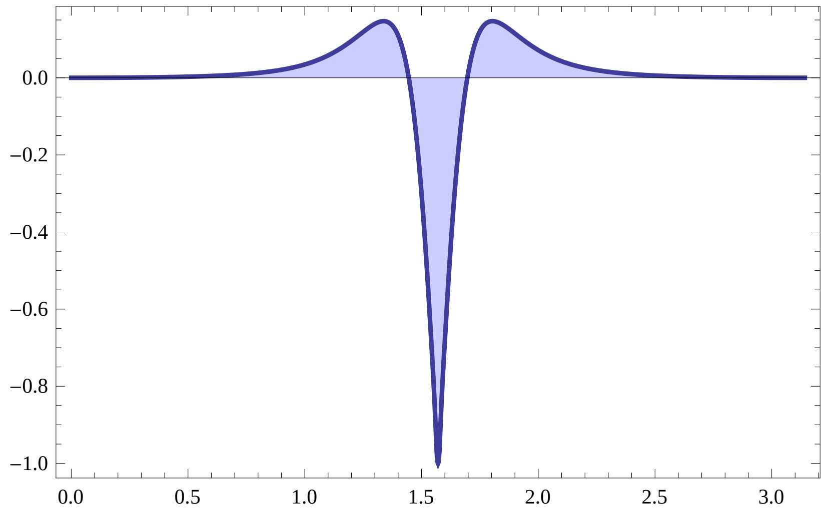
C.1 The case of a single endogenous variable

International trade models are fairly complicated. In order to make the general computational strategy employed in this paper easier to follow, this appendix illustrates some intuition used extensively in the main text with elementary examples, not necessarily coming from trade theory. Readers who find the rest of the paper intuitively clear may prefer to skip this discussion, as it does not contain any novel economic insights.

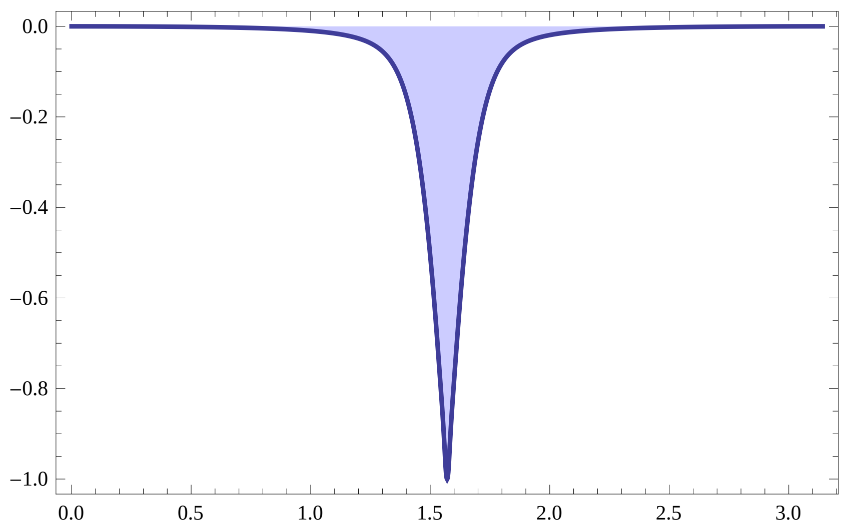
Consider an economic model in which the equilibrium value of an endogenous variable y is given implicitly as a function of an exogenous parameter κ by the equation

$$f(y,\kappa) = 0.$$

where the known function f satisfies the requirements of the implicit function theorem. For example, one can think of $f(y,\kappa) = 0$ as representing the first-order condition of a maximization problem. Suppose that we know the value y_0 corresponding to $\kappa = 0$, i.e. $f(y_0, 0) = 0$. Assume also that it is possible to compute all partial derivatives of f



(a)



(b)

Figure 5: The figure shows the first-order response of GDP to increased border costs. The spatial configuration is a circle parameterized by $\theta \in (-\pi, \pi]$, with only the range $[0, \pi]$ shown in the figure, which is sufficient due to the left-right symmetry. The circle is split into two semi-circular countries with the two border points located at $\theta = \pm\pi/2$. Part (a) plots the first-order change in GDP induced by increasing border costs as calculated ignoring general equilibrium feedback, while part (b) presents the full general equilibrium result. The parameter values used to generate the figure are $\sigma = 6$, $\rho = 0.75$, and $\alpha R = 5$, and the functional form of trade costs is $(1 + 4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$. For simple comparison, the y-axes are linearly transformed.

at $(y, \kappa) = (y_0, 0)$. It may be that the function $f(y, \kappa)$ is hard to invert with respect to its first argument. Under such circumstances, we can still recover the solution to the economic problem $y(\kappa)$ using comparative statics, assuming that $y(\kappa)$ is an analytic function.

First of all, since the first partial derivatives are known, we can use the approximation

$$y(\kappa) = y_0 + \left.\frac{dy}{d\kappa}\right|_{\kappa=0} \kappa + O\left(\kappa^2\right),$$

where the derivative may be computed as

$$\left.\frac{dy}{d\kappa}\right|_{\kappa=0} = -\frac{f_2(y_0, 0)}{f_1(y_0, 0)}. \tag{58}$$

This is what first-order comparative statics teaches us. But, of course, we can go further. With higher precision,

$$y(\kappa) = y_0 + \left.\frac{dy}{d\kappa}\right|_{\kappa=0} \kappa + \left.\frac{1}{2}\frac{d^2y}{d\kappa^2}\right|_{\kappa=0} \kappa^2 + O\left(\kappa^3\right).$$

The second derivative here can be obtained by the standard formula for second-order comparative statics,

$$\left.\frac{d^2y}{d\kappa^2}\right|_{\kappa=0} = -\left.\frac{f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22}}{f_1^3}\right|_{y=y_0,\kappa=0}. \tag{59}$$

In principle we can evaluate any derivative $d^n y/d\kappa^n$, and recover the full solution to the economic problem as the series

$$y(\kappa) = y_0 + y_1\kappa + y_2\kappa^2 + y_3\kappa^3 + \ldots$$

with

$$y_n \equiv \frac{1}{n!} \left.\frac{d^n y}{d\kappa^n}\right|_{\kappa=0}.$$

This is of course not as elegant as inverting $f(y, \kappa)$ with respect to its first argument directly, but it conveys the same information.

Obviously, we need a systematic way to express y_n in terms of partial derivatives of f. But that is not difficult. Substituting the Taylor expansion of $y(\kappa)$ for y into $f(y, \kappa) = 0$, we get

$$f\left(y_0 + y_1\kappa + y_2\kappa^2 + y_3\kappa^3 + ..., \kappa\right) = 0.$$

The Taylor series of the left hand side is

$$\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{f^{(j,k)}\left(y_0, 0\right)}{j!k!}\left(\sum_{l=1}^{\infty} y_l\kappa^l\right)^j \kappa^k,$$

where $f^{(j,k)}$ is the jth partial derivative with respect to the first argument of f and the kth partial derivative with respect to the second argument. The relation must hold for any κ, so for any non-negative integer n, the term proportional to κ^n must vanish. For $n = 0$, this implies $f\left(y_0, 0\right) = 0$, which is satisfied by assumption. For $n = 1$, the requirement becomes

$$f_1\left(y_0, 0\right) y_1\kappa + f_2\left(y_0, 0\right)\kappa = 0,$$

which is equivalent to the first-order comparative statics (58). For $n = 2$,

$$f_1\left(y_0, 0\right) y_2\kappa^2 + \frac{1}{2}f_{11}\left(y_0, 0\right) y_1^2\kappa^2 + f_{12}\left(y_0, 0\right) y_1\kappa^2 + \frac{1}{2}f_{22}\left(y_0, 0\right)\kappa^2 = 0,$$

leading to the second-order comparative statics formula (59). The value of y_3 can be obtained by looking at terms proportional to κ^3, etc.

Of course, in concrete applications of this method, it is very likely that most economic intuition is already contained in the first few terms, say y_0, y_1, and y_2. Only under rare circumstances would one need to compute even higher terms. This makes the present approach useful at the practical level.

C.2 The case of two endogenous variables and its generalization

The case of a single endogenous variable was straightforward. Now suppose that instead, y is a two-dimensional vector, $y \equiv \left(y_{(x_1)}, y_{(x_2)}\right)'$, where the labels x_1 and x_2 can be thought of as two distinct locations in space. Let the equations following from the model take the implicit form,

$$f\left(y, \kappa\right) = 0,$$

where $f(y,\kappa)$ is now a two-dimensional vector as well. Assume also that the first component of this equation can be solved with respect to $y_{(x_1)}$ and the second one with respect to $y_{(x_2)}$, i.e. that for some known functions $g_{(x_1)}$ and $g_{(x_2)}$, the equations

$$y_{(x_1)} = g_{(x_1)}\left(y_{(x_2)},\kappa\right),$$

$$y_{(x_2)} = g_{(x_2)}\left(y_{(x_1)},\kappa\right)$$

are equivalent to $f(y,\kappa)=0$. This assumption is made for expositional purposes only, and can be easily lifted. The task is again the same as in the single endogenous variable case. We are given the solution to these equations $y_0 = \left(y_{0(x_1)}, y_{0(x_2)}\right)'$ corresponding to $\kappa = 0$, and need to find y for non-zero κ, or at least the first-order change y_1 defined by $y = y_0 + y_1\kappa + O\left(\kappa^2\right)$.

One intuitive way to approach the problem is the following. Denote

$$v \equiv \begin{pmatrix} v_{(x_1)} \\ v_{(x_2)} \end{pmatrix} = \begin{pmatrix} \left.\frac{\partial g_{(x_1)}}{\partial\kappa}\right|_{y_{(x_2)}=y_{0(x_2)},\kappa=0} \\ \left.\frac{\partial g_{(x_2)}}{\partial\kappa}\right|_{y_{(x_1)}=y_{0(x_1)},\kappa=0} \end{pmatrix},$$

$$G \equiv \begin{pmatrix} G_{(x_1,x_1)} & G_{(x_1,x_2)} \\ G_{(x_2,x_1)} & G_{(x_2,x_2)} \end{pmatrix} \equiv \begin{pmatrix} 0 & \left.\frac{\partial g_{(x_1)}}{\partial y_{(x_2)}}\right|_{y_{(x_2)}=y_{0(x_2)},\kappa=0} \\ \left.\frac{\partial g_{(x_2)}}{\partial y_{(x_1)}}\right|_{y_{(x_1)}=y_{0(x_1)},\kappa=0} & 0 \end{pmatrix}.$$

If the exogenous parameter changes from 0 to some small value κ, it is natural to make the initial guess that $y_{(x_1)}$ changes from $y_{0(x_1)}$ to

$$y_{0(x_1)} + v_{(x_1)}\kappa,$$

and similarly $y_{(x_2)}$ becomes

$$y_{0(x_2)} + v_{(x_2)}\kappa.$$

But that cannot be the whole story. The fact that $y_{(x_2)}$ is now different will influence $y_{(x_1)}$ through the equation $y_{(x_1)} = g_{(x_1)}\left(y_{(x_2)},\kappa\right)$, and vice versa. So a better guess for $y_{(x_1)}$ and

$y_{(x_2)}$ is

$$y_{0(x_1)} + v_{(x_1)}\kappa + G_{(x_1,x_2)}v_{(x_2)}\kappa$$

and

$$y_{0(x_2)} + v_{(x_2)}\kappa + G_{(x_2,x_1)}v_{(x_1)}\kappa.$$

Repeating this logic indefinitely, we would get a candidate expression for $y_{(x_1)}$ and $y_{(x_2)}$ in the form of an infinite series. It can be succinctly written as

$$y = y_0 + \kappa \sum_{n=0}^{\infty} G^n v + O\left(\kappa^2\right). \tag{60}$$

This expression is of course correct, as can be seen by the standard method of comparative statics. Taking exact differentials of the equations of the problem, we obtain

$$dy_{(x_1)} = G_{(x_1,x_2)}dy_{(x_2)} + v_{(x_1)}d\kappa,$$

$$dy_{(x_2)} = G_{(x_2,x_1)}dy_{(x_1)} + v_{(x_2)}d\kappa,$$

and in matrix notation,

$$(1 - G)(y - y_0) = v\kappa + O\left(\kappa^2\right),$$

$$y = y_0 + \kappa (1 - G)^{-1} v + O\left(\kappa^2\right).$$

This is equivalent to the candidate expression above, thanks to the matrix geometric series identity $(1 - G)^{-1} = \sum_{n=0}^{\infty} G^n$.

We see that to succeed in this kind of task, one must be able to invert the matrix $1 - G$, or equivalently, to sum an infinite series of powers of G. In the two-variable case this is not a problem, of course. When the number of variables is large, however, this becomes a major obstacle.

There is a way to proceed, however. In situations where we can easily diagonalize G, computing $(1 - G)^{-1}$ is straightforward. Diagonalization of G means that we can express it as

$$G = C^{-1}DC,$$

where C is a known matrix and D is a known diagonal matrix with eigenvalues of G on its diagonal: $D = \text{diag}(d_1, d_2, ...)$. In this case

$$(1-G)^{-1} = C^{-1}(1-D)^{-1}C, \quad (1-D)^{-1} = \text{diag}\left(\frac{1}{1-d_1}, \frac{1}{1-d_2}, ...\right).$$

This strategy is extensively used in the main text. The action of the matrix C can be thought of as a change of basis in the vector space of infinitesimal changes in endogenous variables. In concrete examples it corresponds to either Fourier series expansion, or to spherical harmonic expansion. In those cases, G, C, and D are infinite-dimensional. One could also consider the case of discrete Fourier transform where the matrices would be finite-dimensional, but for brevity that case will be omitted.

D Derivation of equation (18)

Start with the GDP equation (17) with trade costs characterized by $(1-\kappa b(x,x'))\,T(x,x')$. Using the Taylor expansion $y(x) = y_0(x) + \kappa y_1(x) + O(\kappa^2)$ on the right-hand side of the GDP equation yields

$$\begin{aligned} y^{\sigma}(x) &= y_0^{\sigma}(x) - \kappa \int \frac{b(x,x')\,T(x,x')\,y_0(x')}{\int T(x'',x')\,y_0^{1-\sigma}(x'')\,dL(x'')} dL(x') \\ &+\kappa \int T(x,x')\,y_0(x') \frac{\int b(x'',x')\,T(x'',x')\,y_0^{1-\sigma}(x'')\,dL(x'')}{\left(\int T(x'',x')\,y_0^{1-\sigma}(x'')\,dL(x'')\right)^2} dL(x') \\ &+\sigma\kappa y_0^{\sigma-1}(x)\,(Gy_1)(x) + O\left(\kappa^2\right). \end{aligned}$$

Remembering the expression (12) for $G_c(x,x')$, this can be written as

$$\begin{aligned} y^{\sigma}(x) &= y_0^{\sigma}(x) - \kappa\sigma y_0^{\sigma-1}(x) \int b(x,x')\,G_c(x,x')\,y_0(x')\,dL(x') \\ &+\kappa\sigma^2 y_0^{\sigma-1}(x) \int G_c(x,x') \left(\int b(x'',x')\,G_c(x'',x')\,dL(x'')\right) y_0(x')\,dL(x') \\ &+\sigma y_0^{\sigma-1}(x)\,(Gy_1)(x) + O\left(\kappa^2\right). \end{aligned}$$

Taylor expanding the left hand side then leads to the final equation,

$$\begin{aligned} y_1(x) &= -\int b(x,x') G_c(x,x') y_0(x') dL(x') \\ &+\sigma \int G_c(x,x') \left(\int b(x'',x') G_c(x'',x') dL(x'') \right) y_0(x') dL(x') + (Gy_1)(x). \end{aligned}$$

Given the definition (19), this is equivalent to (18).

E Local-price-index-adjusted GDP

The welfare of individual agents is characterized by the local GDP per capita adjusted for the local price index, $y^{(P)}(x) \equiv y(x)/P(x)$, with

$$P(x) \equiv \frac{\sigma-1}{\sigma} \left(\frac{1}{\sigma F} \int (1-\kappa b(x',x)) T(x',x) y^{1-\sigma}(x') dL(x') \right)^{\frac{1}{1-\sigma}}.$$

The first-order Taylor expansion of $y^{(P)}(x)$ is

$$y^{(P)}(x) = y_0^{(P)}(x) + \kappa y_0^{(P)}(x) \frac{y_1(x)}{y_0(x)} + \frac{\kappa}{\sigma-1} y_0^{(P)}(x) \frac{N_1(x)}{N(x)} + O\left(\kappa^2\right),$$

where

$$\tilde{N}(x) \equiv \int (1-\kappa b(x',x)) T(x',x) y^{1-\sigma}(x') dL(x'),$$
$$N(x) \equiv \int T(x',x) y_0^{1-\sigma}(x') dL(x'),$$
$$N_1(x) \equiv \lim_{\kappa\to 0} \tfrac{1}{\kappa} \tfrac{\tilde{N}(x)-N(x)}{N(x)}.$$

Given the definition $y_1^{(P)}(x) \equiv \lim_{\kappa\to 0}\left(y^{(P)}(x) - y_0^{(P)}(x) \right)/\kappa$, this implies

$$\frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} = \frac{y_1(x)}{y_0(x)} + \frac{1}{\sigma-1} \frac{N_1(x)}{N(x)},$$

$$\frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} = \frac{y_1(x)}{y_0(x)} - \frac{\int T(x',x) y_0^{-\sigma}(x') y_1(x') dL(x')}{\int T(x',x) y_0^{1-\sigma}(x') dL(x')} - \frac{1}{\sigma-1} \frac{\int b(x',x) T(x',x) y_0^{1-\sigma}(x') dL(x')}{\int T(x',x) y_0^{1-\sigma}(x') dL(x')}.$$

Using the expression (12) for $G_c(x', x)$, this is

$$\frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} = \frac{y_1(x)}{y_0(x)} - \sigma \int G_c(x', x) \frac{y_1(x')}{y_0(x')} dL(x') - \frac{\sigma}{\sigma - 1} \hat{g}_c(x),$$

or in operator notation

$$\frac{y_1^{(P)}}{y_0^{(P)}} = (1 - \sigma G_c) \frac{y_1}{y_0} - \frac{\sigma}{\sigma - 1} \hat{g}_c.$$

The function $\hat{g}_c$ is defined in (26). If $G_c(x', x) = G_c(x, x')$ and $b(x', x) = b(x, x')$, then $\hat{g}_c = \tilde{g}_c$, and we can write

$$\frac{y_1^{(P)}}{y_0^{(P)}} = (1 - \sigma G_c) \frac{y_1}{y_0} - \frac{\sigma}{\sigma - 1} \tilde{g}_c.$$

F Properties of $G_c^{*2}(\theta)$ on the circle for large R

In the present context, the expression (12) for G_c becomes

$$G_c(\theta) = \frac{1}{\sigma L \bar{T}} T(\theta), \tag{61}$$

with the average $T(\theta)$ defined as $\bar{T} \equiv L^{-1} \int_{-\pi}^{\pi} T(\theta) \, dL(\theta)$. The function of interest, $G_c^{*2}(\theta)$, can be rewritten as

$$G_c^{*2}(\theta) \equiv \int_{-\pi}^{\pi} G_c(\theta') G_c(\theta - \theta') \, d\theta' = \frac{1}{\left(\sigma L \bar{T}\right)^2} T^{*2}(\theta) \equiv \frac{1}{\left(\sigma L \bar{T}\right)^2} \int_{-\pi}^{\pi} T(\theta') T(\theta - \theta') \, d\theta'.$$

The transportation costs $\tilde{\tau}(d)$ are asymptotically power-law, in the sense of satisfying condition (29). This implies that $T(\theta)$ is asymptotically power-law as well,

$$a_h^{1-\sigma} \hat{T}(\theta) \leq T(\theta) \leq a_l^{1-\sigma} \hat{T}(\theta), \tag{62}$$

with $\hat{T}(d) = \hat{\tau}^{1-\sigma}(d)$. Since

$$\int_{-\pi}^{\pi} \hat{T}(\theta) \, d\theta = 2 \int_0^{\frac{1}{\alpha R}} d\theta + 2 \int_{\frac{1}{\alpha R}}^{\pi} (\alpha R \theta)^{\rho(1-\sigma)} \, d\theta = \frac{2\delta}{2\delta - 1} \frac{2}{\alpha R} - \frac{2\pi}{2\delta - 1} \frac{1}{(\pi \alpha R)^{2\delta}},$$

a two sided bound on $\bar{T}$ immediately follows,

$$\frac{a_h^{1-\sigma}}{2\delta-1}\left(\frac{2\delta}{\pi\alpha R}-\frac{1}{(\pi\alpha R)^{2\delta}}\right)\leq\bar{T}\leq\frac{a_l^{1-\sigma}}{2\delta-1}\left(\frac{2\delta}{\pi\alpha R}-\frac{1}{(\pi\alpha R)^{2\delta}}\right). \tag{63}$$

Now that we know roughly the magnitude of $\bar{T}$, it remains to understand the nature of $T^{*2}(\theta)$. For this purpose, notice that (62) implies also

$$a_h^{1-\sigma}\left(T*\hat{T}\right)(\theta)\leq T^{*2}(\theta)\leq a_l^{1-\sigma}\left(T*\hat{T}\right)(\theta) \tag{64}$$

and

$$a_h^{2-2\sigma}\hat{T}^{*2}(\theta)\leq T^{*2}(\theta)\leq a_l^{2-2\sigma}\hat{T}^{*2}(\theta). \tag{65}$$

F.1 The peaks of $G_c^{*2}(\theta)$

Equation (61) implies

$$G_c^{*2}(0)=\frac{1}{\left(\sigma L\bar{T}\right)^2}T^{*2}(0).$$

The related function $\hat{T}^2(\theta)$ can be integrated explicitly,

$$\hat{T}^{*2}(0)\equiv\int_{-\pi}^{\pi}\hat{T}^2(\theta)\,d\theta=\int_0^{\frac{1}{\alpha R}}d\theta+2\int_{\frac{1}{\alpha R}}^{\pi}\frac{1}{(\alpha R\theta)^{4\delta}}d\theta=\frac{4\delta}{4\delta-1}\frac{2}{\alpha R}+\frac{1}{1-4\delta}\frac{2\pi}{(\pi\alpha R)^{4\delta}}.$$

In the final expression, the first term comes from $|\theta|$ of order $1/(\alpha R)$. The part of the integration domain responsible for the second term is characterized by $|\theta|$ being of order one. The first term is important for $\delta\in\left(\frac{1}{4},\infty\right)$. The second term dominates if $\delta\in\left(0,\frac{1}{4}\right)$. Combining these two equalities with (65) and (63) gives

$$\frac{(2\delta-1)^2\,4\pi}{(4\delta-1)(\sigma L)^2}\frac{\frac{2\delta}{\pi\alpha R}-\frac{1}{(\pi\alpha R)^{4\delta}}}{\left(\frac{2\delta}{\pi\alpha R}-\frac{1}{(\pi\alpha R)^{2\delta}}\right)^2}\frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}}\leq G_c^{*2}(0)\leq\frac{(2\delta-1)^2\,4\pi}{(4\delta-1)(\sigma L)^2}\frac{\frac{2\delta}{\pi\alpha R}-\frac{1}{(\pi\alpha R)^{4\delta}}}{\left(\frac{2\delta}{\pi\alpha R}-\frac{1}{(\pi\alpha R)^{2\delta}}\right)^2}\frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

Alternatively, using also (61) and (62) with $\theta = 0$, the same relations imply

$$\frac{2\delta - 1}{\sigma L} \frac{\frac{2\delta}{\pi\alpha R} - \frac{1}{(\pi\alpha R)^{4\delta}}}{\frac{2\delta}{\pi\alpha R} - \frac{1}{(\pi\alpha R)^{2\delta}}} G_c(0) \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(0) \leq \frac{2\delta - 1}{\sigma L} \frac{\frac{2\delta}{\pi\alpha R} - \frac{1}{(\pi\alpha R)^{4\delta}}}{\frac{2\delta}{\pi\alpha R} - \frac{1}{(\pi\alpha R)^{2\delta}}} G_c(0) \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

Specializing to various ranges for δ and remembering that R is large, the last two sets of inequalities imply (33) and (34).

F.2 $\delta < \frac{1}{2}$, tails of $G_c^{*2}(\theta)$

Consider θ greater than $2/(\alpha R)$. For simplicity, assume also that it is smaller than $\pi - 1/(\alpha R)$. Then the definition of $\hat{T}^{*2}(\theta)$ gives

$$\begin{aligned} \hat{T}^{*2}(\theta) &= \frac{1}{(\alpha R)^{4\delta}} \int_{-\pi}^{-\frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' + \frac{1}{(\alpha R)^{2\delta}} \int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' \\ &+ \frac{1}{(\alpha R)^{4\delta}} \int_{\frac{1}{\alpha R}}^{\theta - \frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' + \frac{1}{(\alpha R)^{2\delta}} \int_{\theta - \frac{1}{\alpha R}}^{\theta + \frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} d\theta' \\ &+ \frac{1}{(\alpha R)^{4\delta}} \int_{\theta + \frac{1}{\alpha R}}^{\pi} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta'. \end{aligned}$$

It is easy to see that since R is large, the second and the fourth term give a contribution that is negligible relative to the remaining terms.

$$\begin{aligned} (\alpha R)^{4\delta} \hat{T}^{*2}(\theta) &\approx \int_{-\pi}^{-\frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' \\ &\quad + \int_{\frac{1}{\alpha R}}^{\theta - \frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' + \int_{\theta + \frac{1}{\alpha R}}^{\pi} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta'. \end{aligned}$$

Similarly, the remaining integrals will not change much if in their limits $1/(\alpha R)$ is replaced by zero. In that case, the three integrals can be merged into one.

$$(\alpha R)^{4\delta} \hat{T}^{*2}(\theta) \approx \int_{-\pi}^{\pi} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' \equiv I(\theta). \tag{66}$$

The integral[54] $I(\theta)$ is independent of R. The last relation, together with (61), (65), and (63), gives

$$\frac{(1-2\delta)^2}{(\sigma L)^2}\pi^{4\delta}I(\theta)\frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \lesssim G_c^{*2}(\theta) \lesssim \frac{(1-2\delta)^2}{(\sigma L)^2}\pi^{4\delta}I(\theta)\frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

This results, in turn, implies the first lines of (35) and (36).

F.3 $\delta > \frac{1}{2}$, tails of $G_c^{*2}(\theta)$

The definition of $\left(T*\hat{T}\right)(\theta)$ is

$$\left(T*\hat{T}\right)(\theta) = \int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} T(\theta-\theta')\,d\theta' + \int_{\frac{1}{\alpha R}}^{\pi}\frac{T(\theta-\theta')}{|\alpha R\theta'|^{2\delta}}d\theta' + \int_{-\pi}^{-\frac{1}{\alpha R}}\frac{T(\theta-\theta')}{|\alpha R\theta'|^{2\delta}}d\theta'.$$

Assuming for simplicity that T is differentiable (this assumption can be lifted at the cost of a longer explanation), and integrating by parts, we get

$$\begin{aligned}
\left(T*\hat{T}\right)(\theta) \quad &= \quad \int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} T(\theta-\theta')\,d\theta' \\
&\quad +\frac{1}{2\delta-1}\frac{1}{\alpha R}\left(T\left(\theta-\frac{1}{\alpha R}\right)+T\left(-\theta-\frac{1}{\alpha R}\right)\right) \\
&\quad -\frac{1}{(\pi\alpha R)^{2\delta}}\frac{\pi}{2\delta-1}\left(T(\theta-\pi)+T(-\theta-\pi)\right) \\
&\quad -\frac{1}{2\delta-1}\frac{1}{(\alpha R)^{2\delta}}\int_{\frac{1}{\alpha R}}^{\pi}|\theta'|^{1-2\delta}\left(T'(\theta-\theta')+T'(-\theta-\theta')\right)d\theta'.
\end{aligned}$$

Consider θ in absolute value much greater than $1/(\alpha R)$. In that case, T is slowly varying. We can neglect the last two terms because they go to zero faster than $1/R$. In the remaining terms, we can approximate

$$\int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} T(\theta-\theta')\,d\theta' \approx \frac{2}{\alpha R}T(\theta),\ T\left(\theta-\frac{1}{\alpha R}\right)\approx T\left(-\theta-\frac{1}{\alpha R}\right)\approx T(\theta),$$

[54] The integral can be expressed using the gamma function Γ and the incomplete beta function B as $(-1)^{2\delta}|\theta|^{1-2\delta}\left(B_{\frac{\pi}{|\theta|}}(1-2\delta,1-2\delta)-B_{-\frac{\pi}{|\theta|}}(1-2\delta,1-2\delta)\right)+\left((-1)^{2\delta}-1\right)\sqrt{\pi}2^{4\delta}\frac{\delta\Gamma(-2\delta)}{\Gamma\left(\frac{3}{2}-2\delta\right)}|\theta|^{1-4\delta}$.

leading to the result that

$$\left(T * \hat{T}\right)(\theta) \approx \frac{2\delta}{2\delta - 1}\frac{2}{\alpha R}T(\theta).$$

Inequality (64) then becomes

$$a_h^{1-\sigma}\frac{2\delta}{2\delta-1}\frac{2}{\alpha R}T(\theta) \lesssim T^{*2}(\theta) \lesssim a_l^{1-\sigma}\frac{2\delta}{2\delta-1}\frac{2}{\alpha R}T(\theta).$$

Using (61) and (63), the implication for $G_c^{*2}(\theta)$ is

$$\frac{1}{\sigma\rho_L}\left(\frac{a_l}{a_h}\right)^{\sigma-1} G_c(\theta) \lesssim G_c^{*2}(\theta) \lesssim \frac{1}{\sigma\rho_L}\left(\frac{a_h}{a_l}\right)^{\sigma-1} G_c(\theta). \tag{67}$$

This implies the second line of (36). Combining this with (61), (63), and (62) then also gives the second line of (35).

G Properties of $G_c^{*2}(\theta)$ on the sphere for large R

In analogy to the case of the circle, $G_c(\theta) = T(\theta)/\left(\sigma L\bar{T}\right)$, where $T(\theta)$ averaged over the sphere is $\bar{T} \equiv L^{-1}\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} T(\theta)\,dL(\theta,\varphi)$. The function $G_c^{*2}(\theta)$ is defined as

$$G_c^{*2}(\theta) \equiv \int_0^{\pi}\int_0^{2\pi} G_c(\theta')\,G_c(d(\theta,\varphi,\theta',\varphi'))\sin\theta' d\varphi' d\theta'.$$

Note that the right-hand side is independent of φ. It is again simple to find a bound on $\bar{T}$. An upper bound comes from (62), $\sin\theta \leq \theta$, and explicit integration.

$$\bar{T}a_l^{\sigma-1} \leq \frac{1}{2}\int_0^{\pi}\hat{T}(\theta)\sin\theta d\theta \leq \frac{1}{2}\int_0^{\pi}\hat{T}(\theta)\,\theta d\theta = \frac{1}{2}\int_0^{\frac{1}{\alpha R}}\theta d\theta + \frac{1}{2}\frac{1}{(\alpha R)^{2\delta}}\int_{\frac{1}{\alpha R}}^{\pi}\theta^{1-2\delta}d\theta.$$

$$\bar{T} \leq \left(\frac{1}{4}\frac{\delta}{\delta-1}\frac{1}{(\alpha R)^2} + \frac{1}{4}\frac{1}{1-\delta}\frac{\pi^2}{(\pi\alpha R)^{2\delta}}\right)a_l^{1-\sigma}.$$

A lower bound can be obtained by direct analogy, this time using $\sin\theta \geq 2/\pi\theta,\ \theta \in \left[0, \frac{\pi}{2}\right]$ rather than $\sin\theta \leq \theta$.

$$\bar{T}a_h^{\sigma-1} \geq \frac{1}{2}\int_0^{\pi} \hat{T}(\theta)\sin\theta d\theta \geq \frac{1}{\pi}\int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{\pi}\frac{1}{(\alpha R)^{2\delta}}\int_{\frac{1}{\alpha R}}^{\frac{\pi}{2}} \theta^{1-2\delta} d\theta + \frac{1}{2}\frac{1}{(\alpha R)^{2\delta}}\int_{\frac{\pi}{2}}^{\pi} \theta^{-2\delta}\sin\theta d\theta.$$

Omitting the last term, which is positive, and evaluating the others yields

$$\bar{T}a_h^{\sigma-1} \geq \frac{1}{2\pi}\frac{1}{(\alpha R)^2} + \frac{1}{\pi}\frac{1}{(\alpha R)^{2\delta}}\frac{1}{2-2\delta}\left(\frac{\pi}{2}\right)^{2-2\delta} - \frac{1}{\pi}\frac{1}{(\alpha R)^{2\delta}}\frac{1}{2-2\delta}\left(\frac{1}{\alpha R}\right)^{2-2\delta},$$

$$\bar{T} \geq \left(\frac{\delta}{\delta-1}\frac{1}{2\pi(\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta}\frac{1}{(\pi\alpha R)^{2\delta}}\right)a_h^{1-\sigma}.$$

The reader can certainly derive stricter bounds, but to understand the dependence on R, these two are sufficient.

$$\left(\frac{\delta}{\delta-1}\frac{1}{2\pi(\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta}\frac{1}{(\pi\alpha R)^{2\delta}}\right)a_h^{1-\sigma} \leq \bar{T} \leq \left(\frac{1}{4}\frac{\delta}{\delta-1}\frac{1}{(\alpha R)^2} + \frac{1}{4}\frac{1}{1-\delta}\frac{\pi^2}{(\pi\alpha R)^{2\delta}}\right)a_l^{1-\sigma}. \tag{68}$$

G.1 The peaks of $G_c^{*2}(\theta)$

$$G_c^{*2}(0) = \frac{1}{(\sigma L\bar{T})^2}\int_0^{\pi}\int_0^{2\pi} T^2(\theta)\sin\theta d\varphi d\theta = \frac{2\pi}{(\sigma L\bar{T})^2}\int_0^{\pi} T^2(\theta)\sin\theta d\theta.$$

The upper bound is

$$\int_0^{\pi} \hat{T}^2(\theta)\sin\theta d\theta \leq \int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{(\alpha R)^{4\delta}}\int_{\frac{1}{\alpha R}}^{\pi} \theta^{1-4\delta} d\theta,$$

$$\int_0^{\pi} \hat{T}^2(\theta)\sin\theta d\theta \leq \frac{1}{2}\frac{1}{(\alpha R)^2} + \frac{1}{(\alpha R)^{4\delta}}\frac{1}{2-4\delta}\pi^{2-4\delta} - \frac{1}{(\alpha R)^{4\delta}}\frac{1}{2-4\delta}\frac{1}{(\alpha R)^{2-4\delta}},$$

$$\int_0^{\pi} \hat{T}^2(\theta)\sin\theta d\theta \leq \frac{\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{\pi^2}{2}\frac{1}{1-2\delta}\frac{1}{(\pi\alpha R)^{4\delta}}.$$

The lower bound is

$$\int_0^\pi \hat{T}^2(\theta)\sin\theta d\theta \geq \frac{2}{\pi}\int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{(\alpha R)^{4\delta}}\frac{2}{\pi}\int_{\frac{1}{\alpha R}}^{\frac{\pi}{2}} \theta^{1-4\delta} d\theta,$$

$$\int_0^\pi \hat{T}^2(\theta)\sin\theta d\theta \geq \frac{1}{\pi}\frac{1}{(\alpha R)^2} + \frac{1}{(\alpha R)^{4\delta}}\frac{1}{\pi}\frac{1}{1-2\delta}\left(\frac{\pi}{2}\right)^{2-4\delta} - \frac{1}{(\alpha R)^2}\frac{1}{1-2\delta}\frac{1}{\pi},$$

$$\int_0^\pi \hat{T}^2(\theta)\sin\theta d\theta \geq \frac{1}{\pi}\frac{2\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2}\pi^{1-4\delta}}{1-2\delta}\frac{1}{(\alpha R)^{4\delta}}.$$

The bounds combined:

$$\begin{aligned}\left(\frac{1}{\pi}\frac{2\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2}\pi^{1-4\delta}}{1-2\delta}\frac{1}{(\alpha R)^{4\delta}}\right) a_h^{2-2\sigma} \quad &\leq \quad \int_0^\pi T^2(\theta)\sin\theta d\theta \\ &\leq \quad \left(\frac{\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{\pi^2}{2}\frac{1}{1-2\delta}\frac{1}{(\pi\alpha R)^{4\delta}}\right) a_l^{2-2\sigma}.\end{aligned}$$

Combining this with (68) gives

$$\frac{2\pi}{(\sigma L)^2}\frac{\frac{1}{\pi}\frac{2\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2}\pi^{1-4\delta}}{1-2\delta}\frac{1}{(\alpha R)^{4\delta}}}{\left(\frac{1}{4}\frac{\delta}{\delta-1}\frac{1}{(\alpha R)^2} + \frac{1}{4}\frac{1}{1-\delta}\frac{\pi^2}{(\pi\alpha R)^{2\delta}}\right)^2}\frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(0) \leq \frac{2\pi}{(\sigma L)^2}\frac{\frac{\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{\pi^2}{2}\frac{1}{1-2\delta}\frac{1}{(\pi\alpha R)^{4\delta}}}{\left(\frac{\delta}{\delta-1}\frac{1}{2\pi(\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta}\frac{1}{(\pi\alpha R)^{2\delta}}\right)^2}\frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}},$$

or alternatively, also with (62)

$$\frac{2\pi G_c(0)}{\sigma L}\frac{\frac{1}{\pi}\frac{2\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2}\pi^{1-4\delta}}{1-2\delta}\frac{1}{(\alpha R)^{4\delta}}}{\frac{1}{4}\frac{\delta}{\delta-1}\frac{1}{(\alpha R)^2} + \frac{1}{4}\frac{1}{1-\delta}\frac{\pi^2}{(\pi\alpha R)^{2\delta}}}\frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(0) \leq \frac{2\pi G_c(0)}{\sigma L}\frac{\frac{\delta}{2\delta-1}\frac{1}{(\alpha R)^2} + \frac{\pi^2}{2}\frac{1}{1-2\delta}\frac{1}{(\pi\alpha R)^{4\delta}}}{\frac{\delta}{\delta-1}\frac{1}{2\pi(\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta}\frac{1}{(\pi\alpha R)^{2\delta}}}\frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

G.2 $\delta < 1$, tails of $G_c^{*2}(\theta)$

$$\hat{T}^2(\theta) \equiv \int_0^\pi \int_0^{2\pi} \hat{T}(\theta')\hat{T}(d(\theta,\varphi,\theta',\varphi'))\sin\theta' d\varphi' d\theta'.$$

This integral contains regions where either θ' or $d(\theta,\varphi,\theta',\varphi')$ are smaller than $1/(\alpha R)$. As in the case of the circle with $\delta < \frac{1}{2}$, they do not contribute much to the integral when $d(\theta,\varphi,\theta',\varphi') \gg 1/(\alpha R)$, and in this case can be safely ignored. As a result, we get the following approximation.

$$(\alpha R)^{4\delta}\,\hat{T}^2(\theta) \approx \int_0^{\pi}\int_0^{2\pi}\frac{1}{\theta'^{2\delta}}\frac{1}{d^{2\delta}(\theta,\varphi,\theta',\varphi')}\sin\theta' d\varphi' d\theta' \equiv I_{(2)}(\theta).$$

The right-hand side is now independent of R. Together with (68) this implies

$$\frac{2^5\pi^{4\delta-3}}{(\sigma L)^2}(1-\delta)\,I_{(2)}(\theta)\,\frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \le G_c^{*2}(\theta) \le \frac{2^5\pi^{4\delta-3}}{(\sigma L)^2}(1-\delta)\,I_{(2)}(\theta)\,\frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

G.3 $\delta > 1$, tails of $G_c^{*2}(\theta)$

$$\left(T*\hat{T}\right)(\theta) \equiv \int_0^{\pi}\int_0^{2\pi}\hat{T}(\theta')\,T(d(\theta,\varphi,\theta',\varphi'))\sin\theta' d\varphi' d\theta',$$

$$\begin{aligned}\left(T*\hat{T}\right)(\theta) &= \int_0^{\frac{1}{\alpha R}}\int_0^{2\pi}T(d(\theta,\varphi,\theta',\varphi'))\sin\theta' d\varphi' d\theta' \\ &\quad+\frac{1}{(\alpha R)^{2\delta}}\int_{\frac{1}{\alpha R}}^{\pi}\int_0^{2\pi}\frac{1}{\theta'^{2\delta}}T(d(\theta,\varphi,\theta',\varphi'))\sin\theta' d\varphi' d\theta'.\end{aligned}$$

For $\theta \gg 1/(\alpha R)$, T is slowly varying.

$$\begin{aligned}\left(T*\hat{T}\right)(\theta) &\approx \frac{\pi}{(\alpha R)^2}T(\theta) \\ &\quad+\frac{1}{(\alpha R)^{2\delta}}\int_{\frac{1}{\alpha R}}^{\pi}\int_0^{2\pi}\frac{1}{\theta'^{2\delta}}T(d(\theta,\varphi,\theta',\varphi'))\sin\theta' d\varphi' d\theta'.\end{aligned}$$

Using $\sin\theta < \theta$ and integrating by parts,

$$\begin{aligned}\left(T*\hat{T}\right)(\theta) &\le \frac{\pi}{(\alpha R)^2}T(\theta) \\ &\quad+\frac{1}{(\alpha R)^{2\delta}}\frac{1}{2-2\delta}\int_0^{2\pi}\left[\theta'^{2-2\delta}T(d(\theta,\varphi,\theta',\varphi'))\right]_{\frac{1}{\alpha R}}^{\pi}d\varphi' \\ &\quad-\frac{1}{(\alpha R)^{2\delta}}\frac{1}{2-2\delta}\int_{\frac{1}{\alpha R}}^{\pi}\int_0^{2\pi}\theta'^{2-2\delta}\frac{\partial T(d(\theta,\varphi,\theta',\varphi'))}{\partial\theta'}d\varphi' d\theta'.\end{aligned}$$

Since $\theta'^{2-2\delta}$ is small for $\theta' \gg 1/(\alpha R)$ and T varies slowly over the region where θ' is of order $1/(\alpha R)$, the last term can be neglected.

$$\begin{aligned}\left(T * \hat{T}\right)(\theta) \quad \leq \quad & \frac{\pi}{(\alpha R)^2} T(\theta) \\ & + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{1-\delta} \pi^{3-2\delta} T(\pi - \theta) \\ & - \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \int_0^{2\pi} \frac{1}{(\alpha R)^{1-2\delta}} T\left(d\left(\theta, \varphi, \frac{1}{\alpha R}, \varphi'\right)\right) \frac{1}{\alpha R} d\varphi'.\end{aligned}$$

In the last term, the dependence on φ' is very weak since θ' is set to $1/(\alpha R)$. We can approximate

$$\begin{aligned}\left(T * \hat{T}\right)(\theta) \quad \leq \quad & \frac{\pi}{(\alpha R)^2} T(\theta) \\ & + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{1-\delta} \pi^{3-2\delta} T(\pi - \theta) \\ & - \frac{1}{(\alpha R)^2} \frac{1}{1-\delta} \pi T(\theta).\end{aligned}$$

Since R is large

$$\left(T * \hat{T}\right)(\theta) \leq \frac{\pi}{(\alpha R)^2} \frac{\delta}{\delta - 1} T(\theta).$$

A lower bound can be obtained analogously using $\sin\theta < \frac{2}{\pi}\theta$.

$$\left(T * \hat{T}\right)(\theta) \geq \frac{\pi}{(\alpha R)^2} T(\theta) + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \frac{2}{\pi} \int_0^{2\pi} \left[\theta'^{2-2\delta} T\left(d\left(\theta, \varphi, \theta', \varphi'\right)\right)\right]_{\frac{1}{\alpha R}}^{\frac{\pi}{2}} d\varphi',$$

$$\left(T * \hat{T}\right)(\theta) \geq \left(\pi + \frac{2}{\delta - 1}\right) \frac{1}{(\alpha R)^2} T(\theta).$$

Together with (68), this leads to

$$\frac{2}{\sigma L} \frac{(\pi(\delta - 1) + 2)}{\pi\delta} G_c(\theta) \frac{a_l^{\sigma-1}}{a_h^{\sigma-1}} \leq G_c^{*2}(\theta) \leq \frac{2}{\sigma L} G_c(\theta) \frac{a_h^{\sigma-1}}{a_l^{\sigma-1}},$$

or alternatively, to

$$\frac{8}{(\sigma L)^2}\frac{\delta-1}{\delta}\frac{(\pi(\delta-1)+2)}{\pi\delta}\frac{1}{(\pi\alpha R)^{2\delta-2}}\frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(\pi) \leq \frac{4\pi}{(\sigma L)^2}\frac{\delta-1}{\delta}\frac{1}{(\pi\alpha R)^{2\delta-2}}\frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

H Fourier series expansions

H.1 Fourier series expansions of country indicator functions

In the case of the indicator function $1_{C_A}(\theta)$ of the set $C_A = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ characterizing country A, the standard formula for Fourier coefficients (38) specializes to

$$1_{C_A,n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} 1_{C_A}(\theta)\, e^{-in\theta}d\theta = \frac{1}{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-in\theta}d\theta.$$

Evaluating the integral in various cases,

$$1_{C_A,n} = \begin{cases} \frac{1}{2} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even and nonzero}, \\ \frac{(-1)^{\frac{n-1}{2}}}{\pi n} & \text{for } n \text{ odd}. \end{cases} \tag{69}$$

Now consider the indicator function $1_{C_B}(\theta)$ of country B with $C_B = \left(-\pi, -\frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right]$. Because almost everywhere $1_{C_A}(\theta) + 1_{C_B}(\theta) = 1$, it must be that $1_{C_A,0} + 1_{C_A,0} = 1$ and for nonzero n, $1_{C_A,n} + 1_{C_A,n} = 0$. This implies

$$1_{C_B,n} = \begin{cases} \frac{1}{2} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even and nonzero}, \\ \frac{(-1)^{\frac{n+1}{2}}}{\pi n} & \text{for } n \text{ odd}. \end{cases} \tag{70}$$

We see that the Fourier series expansions of the country indicator functions are

$$\begin{aligned} 1_{C_A}(\theta) &= \tfrac{1}{2} + \tfrac{1}{\pi}\textstyle\sum_{n\in\mathbb{Z},\ n \text{ odd}} (-1)^{\frac{n-1}{2}}\, \tfrac{1}{n}e^{in\theta}, \\ 1_{C_B}(\theta) &= \tfrac{1}{2} - \tfrac{1}{\pi}\textstyle\sum_{n\in\mathbb{Z},\ n \text{ odd}} (-1)^{\frac{n-1}{2}}\, \tfrac{1}{n}e^{in\theta}. \end{aligned} \tag{71}$$

For future convenience, multiply the expression for $1_{C_A}(\theta)$ by $e^{im\theta}$ and replace $n \to n-m$ to arrive at the following identity

$$1_{C_A}(\theta)\, e^{im\theta} = \frac{1}{2}e^{im\theta} + \frac{1}{\pi} \sum_{n\in\mathbb{Z},\ n-m \text{ odd}} (-1)^{\frac{n-m+1}{2}} \frac{1}{n-m} e^{in\theta}. \tag{72}$$

H.2 Fourier expansion of $\tilde{h}(\theta)$

For a symmetric function $H(\theta)$ on the circle (extended periodically to the real line), let us evaluate $\tilde{h}_n \equiv \left(\tilde{H}1\right)_n$ where $\tilde{H}$ is the integral operator with kernel $\rho_L \tilde{H}(d(\theta,\theta'))\, b(\theta,\theta')$. The function $b(\theta,\theta')$ is one whenever θ and θ' lie on opposite sides of the border and zero otherwise. It can be written as $b(\theta,\theta') = b_{AB}(\theta,\theta') + b_{BA}(\theta,\theta')$ with $b_{AB}(\theta,\theta') \equiv 1_{C_A}(\theta)\, 1_{C_B}(\theta')$ and $b_{BA}(\theta,\theta') = 1_{C_B}(\theta)\, 1_{C_A}(\theta')$. For ease of notation, define also $\tilde{h}_{AB} \equiv \tilde{H}_{AB}1$ with the kernel of the operator $\tilde{H}_{AB}$ being $\rho_L \tilde{H}(d(\theta,\theta'))\, b_{AB}(\theta,\theta')$, and analogously $\tilde{h}_{BA} \equiv \tilde{H}_{BA}1$. These two functions add up to $\tilde{h}$, so $\tilde{h}_n = \tilde{h}_{AB,n} + \tilde{h}_{BA,n}$.

First, compute $\tilde{h}_{AB,n}$.

$$\tilde{h}_{AB}(\theta) = \rho_L 1_{C_A}(\theta) \int_{-\pi}^{\pi} H(\theta-\theta')\, 1_{C_B}(\theta')\, d\theta'.$$

Fourier expanding the function $H(\theta-\theta')$,

$$\tilde{h}_{AB}(\theta) = \rho_L 1_{C_A}(\theta) \sum_{m\in\mathbb{Z}} H_m e^{im\theta} \int_{-\pi}^{\pi} e^{-im\theta'} 1_{C_B}(\theta')\, d\theta' = L \sum_{m\in\mathbb{Z}} H_m 1_{C_B,m} 1_{C_A}(\theta)\, e^{im\theta}.$$

Note that because $H(\theta)$ is symmetric, $H_m = H_{-m}$. Substituting for $1_{C_A}(\theta)\, e^{im\theta}$ from (72) gives

$$\tilde{h}_{AB}(\theta) = L \sum_{m\in\mathbb{Z}} A_m 1_{C_B,m} \left(\frac{1}{2}e^{im\theta} + \frac{1}{\pi} \sum_{n\in\mathbb{Z},\ n-m \text{ odd}} (-1)^{\frac{n-m+1}{2}} \frac{1}{n-m} e^{in\theta} \right).$$

Exchanging the order of summations,

$$\begin{aligned}\tilde{h}_{AB}(\theta) &= \frac{1}{2}L\sum_{m\in\mathbb{Z}} H_m 1_{C_B,m} e^{im\theta} \\ &+ \frac{1}{\pi}L\sum_{n\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z},\ m-n\ \mathrm{odd}} (-1)^{\frac{n-m+1}{2}}\frac{1}{n-m}H_m 1_{C_B,m}\right)e^{in\theta}.\end{aligned}$$

The Fourier series expansion $\tilde{h}_{BA}(\theta)$ follows from the one for $\tilde{h}_{AB}(\theta)$ because these two functions are related to each other by the shift $\theta \to \theta + \pi$,

$$\begin{aligned}\tilde{h}_{BA}(\theta) &= \frac{1}{2}L\sum_{m\in\mathbb{Z}} (-1)^m H_m 1_{C_B,m} e^{im\theta} \\ &+ \frac{1}{\pi}L\sum_{n\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z},\ m-n\ \mathrm{odd}} (-1)^{\frac{-n-m+1}{2}}\frac{1}{n-m}H_m 1_{C_B,m}\right)e^{in\theta}.\end{aligned}$$

According to (70), $1_{C_B,m}$ with even m is nonzero only for $m = 0$, in which case $1_{C_B,0} = \frac{1}{2}$. This means that after adding the two equations, we obtain

$$\begin{aligned}\tilde{h}(\theta) &= \frac{1}{2}LH_0 \\ &+ \frac{1}{\pi}L\sum_{n\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z},\ m-n\ \mathrm{odd}} \frac{(-1)^{\frac{n-m+1}{2}} + (-1)^{\frac{-n-m+1}{2}}}{n-m}H_m 1_{C_B,m}\right)e^{in\theta}.\end{aligned}$$

This is the desired Fourier expansion of $\tilde{h}(\theta)$. From here we can read off the individual Fourier coefficients.

$$\tilde{h}_n = \begin{cases} \frac{1}{2}LH_0\delta_{0n} + \frac{2}{\pi}L\sum_{m\in\mathbb{Z},\ m\ \mathrm{odd}} \frac{(-1)^{\frac{n-m+1}{2}}}{n-m}H_m 1_{C_B,m} & \text{for } n \text{ even,} \\ 0 & \text{for } n \text{ odd.}\end{cases}$$

Here δ_{0n} is the Kronecker delta, equal to one if $n = 0$ and zero otherwise. Now we can substitute the explicit expressions (70) for $1_{C_B,m}$ and use the relabeling

$$\sum_{m\in\mathbb{Z},\ m\ \mathrm{odd}} \frac{1}{n-m}\frac{H_m}{m} = \sum_{m\in\mathbb{Z},\ m\ \mathrm{odd\ positive}} \left(\frac{1}{n-m}\frac{H_m}{m} - \frac{1}{n+m}\frac{H_m}{m}\right) = -2\sum_{m=0}^{\infty}\frac{H_{2m+1}}{(2m+1)^2 - n^2}$$

to get the final expression

$$\tilde{h}_n = \begin{cases} 0 & \text{for } n \text{ odd,} \\ \frac{1}{2} L H_0 \delta_{0n} - \frac{4}{\pi^2} (-1)^{\frac{n}{2}} L \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} H_{2m+1} & \text{for } n \text{ even.} \end{cases}$$

H.3 Fourier expansion of $\tilde{g}_c(\theta)$

The discussion above was for an unspecified function $H(\theta)$ on the circle. Specializing to $G_c(\theta)$, we get the result

$$\tilde{g}_{c,n} = \begin{cases} 0 & \text{for } n \text{ odd,} \\ \frac{1}{2\sigma} \delta_{0n} - \frac{4}{\pi^2} (-1)^{\frac{n}{2}} L \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} G_{c,2m+1} & \text{for } n \text{ even.} \end{cases} \quad (73)$$

I Derivation of the expression for T_n on circle

The goal here is to evaluate the Fourier coefficients of

$$T(\theta) = \left(\frac{1}{1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}} \right)^{\delta}.$$

The standard formula (38) for Fourier coefficients implies

$$T_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-in\theta}}{\left(1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}\right)^{\delta}} d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\cos n\theta}{\left(1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}\right)^{\delta}} d\theta,$$

where the second equality follows from the Euler formula $e^{ix} = \cos x + i \sin x$ and from the fact that $T(\theta)$ is symmetric while $\sin n\theta$ is antisymmetric. Taking advantage of the symmetry of the final integrand to adjust the integration range and using the identity $\sin^2(\theta/2) = (1 - \cos\theta)/2$,

$$T_n = \frac{1}{\pi} \int_0^{\pi} \frac{\cos n\theta}{\left(1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 \cos\theta\right)^{\delta}} d\theta.$$

Define $Z \equiv 1/\sqrt{1+4\alpha^2 R^2}$. Then $2\alpha^2 R^2 = (1-Z^2)/(2Z^2)$, and the integral can be rewritten as

$$T_n = Z^\delta \frac{1}{\pi}\left(\frac{2Z}{1+Z^2}\right)^\delta \int_0^\pi \frac{\cos n\theta}{\left(1-\frac{1-Z^2}{1+Z^2}\cos\theta\right)^\delta} d\theta.$$

The (corrected[55] version) of second equation in paragraph 9.131 on p. 1008 of Gradshteyn and Ryzhik (2007) states that

$$P_\nu^m(z) = \frac{\nu(\nu-1)\ldots(\nu-m+1)}{\pi}\int_0^\pi \frac{\cos m\varphi}{\left(z-\sqrt{z^2-1}\cos\varphi\right)^{\nu+1}} d\varphi,$$

where $P_\nu^m(z)$ denotes[56] associated Legendre functions of the first kind. Using this equation with the replacement $\{m,\nu,\varphi\} \to \{n,\delta-1,\theta\}$, gives

$$\frac{(-1)^n}{(1-\delta)_n} P_{\delta-1}^n(z) = \frac{1}{\pi}\frac{1}{z^\delta}\int_0^\pi \frac{\cos n\theta}{\left(1-\frac{\sqrt{z^2-1}}{z}\cos\theta\right)^\delta} d\theta,$$

where $(1-\delta)_n$ is the Pochhammer symbol. Replacing also $z \to \frac{1+Z^2}{2Z}$ and noticing that this corresponds to $\frac{\sqrt{z^2-1}}{z} \to \frac{1-Z^2}{1+Z^2}$, one gets the identity

$$\frac{(-1)^n}{(1-\delta)_n} P_{\delta-1}^n\left(\frac{1+Z^2}{2Z}\right) = \frac{1}{\pi}\left(\frac{2Z}{1+Z^2}\right)^\delta \int_0^\pi \frac{\cos n\theta}{\left(1-\frac{1-Z^2}{1+Z^2}\cos\theta\right)^\delta} d\theta.$$

The integral on the right-hand side has the same form as the one in the expression for T_n, which leads to the conclusion

$$T_n = \frac{(-1)^n}{(1-\delta)_n} Z^\delta P_{\delta-1}^n\left(\frac{1+Z^2}{2Z}\right).$$

J Spherical harmonic expansions

J.1 Spherical harmonic expansions of country indicator func-

[55] The formula in the book contains an additional factor of $(-1)^m$, which is a typo.

[56] The Mathematica notation for this function is LegendreP[ν,μ,3,z].

tions

Let us find the spherical harmonic expansion of the indicator function $1_{C_A}(\theta)$ of the set $C_A = \left\{(\theta,\varphi)\,|\theta \in \left|0,\frac{\pi}{2}\right)\right\}$, which corresponds to country A. The general formula for spherical harmonic coefficients (51) gives

$$(1_{C_A})_l^m = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} Y_l^{m*}(\theta,\varphi)\, d\varphi \sin\theta d\theta.$$

This vanishes for non-zero m. For $m = 0$, we can use the expression (52) to simplify the integral to

$$(1_{C_A})_l^0 = \sqrt{\pi}\sqrt{2l+1}\int_0^{\frac{\pi}{2}} P_l(\cos\theta)\sin\theta d\theta = \sqrt{\pi}\sqrt{2l+1}\int_0^1 P_l(t)\, dt.$$

The last integral can be evaluated explicitly, with the result

$$(1_{C_A})_l^0 = \begin{cases} \sqrt{\pi} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero,} \\ \sqrt{\pi}\sqrt{2l+1}\frac{(-1)^{\frac{l-1}{2}}}{2^l}\frac{(l-1)!}{\frac{l-1}{2}!\frac{l+1}{2}!} & \text{for } l \text{ odd.} \end{cases}$$

Because up to a set of measure zero $1_{C_A}(\theta,\varphi) + 1_{C_B}(\theta,\varphi) = 1 = 2\sqrt{\pi}Y_0^0(\theta,\varphi)$, the spherical harmonic coefficients of 1_{C_B} with $C_B = \left\{(\theta,\varphi)\,|\theta \in \left(\frac{\pi}{2},\pi\right]\right\}$ follow. $(1_{C_B})_l^m$ with non-zero m vanishes, and

$$(1_{C_B})_l^0 = \begin{cases} \sqrt{\pi} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero,} \\ \sqrt{\pi}\sqrt{2l+1}\frac{(-1)^{\frac{-l-1}{2}}}{2^l}\frac{(l-1)!}{\frac{l-1}{2}!\frac{l+1}{2}!} & \text{for } l \text{ odd.} \end{cases} \tag{74}$$

J.2 Spherical harmonic expansion of $Y_{l'}^0(\theta,\varphi)\, C_B(\theta,\varphi)$ and $Y_{l'}^0(\theta,\varphi)\, C_A(\theta,\varphi)$

To find spherical harmonic coefficients of $Y_{l'}^0(\theta,\varphi)\, 1_{C_B}(\theta,\varphi)$ we may again use (51). The coefficients with nonzero m vanish, because $Y_0^0(\theta,\varphi)\, 1_{C_B}(\theta,\varphi)$ is independent of φ. For the remaining coefficients,

$$\left[Y_{l'}^0 1_{C_B}\right]_l^0 = 2\pi \int_0^\pi Y_{l'}^0(\theta,0)\, 1_{C_B}(\theta,0)\, Y_l^0(\theta,0) \sin\theta d\theta.$$

Due to (52) this is

$$\left[Y_{l'}^0 1_{C_B}\right]_l^0 = \frac{1}{2}\sqrt{2l'+1}\sqrt{2l+1}\int_0^1 P_{l'}(t)\, P_l(t)\, dt.$$

It is not hard to evaluate the integral for any given pair l, l' using the standard definition of Legendre polynomials. An alternative expression may be obtained as follows.

$$\left[Y_{l'}^0 1_{C_B}\right]_l^0 = 2\pi \sum_{l''=0}^{\infty} (1_{C_B})_{l''}^0 \int_0^\pi Y_{l'}^0(\theta,0)\, Y_{l''}^0(\theta,0)\, Y_l^0(\theta,0) \sin\theta d\theta.$$

$$\left[Y_{l'}^0 1_{C_B}\right]_l^0 = \frac{\sqrt{2l+1}\sqrt{2l'+1}}{\sqrt{4\pi}} \sum_{l''=0}^{\infty} (1_{C_B})_{l''}^0 \sqrt{2l''+1}\frac{1}{2}\int_0^\pi P_{l'}(\cos\theta)\, P_{l''}(\cos\theta)\, P_l(\cos\theta) \sin\theta d\theta.$$

$$\left[Y_{l'}^0 1_{C_B}\right]_l^0 = \frac{\sqrt{2l+1}\sqrt{2l'+1}}{\sqrt{4\pi}} \sum_{l''=0}^{\infty} (1_{C_B})_{l''}^0 \sqrt{2l''+1} \begin{pmatrix} l' & l'' & l \\ 0 & 0 & 0 \end{pmatrix}^2,$$

where $\begin{pmatrix} l' & l'' & l \\ 0 & 0 & 0 \end{pmatrix}$ is the Wigner 3j symbol (closely related to Clebsch–Gordan coefficients). Substituting the explicit expressions (74) for $(1_{C_B})_{l''}^0$ leads to

$$\begin{aligned}\left[Y_{l'}^0 1_{C_B}\right]_l^0 &= \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \begin{pmatrix} l' & 0 & l \\ 0 & 0 & 0 \end{pmatrix}^2 \\ &\quad + \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \sum_{l''=1,\ l''\ \text{odd}}^{\infty} \frac{(-1)^{\frac{-l''-1}{2}}(2l''+1)}{2^{l''}} \frac{(l''-1)!}{\frac{l''-1}{2}!\frac{l''+1}{2}!} \begin{pmatrix} l' & l'' & l \\ 0 & 0 & 0 \end{pmatrix}^2.\end{aligned}$$

Because the Wigner 3j symbol vanishes whenever the triangle inequality between l, l', and l'' is not satisfied, the infinite sum reduces to a finite one:

$$\left[Y_{l'}^0 1_{C_B}\right]_l^0 = \frac{1}{2}\delta_{ll'} + \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \sum_{l''=|l+l'|,\ l''\ \text{odd}}^{l+l'} \frac{(-1)^{\frac{-l''-1}{2}}(2l''+1)}{2^{l''}} \frac{(l''-1)!}{\frac{l''-1}{2}!\frac{l''+1}{2}!} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2.$$

Here $\delta_{ll'}$ is the Kronecker delta, equal to one when $l = l'$, and zero otherwise. Since up to a set of measure zero $1_{C_A}(\theta,\varphi) + 1_{C_B}(\theta,\varphi) = 1 = 2\sqrt{\pi}Y_0^0(\theta,\varphi)$, this result also implies

$$\left[Y_{l'}^0 1_{C_A}\right]_l^0 = \frac{1}{2}\delta_{ll'} + \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \sum_{l''=|l+l'|,\ l''\text{ odd}}^{l+l'} \frac{(-1)^{\frac{-l''+1}{2}}(2l''+1)}{2^{l''}} \frac{(l''-1)!}{\frac{l''-1}{2}!\frac{l''+1}{2}!} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2 .$$

J.3 Spherical harmonic expansions used to analyze the impact of border costs

Let us find certain spherical harmonic expansions needed to evaluate the impact of changes in border costs. The 'border indicator function' $b(x,x') \equiv b(\theta,\varphi,\theta',\varphi')$ may be decomposed into two parts

$$\begin{aligned} b(\theta,\varphi,\theta',\varphi') &= b_{AB}(\theta,\varphi,\theta',\varphi') + b_{BA}(\theta,\varphi,\theta',\varphi'), \\ b_{AB}(\theta,\varphi,\theta',\varphi') &\equiv 1_{C_A}(\theta,\varphi)\, 1_{C_B}(\theta',\varphi'), \\ b_{BA}(\theta,\varphi,\theta',\varphi') &\equiv 1_{C_B}(\theta,\varphi)\, 1_{C_A}(\theta',\varphi'). \end{aligned}$$

Consider a function $A(\theta,\varphi) \equiv A(\theta)$ on the on the sphere that is independent of φ. Denote the spherical angle (i.e. $1/R$ times the spherical distance) between points $x \equiv (\theta,\varphi)$ and $x' \equiv (\theta',\varphi')$ as $\tilde{d}(x,x') \equiv \tilde{d}(\theta,\varphi,\theta',\varphi')$. This angle can be computed with the help of the identity

$$\cos\tilde{d}(x,x') = \sin\theta\sin\theta' + \cos\theta\cos\theta'\sin(\varphi-\varphi').$$

The function whose spherical harmonic expansion we need to evaluate is $a(\theta,\varphi) \equiv a(\theta)$ defined[57] by the equation

$$a(\theta) \equiv \frac{1}{L}\int A\left(\tilde{d}(\theta,\varphi,\theta',\varphi')\right) b(\theta,\varphi,\theta',\varphi')\, dL(\theta',\varphi').$$

[57]Note that the integral on the right-hand side is independent of φ due to the rotational symmetry of each factor inside the integral.

It will be convenient to introduce also notation for its two parts corresponding to the decomposition of b in terms of b_{AB} and b_{BA}:

$$\begin{aligned} a_{AB}(\theta) &\equiv \tfrac{1}{L}\int A\left(\tilde{d}(\theta,\varphi,\theta',\varphi')\right) b_{AB}(\theta,\varphi,\theta',\varphi')\, dL(\theta',\varphi'), \\ a_{BA}(\theta) &\equiv \tfrac{1}{L}\int A\left(\tilde{d}(\theta,\varphi,\theta',\varphi')\right) b_{BA}(\theta,\varphi,\theta',\varphi')\, dL(\theta',\varphi').. \end{aligned}$$

Because $a(\theta,\varphi)$ is independent of φ, its spherical harmonic coefficients a_l^m with nonzero m vanish. The definition of may be rewritten as

$$a_{AB}(\theta) = \rho_L 1_{C_A}(\theta,\varphi)\int_0^{\pi}\int_0^{2\pi} A\left(d(\theta,\varphi,\theta',\varphi')\right) 1_{C_B}(\theta',\varphi')\sin\theta' d\varphi' d\theta'.$$

The integral on the right-hand side depends only on θ, the dependence on φ is trivial. To find its value, notice that it is equal to the spherical convolution $(A * 1_{C_B})(\theta,\varphi)$. With the help of the formula (53), its spherical harmonic coefficients are simply

$$(A * 1_{C_B})_l^0 = \frac{\sqrt{4\pi}}{\sqrt{2l+1}} A_l^0 (1_{C_B})_l^0,$$

which means that, according to (50),

$$\int_0^{\pi}\int_0^{2\pi} A\left(d(\theta,\varphi,\theta',\varphi')\right) 1_{C_B}(\theta',\varphi')\, d\varphi' d\theta' = \sum_{l=0}^{\infty}\frac{\sqrt{4\pi}}{\sqrt{2l+1}} A_l^0 (1_{C_B})_l^0\, Y_l^0(\theta,\varphi).$$

As a result, the expression for $a_{AB}(\theta)$ becomes

$$a_{AB}(\theta) = \rho_L \sum_{l=0}^{\infty}\frac{\sqrt{4\pi}}{\sqrt{2l+1}} A_l^0 (1_{C_B})_l^0\, Y_l^0(\theta,\varphi)\, 1_{C_A}(\theta,\varphi),$$

or equivalently,

$$a_{AB}(\theta) = \frac{L}{\sqrt{4\pi}}\sum_{l=0}^{\infty}\left(\sum_{l'=0}^{\infty}\frac{1}{\sqrt{2l'+1}} A_{l'}^0 (1_{C_B})_{l'}^0 \left[Y_{l'}^0 1_{C_A}\right]_l^0\right) Y_l^0(\theta,\varphi).$$

Analogously,

$$a_{BA}(\theta) = \frac{L}{\sqrt{4\pi}} \sum_{l=0}^{\infty} \left(\sum_{l'=0}^{\infty} \frac{1}{\sqrt{2l'+1}} A_{l'}^{0} \left(1_{C_A}\right)_{l'}^{0} \left[Y_{l'}^{0} 1_{C_B}\right]_{l}^{0} \right) Y_{l}^{0}(\theta, \varphi).$$

Adding the last two equations and comparing the result to (50) yields the following expression for the spherical harmonic coefficients of $a(\theta, \varphi)$:

$$a_{l}^{0} = \frac{L}{\sqrt{4\pi}} \sum_{l'=0}^{\infty} \frac{A_{l'}^{0}}{\sqrt{2l'+1}} \left(\left(1_{C_B}\right)_{l'}^{0} \left[Y_{l'}^{0} 1_{C_A}\right]_{l}^{0} + \left(1_{C_A}\right)_{l'}^{0} \left[Y_{l'}^{0} 1_{C_B}\right]_{l}^{0} \right).$$

The values of $\left(1_{C_A}\right)_{l'}^{0}$, $\left(1_{C_B}\right)_{l'}^{0}$, $\left[Y_{l'}^{0} 1_{C_A}\right]_{l}^{0}$, and $\left[Y_{l'}^{0} 1_{C_B}\right]_{l}^{0}$ were computed in earlier parts of this appendix.

J.4 Spherical harmonic expansion of $\tilde{g}_c(x)$

This result can be immediately applied (in the case of border costs) to the function $\tilde{g}_c(x) \equiv \int \tilde{G}_c(x, x')\, dL(x')$ defined in (23) as $\tilde{g}_c(x) \equiv \int \tilde{G}_c(x, x')\, dL(x')$:

$$\tilde{g}_{c,l} \equiv \left(\tilde{g}_c\right)_{l}^{0} = \frac{1}{\sqrt{4\pi}} \sum_{l'=0}^{\infty} \frac{G_{c,l'}}{\sqrt{2l'+1}} \left(\left(1_{C_B}\right)_{l'}^{0} \left[Y_{l'}^{0} 1_{C_A}\right]_{l}^{0} + \left(1_{C_A}\right)_{l'}^{0} \left[Y_{l'}^{0} 1_{C_B}\right]_{l}^{0} \right). \tag{75}$$

Of course, due to rotational symmetry, $\left(\tilde{g}_c\right)_{l}^{m} = 0$ for $m \neq 0$. Analogously to the case of the circle, $\left(\hat{g}_c\right)_{l}^{m} = \left(\tilde{g}_c\right)_{l}^{m}$ for any l and m.

K Derivation of the expression for T_l for the sphere

The spherical harmonic coefficients T_{l}^{m} of

$$T(\theta) = \left(\frac{1}{1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}} \right)^{\delta} = \left(\frac{1}{1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 \cos\theta} \right)^{\delta}$$

can be computed using (51). For nonzero m T_l^m vanishes because $T(\theta)$ is independent of φ. For zero m, write $T_l \equiv T_l^0$. In this case (51) and (52) give

$$T_l = \int_0^{\pi} \int_0^{2\pi} T(\theta) Y_l^{0*}(\theta, \varphi) \, d\varphi \sin\theta d\theta = \sqrt{\pi}\sqrt{2l+1} \int_0^{\pi} T(\theta) P_l(\cos\theta) \sin\theta d\theta.$$

Performing the substitution $t \equiv \cos\theta$ in the integral gives

$$T_l = \sqrt{\pi}\sqrt{2l+1} \int_0^{\pi} \frac{P_l(\cos\theta)\sin\theta d\theta}{\left(1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 \cos\theta\right)^{\delta}} = \sqrt{\pi}\sqrt{2l+1} \int_{-1}^{1} \frac{P_l(t)\, dt}{\left(1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 t\right)^{\delta}}.$$

As in the case of the circle, define $Z \equiv 1/\sqrt{1 + 4\alpha^2 R^2}$, which implies $2\alpha^2 R^2 = \frac{1-Z^2}{2Z^2}$.

$$T_l = \sqrt{\pi}\sqrt{2l+1} \left(\frac{2Z^2}{1-Z^2}\right)^{\delta} \int_{-1}^{1} \frac{P_l(t)}{\left(\frac{1+Z^2}{1-Z^2} - t\right)^{\delta}} dt.$$

The value of the integral can be found in Gradshteyn and Ryzhik (2007), where equation 7.228 on p. 791 states that

$$\frac{1}{2}\Gamma(1+\mu) \int_{-1}^{1} P_l(x)(z-x)^{-\mu-1}\, dx = \left(z^2 - 1\right)^{-\frac{\mu}{2}} e^{-i\pi\mu} Q_l^{\mu}(z).$$

With the replacement $\{\mu, z, x\} \to \left\{\delta - 1, \frac{1+Z^2}{1-Z^2}, t\right\}$ (which also means $z^2 - 1 \to \frac{4Z^2}{(1-Z^2)^2}$), this is

$$\int_{-1}^{1} \frac{P_l(t)}{\left(\frac{1+Z^2}{1-Z^2} - t\right)^{\delta}} dt = \frac{2}{\Gamma(\delta)} \left(\frac{1-Z^2}{2Z}\right)^{\delta-1} e^{-i\pi(\delta-1)} Q_l^{\delta-1}\left(\frac{1+Z^2}{1-Z^2}\right).$$

An alternative form of the right-hand side may be found using Gradshteyn and Ryzhik (2007), p. 959, eq. 8.703,

$$Q_{\nu}^{\mu}(z) = \frac{e^{\mu\pi i}\Gamma(\nu+\mu+1)\Gamma\left(\frac{1}{2}\right)}{2^{\nu+1}\Gamma\left(\nu+\frac{3}{2}\right)} \left(z^2-1\right)^{\frac{\mu}{2}} z^{-\nu-\mu-1} F\left(\frac{\nu+\mu+2}{2}, \frac{\nu+\mu+1}{2}, \nu+\frac{3}{2}; \frac{1}{z^2}\right).$$

Replacing $\{\mu, \nu, z\} \rightarrow \left\{\delta-1, l, \frac{1+Z^2}{1-Z^2}\right\}$ and noting that $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$,

$$\begin{aligned} Q_l^{\delta-1}\left(\frac{1+Z^2}{1-Z^2}\right) &= \frac{e^{(\delta-1)\pi i}\sqrt{\pi}\Gamma\left(l+\delta\right)}{2^{l+1}\Gamma\left(l+\frac{3}{2}\right)}\frac{\left(2Z\right)^{\delta-1}\left(1-Z^2\right)^{l+1}}{\left(1+Z^2\right)^{l+\delta}} \\ &\quad \times F\left(\frac{l+\delta+1}{2}, \frac{l+\delta}{2}, l+\frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right). \end{aligned}$$

As a result, the integral can be rewritten as

$$\int_{-1}^{1}\frac{P_l\left(t\right)dt}{\left(\frac{1+Z^2}{1-Z^2}-t\right)^{\delta}} = \frac{\sqrt{\pi}\Gamma\left(l+\delta\right)}{2^l\Gamma\left(\delta\right)\Gamma\left(l+\frac{3}{2}\right)}\left(\frac{1-Z^2}{1+Z^2}\right)^{l+\delta} F\left(\frac{l+\delta+1}{2}, \frac{l+\delta}{2}, l+\frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right),$$

and spherical harmonic coefficient T_l becomes

$$T_l = \frac{\pi\sqrt{2l+1}\Gamma\left(l+\delta\right)}{2^l\Gamma\left(\delta\right)\Gamma\left(l+\frac{3}{2}\right)}\frac{\left(2Z^2\right)^{\delta}\left(1-Z^2\right)^{l}}{\left(1+Z^2\right)^{l+\delta}}F\left(\frac{l+\delta+1}{2}, \frac{l+\delta}{2}, l+\frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right). \quad (76)$$

The Gauss hypergeometric function on the right-hand side may be further manipulated using several other identities. Gradshteyn and Ryzhik (2007), p. 1009, equation 9.134(2), reads

$$F\left(2\alpha, 2\alpha+1-\gamma, \gamma; z\right) = \left(1+z\right)^{-2\alpha} F\left(\alpha, \alpha+\frac{1}{2}, \gamma; \frac{4z}{\left(1+z\right)^2}\right).$$

Replacement $\{\alpha, \gamma, z\} \rightarrow \left\{\frac{l+\delta}{2}, l+\frac{3}{2}, \left(\frac{1-Z}{1+Z}\right)^2\right\}$ (which implies also $1+z \rightarrow 2\frac{1+Z^2}{(1+Z)^2}$ and $\frac{4z}{(1+z)^2} \rightarrow \left(\frac{1-Z^2}{1+Z^2}\right)^2$) leads to

$$\begin{aligned} F\left(l+\delta, \delta-\frac{1}{2}, l+\frac{3}{2}; \left(\frac{1-Z}{1+Z}\right)^2\right) &= \frac{\left(1+Z\right)^{2l+2\delta}}{2^{l+\delta}\left(1+Z^2\right)^{l+\delta}} \\ &\quad \times F\left(\frac{l+\delta}{2}, \frac{l+\delta+1}{2}, l+\frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right). \end{aligned} \quad (77)$$

Equation 9.131(1) on p. 1008 of Gradshteyn and Ryzhik (2007) states that

$$F\left(\alpha, \beta, \gamma; z\right) = \left(1-z\right)^{-\beta} F\left(\beta, \gamma-\alpha, \gamma; \frac{z}{z-1}\right).$$

Replacing $\{\alpha,\beta,\gamma,z\} \to \left\{l+\delta,\delta-\frac{1}{2},l+\frac{3}{2},\left(\frac{1-Z}{1+Z}\right)^2\right\}$ (and consequently $1-z \to \frac{4Z}{(1+Z)^2}$ and $\frac{z}{z-1} \to -\frac{(1-Z)^2}{4Z}$) gives

$$F\left(l+\delta,\delta-\frac{1}{2},l+\frac{3}{2};\left(\frac{1-Z}{1+Z}\right)^2\right) = \frac{(4Z)^{\frac{1}{2}-\delta}}{(1+Z)^{1-2\delta}} F\left(\delta-\frac{1}{2},\frac{3}{2}-\delta,l+\frac{3}{2};-\frac{(1-Z)^2}{4Z}\right). \tag{78}$$

Equation 8.702 on p. 959 of Gradshteyn and Ryzhik (2007) reads

$$P_\nu^\mu(z) = \frac{1}{\Gamma(1-\mu)}\left(\frac{z+1}{z-1}\right)^{\frac{\mu}{2}} F\left(-\nu,\nu+1,1-\mu;\frac{1-z}{2}\right).$$

Replacement $\{\mu,\nu,z\} \to \left\{-l-\frac{1}{2},\delta-\frac{3}{2},\frac{1+Z^2}{2Z}\right\}$ (and $\frac{z+1}{z-1} \to \left(\frac{1+Z}{1-Z}\right)^2$, $\frac{1-z}{2} \to -\frac{(1-Z)^2}{4Z}$) leads to

$$P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right) = \frac{1}{\Gamma\left(l+\frac{3}{2}\right)}\left(\frac{1-Z}{1+Z}\right)^{l+\frac{1}{2}} F\left(\frac{3}{2}-\delta,\delta-\frac{1}{2},l+\frac{3}{2};-\frac{(1-Z)^2}{4Z}\right). \tag{79}$$

The definition of the Gauss hypergeometric function (e.g. in Section 9.1 of Gradshteyn and Ryzhik (2007)) implies that the function is invariant under the exchange of its first two arguments. For this reason, (78) and (79) give

$$P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right) = \frac{(4Z)^{\delta-\frac{1}{2}}}{\Gamma\left(l+\frac{3}{2}\right)}\frac{(1-Z)^{l+\frac{1}{2}}}{(1+Z)^{l+2\delta-\frac{1}{2}}} F\left(l+\delta,\delta-\frac{1}{2},l+\frac{3}{2};\left(\frac{1-Z}{1+Z}\right)^2\right).$$

This can be combined with (77) to give

$$P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right) = \frac{(4Z)^{\delta-\frac{1}{2}}}{\Gamma\left(l+\frac{3}{2}\right)}\frac{(1-Z^2)^{l+\frac{1}{2}}}{2^{l+\delta}(1+Z^2)^{l+\delta}} F\left(\frac{l+\delta}{2},\frac{l+\delta+1}{2},l+\frac{3}{2};\left(\frac{1-Z^2}{1+Z^2}\right)^2\right).$$

Recalling that F is symmetric in its first two arguments and substituting the last equation into (76) leads to the final result

$$T_l = 2\pi\sqrt{2l+1}\,(\delta)_l \frac{Z^{\frac{1}{2}-\delta}}{\sqrt{1-Z^2}} P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right). \tag{80}$$

Here the Pochhammer symbol $(\delta)_l$ is defined as $\Gamma(l+\delta)/\Gamma(\delta) = \delta(\delta+1)...(\delta+l-1)$.

L Relation to fields in anti de Sitter space

The parameter threshold discussed in this paper has a counterpart in physics, namely the Breitenlohner and Freedman (1982a,b) bound that applies to fields in anti de Sitter space. The variables of the economic models with asymptotically power-law trade costs share one important property with fields in anti de Sitter space, namely the behavior of their propagators at long distances. The relevant comparison here is between a d_s-dimensional economic model and fields in a (d_s+1)-dimensional anti de Sitter space, which has a d_s-dimensional boundary where exogenous changes can be introduced.

Scalar fields in (d_s+1)-dimensional anti de Sitter space have propagators that at large distances d scale like $d^{-2\Delta}$ for a definite parameter Δ, which depends on their mass. The minimum mass-squared that the stability of the system allows is given by the Breitenlohner-Freedman value of $-d_s^2/\left(4R_{AdS}^2\right)$, where R_{AdS} is the curvature radius of the anti de Sitter space; see eq. (2.42) of Aharony et al. (2000). Due to eq. (3.14) of Aharony et al. (2000), this corresponds to $\Delta = d_s/2$.

In the economics situation of Section 4, the consumption part of the GDP propagator behaves at long distances like $d^{-2\delta}$, which means that δ can be thought of as the economics counterpart of Δ. Via this identification the physics relation $\Delta = d_s/2$ translates to the economics relation $\delta = d_s/2$, which is precisely the threshold where the qualitative behavior of the trade model changes.

Note that the explicit form of the propagator (3.42) of Aharony et al. (2000) is the same as the consumption part (12) of the GDP propagator when the trade costs are $\tilde{\tau} = \left(1+(\alpha d)^2\right)^{\rho/2}$. The same propagator (3.42) may be interpreted also from the point of view of the global anti de Sitter space, instead of the Poincaré coordinate patch perspective. When translated to the corresponding global anti de Sitter coordinates, the propagator acquires the same functional form as the consumption part (12) of the GDP propagator when the trade costs are $\tilde{\tau}(d) = \left(1+4\alpha^2 R^2 \sin^2\left(d/\left(2R\right)\right)\right)^{\rho/2}$.

References

Aharony, O., S.S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz, "Large N field theories, string theory and gravity," *Physics Reports*, 2000, *323* (3-4), 183–386.

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

Anderson, J.E., "A theoretical foundation for the gravity equation," *The American Economic Review*, 1979, *69* (1), 106–116.

_ and E. van Wincoop, "Gravity with gravitas: a solution to the border puzzle," *The American Economic Review*, 2003, *93* (1), 170–192.

_ and _ , "Trade costs," *Journal of Economic Literature*, 2004, *42*, 691–751.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare, "New trade models, same old gains?," *The American Economic Review*, forthcoming.

Armington, P.S., "A theory of demand for products distinguished by place of production," *IMF Staff papers*, 1969, *17*, 159–176.

Bak, P., K. Chen, J. Scheinkman, and M. Woodford, "Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics," *Ricerche Economiche*, 1993, *47* (1), 3–30.

Baldwin, R.E., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud, *Economic geography and public policy*, Princeton University Press, 2005.

Balistreri, E.J., R.H. Hillberry, and T.F. Rutherford, "Structural estimation and solution of international trade models with heterogeneous firms," *Journal of International Economics*, 2011, *83* (2), 95–108.

Behrens, K., C. Ertur, and W. Koch, "Dual gravity: using spatial econometrics to control for multilateral resistance," *Journal of Applied Econometrics*, 2007.

Bernard, A.B., J.B. Jensen, S.J. Redding, and P.K. Schott, "Firms in international trade," *The Journal of Economic Perspectives*, 2007, *21* (3), 105–130.

Breitenlohner, P. and D.Z. Freedman, "Positive energy in anti-de Sitter backgrounds and gauged extended supergravity," *Physics Letters B*, 1982, *115* (3), 197–201.

_ **and** _ , "Stability in gauged extended supergravity," *Annals of Physics*, 1982, *144* (2), 249–281.

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

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

_ **and** _ , *Technology and the global economy: a framework for quantitative analysis*, Princeton University Press, in progress. The manuscript is available online.

Engel, C. and J.H. Rogers, "How wide is the border?," *The American Economic Review*, 1996, *86* (5), 1112–1125.

Fabinger, M., "An analytic approach to spatial propagation of macroeconomic shocks," 2011. This work is currently in progress.

Fujita, M., P. Krugman, and A. Venables, *The spatial economy: cities, regions and international trade*, MIT Press, 1999.

Gopinath, G., P.O. Gourinchas, C.T. Hsieh, and N. Li, "International prices, costs and markup differences," *The American Economic Review*, forthcoming.

Gorodnichenko, Y. and L.L. Tesar, "Border effect or country effect? Seattle may not be so far from Vancouver after all," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 219–241.

Gradshteyn, I.S. and I.M. Ryzhik, *Table of integrals, series, and products*, seventh ed., Academic Press, 2007.

Helpman, E. and P.R. Krugman, *Market structure and foreign trade: increasing returns, imperfect competition, and the international economy*, The MIT Press, 1987.

_ , M.J. Melitz, and S.R. Yeaple, "Export versus FDI with heterogeneous firms," *The American Economic Review*, 2004, *94* (1), 300–316.

_ , _ , and Y. Rubinstein, "Estimating trade flows: trading partners and trading volumes," *The Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

Hillberry, R. and D. Hummels, "Trade responses to geographic frictions: a decomposition using micro-data," *European Economic Review*, 2008, *52* (3), 527–550.

Hummels, D., "Toward a geography of trade costs," 2001.

Kanwal, R.P., *Linear integral equations*, second ed., Birkhäuser, 1997.

Krugman, P., "Scale economies, product differentiation, and the pattern of trade," *The American Economic Review*, 1980, *70* (5), 950–959.

Martin, P. and C.A. Rogers, "Industrial location and public infrastructure," *Journal of International Economics*, 1995, *39* (3-4), 335–351.

Matsuyama, K., "Geography of the world economy," 1999.

McCallum, J., "National borders matter: Canada-US regional trade patterns," *The American Economic Review*, 1995, *85* (3), 615–623.

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

_ and G.I.P. Ottaviano, "Market size, trade, and productivity," *Review of Economic studies*, 2008, *75* (1), 295–316.

Okawa, Y. and E. van Wincoop, "Gravity in international finance," 2010.

Romalis, J., "Factor proportions and the structure of commodity trade," *American Economic Review*, 2004, pp. 67–97.

Rossi-Hansberg, E., "A spatial theory of trade," *American Economic Review*, 2005, *95* (5), 1464–1491.

Stokey, N.L., R.E. Lucas, and E.C. Prescott, *Recursive methods in economic dynamics*, Harvard University Press, 1989.