Inattentive Importers *

Kunal Dasgupta
University of Toronto

Jordi Mondria
University of Toronto

NOVEMBER 2017

Abstract

Information frictions prevent importers from observing the price of a good in every market. In this paper, we seek to explain how the presence of such frictions shapes the flow of goods between countries. To this end, we introduce rationally inattentive importers in a multi-country Ricardian trade model. The amount of information importers process is endogenous and reacts to changes in observable trade costs. Unlike traditional trade costs, changes in information processing costs have non-monotonic and asymmetric effects on bilateral trade flows. The model generates a novel prediction regarding the relationship between information processing costs and concentration of imports that finds support in the data. We calibrate the model, perform counterfactuals and show quantitatively how the response of trade flows to exogenous trade shocks gets magnified under inattention.

KEYWORDS : Rational inattention, information costs, magnification effect.
JEL Classification : D83, F10, F19, L15.

*We would like to thank participants at various conferences and workshops, and especially Victor Aguirregabiria, Treb Allen, Thomas Chaney, Gaelan MacKenzie, Peter Morrow, Colin Stewart and two anonymous referees for valuable suggestions. Kunal Dasgupta and Jordi Mondria gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC).

1 Introduction

Incomplete information plagues international commerce. Importers rarely observe the price and attributes of a good in every market. These informational barriers are bound to have an impact on the flow of goods between countries. Yet, despite a widespread agreement among economists that incomplete information could create significant barriers to trade, we lack a framework that formalizes the link between information and trade.[1]

In this paper, we attempt to develop such a framework. Our paper makes three main contributions. First, we formally derive a relationship between the probability of importing a product from a particular source country and the cost of processing information. A characteristic feature of our framework is that these probabilities not only respond directly to any change in model parameters, but also indirectly through a change in information processed. Second, we provide evidence that the import distribution for a product is, on average, less concentrated in countries with intermediate levels of information processing costs. We argue that this finding is consistent with our model of inattention, but is not predicted by standard full information models of trade. Third, we show quantitatively how, in the presence of information processing costs, a small increase in tariffs gets translated into a decline in imports that is significantly larger relative to a model without information frictions.

Specifically, we introduce rational inattention [Sims, 2003, 2006] into a multi-country, Ricardian model of trade. Every period, producers draw productivity stochastically. Importers would like to import a product from the country that has the lowest price. But importers have a limited capacity to process information about prices. Faced with a capacity constraint, importers must decide how much information to process about prices in each country. More information increases the precision of the noisy signals received by the importers, but comes at a higher cost. The rational importer weighs the marginal benefit of an extra unit of information against the marginal cost.

A key insight of our model is that the endogenous processing of information affects the response of trade flows to a change in observable trade costs between trading partners. When a trade cost, such as transport cost, between importing country j and exporting country i declines, country j importers start to purchase more from country i because the expected price offered by country i producers is now lower. This is the standard effect of trade costs on trade flows present in any trade model. Our model has an additional information effect. Faced with a cost of processing information, importers in country j choose how much information to process about every

[1] In their survey on trade costs, Anderson and van Wincoop [2004] highlight the need for more careful modelling of information frictions.

source country. A lower expected price in country i raises the expected benefit of processing information about country i. Country j importers respond by paying more attention to country i and less attention to every other country, thereby boosting the volume of trade between j and i further. Thus, when importers are rationally inattentive, small differences in observable trade costs can have large effects on trade flows – there is a *magnification effect*.

Following Matejka and McKay [2014], we show that the optimal solution of a rationally inattentive importer is to choose probabilistically the country from where to buy a given product, with this probability distribution following an adjusted multinomial logit. In the full information model of Eaton and Kortum [2002], while the prior probability that country j imports a product from country i is positive for every i, the corresponding posterior probability is either zero or one. In our model, however, the posterior probability is also positive for every i. This is because, even after productivity draws are realized, importers in country j do not perfectly observe prices and hence attach a positive probability to every country i having the lowest price. The implications are twofold. First, a country can buy the same product from different source countries. Second, a country can import and export the same product at the same time. Currently, such patterns in the data are rationalized by appealing to intra-industry trade.[2]

The key parameter in our model is the cost of processing information. We show that, unlike traditional trade costs, information costs may have non-monotonic effects on bilateral trade flows as the share of imports first rises but eventually declines when information costs increase. We also show that, unlike traditional trade costs, information costs may have asymmetric effects on bilateral trade flows. An increase in information costs may lead importers to choose to process more information about countries that have lower expected prices.[3] This will result in an increase in imports from these countries, to the detriment of countries that have higher expected price – it is *as if* the importing country has imposed import tariffs that are higher for countries that have higher expected price. A uniform increase in standard trade costs can not generate such an outcome.

Our model generates a novel prediction linking the concentration of imports with the cost of processing information. In the absence of information costs, importers purchase a product from one country only – the country offering the lowest price. Accordingly, the import distribution is degenerate. But when information costs are infinitely high, importers again purchase from one country only – the country with the lowest expected price. For intermediate values of information costs, importers diversify. In fact, we show that the concentration of the import distribution for a

[2] But see Allen [2014] for an exception.

[3] A lower expected price could arise either due to lower bilateral trade cost or higher average productivity.

given product and importing country exhibits a U-shape with respect to information costs, where concentration is measured using the Herfindahl-Hirschman index. We also note that none of the standard models of trade generate a systematic relation between the concentration of imports on the one hand, and importing country characteristics on the other. In the final part of the paper, we test this prediction. We postulate that countries differ in terms of their costs of processing information. We measure information costs with international bandwidth, which is a country-specific variable that determines the speed of data flow. To allow as much flexibility as possible, we carry out a non-parametric approach to examine the relationship between information costs and the concentration of imports. We show that the concentration of imports declines for small values of the information cost but rises for large values. This relationship is robust to a number of controls.

In the last part of the paper, we examine whether information costs matter quantitatively. In order to do so, we numerically solve two 25-country models – one with full information, and the other with inattentive importers. We calibrate the parameters of the two models such that they match the same set of moments. We then perform two counterfactuals: NAFTA termination and Brexit. In the first exercise, we raise import tariffs from 0 to 5 percent between the U.S., Canada and Mexico, the three members of the North American Free Trade Agreement (NAFTA). In the second exercise, we raise import tariffs between the U.K. and the other member countries of the European Union in our sample.

An increase in U.S. tariffs on imports from Canada and Mexico reduces import shares from those countries, while raising the own import share of the U.S. This effect is significantly larger when importers endogenously process information. As imports from its neighbours become more expensive on average, U.S. importers start paying less attention to those countries. This ends up magnifying the effect of the tariffs on trade flows. We get qualitatively similar results in the second counterfactual. As U.K. tariffs on imports from E.U. member countries rise, imports from Germany and France decline while U.K.'s own imports rise, with these effects being much larger when importers are rationally inattentive.

To the best of our knowledge, we are the first to apply the theory of rational inattention to the study of international trade. Our decision to model information as a theory of attention allocation is guided by the following consideration: attention is a major area of investigation in education, psychology and neuroscience, and its influence is growing in economics and finance. As suggested by Kahneman [1973], the human mind is bounded by cognitive limits and even if individuals had access to full information, their mind would be unable to process all the available information. Individuals would then have to choose how to allocate their limited cognitive attention resources to process information when making decisions. Hence, selectively focusing more cognitive resources on one option would result in a decrease of cognitive attention to alternative options. In the context of international trade, a consequence of rational inattention is that unlike most papers that deal with information frictions, importers in our model choose to process different amounts of information about product prices in different source countries.

In one of the first papers to highlight the role of information frictions in shaping international trade flows, Rauch [1999] provided evidence that proximity, common language and colonial ties are more important for trade in differentiated products, which are presumably more dependent on information, than for products traded on organized exchanges and those that have reference prices. Chaney [2014] incorporates exporter networks into a model of trade. Among other things, he shows that his network model can explain the distribution of foreign markets accessed by individual exporters – a fact suggestive of the presence of informational barriers. Drawing an analogy with astrophysics, Head and Mayer [2013] point out that at most 30 percent of the variation in trade flows can be explained by observable freight costs, while the remaining 70 percent of the variation is due to a "dark" trade cost. The authors argue that one significant component of these dark costs must be information costs.

Two recent papers have provided further evidence of informational barriers in goods trade. Looking at the market for agricultural goods in Philippines, Allen [2014] demonstrated that a number of features of the data can be explained by a model with information frictions, but are not consistent with a full information model. Steinwender shows how the establishment of trans-Atlantic telegraph lines, that speeded up the flow of information between the U.S. and U.K., led to a convergence in prices and higher trade volumes for cotton.[4]

The paper that is closest in spirit to our paper is Allen [2014]. Unlike our static model though, he considers a model where producers sequentially search for the highest price across markets. In Allen's paper, information frictions manifest in the form of (i) a fixed cost that producers have to pay to learn about prices in each market, and (ii) an exogenous probability of searching each market. Allen [2014] goes on to show that the probability that producers in market j will search market i depends on a number of bilateral variables, the most important being distance. Our model suggests why this might be the case. *Ceteris paribus*, rationally inattentive importers process more information about markets that are close, or in other words, markets with low expected prices. This, in turn, makes it more likely that there will be a transaction between two markets, over and above what can be explained by pure transport costs.

[4]Other papers that provide evidence that is consistent with the presence of information frictions in trade include Gould [1994], Head and Ries [1998], Rauch and Trindade [2002], Freund and Weinhold [2004], Fink et al. [2005], Combes et al. [2005] and Chan [2016].

In a related paper, Arkolakis et al. [2012] introduce staggered adjustment in the Eaton-Kortum model of trade. They assume that in each period, consumers continue to buy from the same supplier with some probability – consumers are inattentive. Accordingly, with some probability, consumers do not respond to price shocks that hit other suppliers. Arkolakis et al. takes the inattention as given, and is therefore silent on how the degree of inattention itself could respond to trade costs, which is a feature that generates many of the novel results in our paper.

The rest of the paper is organized as follows. In Section 2, we specify the information structure and incorporate it into an otherwise standard Ricardian trade model. In Section 3, we solve for the equilibrium import probabilities and discuss some of their novel properties. In Section 4, we examine an empirical implication of our model. In Section 5, we quantitatively evaluate the role of information processing costs. Section 6 concludes. All the proofs are in the Appendix.

2 The Model

We consider a static framework.[5] There are N countries, where each country is populated with a positive measure of workers who also consume.

Preferences. Consumers (or importers) have preferences defined over a continuum of products, which is the same across countries. We assume that preferences are such that consumers want to consume positive amounts of every product, i.e., preferences display a love for variety. The preferences generate an indirect utility function $v(\mathbb{P})g(Y)$, where $p(\omega) \in \mathbb{P}$ is the price of product ω and Y is the total income of a representative individual. We assume that there exists a monotonic transformation of v, denoted by T such that

$$T(v(.)) = \int_{\omega} u(1/p(\omega))d\omega. \tag{1}$$

We make the following assumptions about the function u: $u(1/p)$ is decreasing and convex in p.

Technology. The markets for the different products are perfectly competitive. Instead of defining the production function of a product, we consider its dual, the cost function. The cost of importing one unit of product ω into country j from country i is given by $1/\tilde{z}_{ij}(\omega)$, with

$$\tilde{z}_{ij}(\omega) = \frac{z_i(\omega)}{c_i \tau_{ij}}. \tag{2}$$

[5]In Appendix D, we sketch out a dynamic extension of our model.

c_i is the average cost of a standardized bundle of inputs required for producing one unit of any product in country i. For now, we take c_i as given, but endogenize it in Section 5. The observable trade cost between exporting country i and importing country j is captured by the iceberg cost τ_{ij}, i.e., country i has to ship τ_{ij} units in order to sell one unit of a good in country j. The trade cost τ_{ij} includes both policy barriers such as import tariffs and export subsidies as well as non-policy barriers such as transportation costs, border costs and time costs. Importantly, τ_{ij} does not include information costs. Finally, $z_i(\omega)$ is a random productivity draw for product ω in country i.

Information frictions. We introduce information frictions by assuming that the productivity realizations, $z_i(\omega)$, are not perfectly observable at the decision stage. We also assume that the cost of product ω produced in country i is fully revealed to importers in country j once country i has been chosen to supply the product. This assumption of perfect observability *ex-post*, combined with perfect competition in the market for each product, implies that the producers in any country do not engage in strategic price setting.[6] The price at which producers in country i are willing to sell product ω to importers in country j is then given by

$$p_{ij}(\omega) = \frac{1}{\tilde{z}_{ij}(\omega)},$$

i.e., producers choose to sell their goods at marginal cost. It must be emphasized that $p_{ij}(\omega)$ is the price that is *actually paid* by country j importers if they choose to purchase the manufactured good from country i. In a full information world, this would be equivalent to the lowest price for that product faced by importers in j [Eaton and Kortum, 2002]. But the un-observability of prices *ex-ante* implies that in this model, $p_{ij}(\omega)$ may not be the lowest price for product ω faced by importers in country j.

It is worth pointing out two observations. First, once an importer in country j chooses to purchase from country i, the transaction always takes place. This is because the preferences in (1) imply that importers always want to purchase a positive quantity.[7] Second, we do not make an ad-hoc assumption that importers have more information about the productivity draws in their own country relative to foreign countries. Rather, as we shall see below, this scenario may arise as an equilibrium outcome.

[6] In the presence of information frictions, firms selling a homogeneous product might choose to charge a price greater than marginal cost even with free entry.

[7] We implicitly rule out the possibility that within a given period, the importer can choose a country other than i in the event that the price in i is revealed to be too high. This assumption is not as restrictive as it seems. The final price paid by importers could involve rounds of negotiations. If these negotiations take time and the importer has to purchase within a given period, the importer may not want to switch suppliers even if the realized price turns out to be much higher than expected.

Importer's problem. Since the products are symmetric, we can simply focus on a representative product. Hence, we can drop the notation ω. For each product, a positive measure of importers in country j choose the source country for the product. Let Z be the vector with the random productivity draws of all countries such that $Z \in \mathbb{R}^N$, where z_i is a random productivity draw in country i. Z is drawn independently for each product from a distribution $G(Z) \in \Delta(\mathbb{R}^N)$, where $\Delta(\mathbb{R}^N)$ is the set of all probability distributions on $\mathbb{R}^N$. Each importer independently chooses his information strategy about the random productivity draws for every product-country pair. Importers are not constrained to learn about these productivity draws with a particular signal structure; rather, they are allowed to choose the optimal mechanism to process information. Given their information strategy, importers receive independent signals about Z, update their information using Bayesian updating and then choose the source country.

The importer's problem has two stages. In the first stage, importers choose their information strategy about Z, taking into account that information is costly to acquire, and update their beliefs. Importers have some prior belief $G(Z)$ and receive signals $S \in \mathbb{R}^N$ about Z to update their beliefs. The information strategy is a joint distribution $F(S, Z) \in \Delta(\mathbb{R}^{2N})$ such that $\int_S F(dS, Z) = G(Z)$ for any $Z \in \mathbb{R}^N$. Given this restriction, the strategy is equivalent to choosing the conditional distribution $F(S|Z)$.[8] In the second stage, importers choose the source country from the set $C = \{1, \cdots, N\}$.

Following Sims [2003], we use tools from information theory to model the limited information processing capabilities of importers. At this point, we define two mathematical objects that form an integral part of our analysis.

Definition. The *entropy* $H(X)$ of a discrete random variable X that takes values x in $\mathbb{X}$ is

$$H(X) = -\sum_{x \in \mathbb{X}} \mathcal{P}(x) \ln \mathcal{P}(x),$$

where $\mathcal{P}(x)$ is the probability mass function of X.

Definition. The *mutual information* of two random variables X and Y (taking values y in $\mathbb{Y}$) is given by

$$I(X; Y) = H(X) - E_y[H(X|Y)],$$

where $H(X|Y) = -\sum_x \mathcal{P}(x|y) \ln \mathcal{P}(x|y)$ is the entropy of X conditional on Y.

[8] For example, a standard exogenous information strategy that is frequently assumed would consist of observing one imperfect signal for each productivity realization in every country, i.e., $y_i = z_i + \epsilon_i$. This information strategy is a feasible strategy in our setup, but it is not optimal.

Intuitively, mutual information measures the reduction in the entropy of X caused by the knowledge of Y. We use entropy as the measure of uncertainty about the productivity draws and mutual information as the measure of uncertainty reduction or information [Shannon, 1948]. The following property of mutual information will be useful later on:

PROPERTY 1: $H(X) - E_y[H(X|Y)] = H(Y) - E_x[H(Y|X)]$.

Importers in country j process information by receiving signals S about Z to reduce the entropy $H(Z)$. Hence, the uncertainty reduction of importers in country j, denoted as κ_j, about Z through the observation of S is the mutual information between productivities Z and signals S

$$\kappa_j = I(Z;S) = H(Z) - E_S\big[H\big(Z|S\big)\big]. \tag{3}$$

If information could be processed freely, an importer would find out the true realization of Z. There are, however, a multitude of costs involved in processing information about the true productivity of a supplier. We assume that such costs are specific to the destination country and denote these costs by λ_j. The assumption of a constant information cost is made for tractability reasons and is common in the literature on rational inattention. By paying a cost $\lambda_j\kappa_j$, country j importers can reduce their uncertainty about the realization of Z by κ_j.[9]

Given the additive preference structure, importers maximize the expected utility of each product, taking into account that processing information about the productivity draws is costly. That is, importers in country j solve the following optimization problem:

$$\max_{F, i\in C} E\big[u(\tilde{z}_{ij}) - \lambda_j\kappa_j\big]$$

where $\tilde{z}_{ij}$ is given by (2), κ_j is given by (3) and the expectation is taken with respect to the distribution over (Z, S) induced by the prior G.

Despite the added complexity of not being constrained to learn about productivity draws with a particular signal structure, however, Matejka and McKay [2014] show that the importer's problem can be reduced to a simpler maximization problem.[10] Specifically, it is enough to solve for the optimal distribution of actions conditional on the realization of the variables of interest [Matejka and McKay, 2014]. Intuitively, two different signals that lead to the same action is not the most efficient information choice as there is information that is acquired but unused. Hence, it is

[9]Note that λ_j is a parameter while κ_j is a variable.
[10]See LEMMA 1 of Matejka and McKay [2014].

optimal for importers to associate one action (source country selection) with at most one particular signal. As actions are associated with at most one specific signal, the information processed by importers in country j can be calculated as the mutual information between productivities Z (variable of interest) and the selected country i chosen by importers in country j (action).

In our model, importers in country j choose the probability that a product is purchased from country i, conditional on the productivity realizations. Let us define $f_{ij}(Z)$ as the posterior probability that country j importers purchase a product from country i conditional on a particular productivity realization for that product across countries, Z. Defining π_{ij} as the unconditional probability that country j importers buy the product from country i, we have

$$\pi_{ij} = \int_Z f_{ij}(Z)dG(Z), \tag{4}$$

where $G(Z)$ is the distribution of Z across products. Note that π_{ij} is also the prior or expected probability that country j importers purchase any product from country i. As stated above, the information processed by importers in country j can be calculated as the mutual information between productivities Z and the country i chosen by importers in country j:

$$\kappa_j = H(Z) - E\big[H\big(Z|i \in C\big)\big],$$

where $H\big(Z|i \in C\big)$ is the entropy of Z, conditional on country j importers purchasing a product from country i. The revised importer's problem is given by

$$\max_{f_{ij}(Z)} \sum_{i=1}^{N} \int_Z u(\tilde{z}_{ij}) f_{ij}(Z)dG(Z) - \lambda_j \kappa_j,$$

subject to

$$\kappa_j = -\sum_{i=1}^{N} \pi_{ij} \ln \pi_{ij} + \int_Z \Big(\sum_{i=1}^{N} f_{ij}(Z) \ln f_{ij}(Z)\Big) dG(Z), \tag{5}$$

$$f_{ij}(Z) \geq 0 \quad \forall i, \tag{6}$$

$$\sum_{i=1}^{N} f_{ij}(Z) = 1, \tag{7}$$

where $\tilde{z}_{ij}$ is given by (2). The first term in the objective function is the expected utility of importers from purchasing a product, while the second term is the cost of processing information. Rationally inattentive importers in country j choose the probability of importing from country i conditional on the realization of Z. In deriving the amount of information processed in (5), we have used Property 1. Equations (6) and (7) simply say that $f_{ij}(Z)$ must be a probability mass function.

3 Equilibrium.

Following Matejka and McKay [2014], the next proposition derives the equilibrium posterior probability of purchasing a given product:

Proposition 1. *If $\lambda_j > 0$, then conditional on the realization of Z, the probability that importers in country j choose to purchase a product from country i is given by*

$$f_{ij}(Z) = \frac{\pi_{ij} e^{u(\tilde{z}_{ij})/\lambda_j}}{\sum_{k=1}^{N} \pi_{kj} e^{u(\tilde{z}_{kj})/\lambda_j}}, \tag{8}$$

where π_{ij} is given by (4).

The posterior choice probabilities have a structure similar to a multinomial logit [McFadden et al., 1973], except that they are adjusted by the prior probabilities, π_{ij}. These π_{ij}-s are independent of productivity realizations of individual products and only depend on exogenous objects such as the productivity distribution, informations costs, preferences, and input costs. When the cost of information is high, posterior choice probabilities $f_{ij}(Z)$ attach a high weight to prior probabilities as importers process small amounts of information. In this case, if a country i is perceived as being highly productive *ex-ante*, then it has a high probability of being chosen as the source for a product even if its actual productivity in that product is low. When the cost of information is low, the posterior choice probabilities attach a high weight to the actual productivity realizations, Z, as importers process large amounts of information and receive signals about Z that are much more precise. The following proposition discusses an important property of $f_{ij}(Z)$.

Proposition 2. *$f_{ij}(Z) > 0$ has the following properties:*

1. *If $\pi_{ij} > 0$, then $f_{ij}(Z) > 0$.*

2. *If there exists a unique i such that $\tilde{z}_{ij} = \max \tilde{z}_{kj}$ for all k, then as $\lambda_j \to 0$, $f_{ij}(Z) \to 0$ for all $k \neq i$ and $f_{ij}(Z) \to 1$.*

An implication of the above proposition is that importers in one country could buy the same product from different countries.[11] Notice that Proposition 1 contrasts sharply with the result in full information Ricardian models such as Eaton and Kortum [2002]. In that paper, even though *a priori* importers in country j can buy a given product from any country, after the productivity draws are realized, this probability drops to zero for every exporting country but the one with the lowest price. In fact, as Proposition 1 shows, as the cost of information becomes negligible and our model converges to a full information model, the conditional probabilities converge to either zero or one. But as long as there are positive information costs, this is not true any more. Importers never observe the true productivity draws and believe that every country can have the cheapest product with some probability.

In the literature, when a narrowly defined product is imported from many countries, it is usually assumed that different countries produce different varieties of the same product [Klenow and Rodriguez-Clare, 1997]. In our model, the exact same product could still be imported from multiple countries because of information frictions. Furthermore, if the prior probabilities that country j both imports from as well as exports to country i are positive, then so are the posterior probabilities. Hence, in equilibrium, we could observe the same product being traded in both directions by two countries. This feature, which is shared by Allen [2014], cannot be generated in a full information model of trade. The next proposition discusses some properties of the prior probability π_{ij}.

Proposition 3. *π_{ij} has the following properties:*

1. *π_{ij} is decreasing in input costs c_i and trade costs τ_{ij}.*
2. *If there exists a unique i such that $c_i\tau_{ij} = \min c_k\tau_{kj}$ for all k, then as $\lambda_j \to \infty$, $\pi_{kj} \to 0$ for all $k \neq i$ and $\pi_{ij} \to 1$.*
3. *If trade is frictionless and countries are ex-ante identical ($\tau_{ij} = 1 \forall i, j$ and $c_i = c \forall i$), then $\pi_{ij} = 1/N$ for all i.*

The first property of Proposition 3 states that *ex-ante*, importers in country j are less likely to purchase a product from countries with high expected costs. Holding everything else constant, an

[11] All importers in a given country have the same initial beliefs about which source country has the lowest price for a product. Their actions, however, may be heterogeneous. If $f_{ij}(Z) > 0$ and $f_{hj}(Z) > 0$, then a fraction $f_{ij}(Z)$ of importers in country j will choose to purchase the product from country i, while a fraction $f_{hj}(Z)$ of importers will choose to import from country h. Intuitively, even though all importers in a country choose the same signal structure, different importers could receive different signals about the productivity draws at a given point in time, and could end up buying from different countries based on these signals.

increase in c_i or τ_{ij} reduces the probability that the price of that product in country i is the lowest price among all countries.

The second property of Proposition 3 demonstrates that all else equal, if the information processing cost becomes extremely large, importers tend to purchase from only one country. Intuitively, when the information processing cost is high, importers incorporate less information into their decision making and attach a greater weight to the primitives, c_i and τ_{ij}. If the expected cost of importing from country i is the lowest, then an increase in importance of the primitives raises the likelihood that country j importers will buy from country i.

The third property of Proposition 3 establishes that in a world with no trade costs and where countries are *a priori* identical, all countries have the same *ex-ante* probability of being selected as the source for a product by importers in country j. In this case, the choice probabilities $f_{ij}(Z)$ in equation (8) follow a standard multinomial logit.

A novel property of our model concerns the effect of information costs on the concentration of the import distribution for a product. To see this, consider a world with complete information ($\lambda_j = 0$). In such a world, importers in country j buy a product almost surely from just the country with the lowest price, as shown in Proposition 2. In this case, the distribution of $f_{ij}(Z)$ for that product will be degenerate. But as the information cost rises, the true productivity realizations are not observed any more. Accordingly, importers diversify their purchases, causing the import distribution to become non-degenerate. At the same time, part 2 of Proposition 3 also shows that importers in country j buy a product almost surely from just one country if they face arbitrarily high costs of processing information ($\lambda \to \infty$). We state this result formally in the next proposition:

Proposition 4. *Starting from a zero information cost, an increase in the information cost causes the distribution of imports of a given product to become less concentrated. Starting from an infinitely high information cost, a reduction in the information cost causes the distribution of imports of a given product to become less concentrated.*

Note that while deriving Propositions 1, 2, 3 and 4, we did not specify a distribution of productivity $G(Z)$. In particular, these results are satisfied for any $G(Z)$. None of the known family of distributions, however, permit analytical solutions for the π_{ij}-s and the f_{ij}-s as there is no general solution for the integral in (4). Hence, we use numerical integration to derive more comparative statics results. In Appendix C, under a non-standard distribution and a restrictive parameter constraint, we derive a closed-form solution to the problem.

Numerical Exercise. For this exercise, we assume there are four countries, indexed by 1,...,4, that have identical input costs (i.e., we set $c_i = 1$ for all i). We order countries by their cost of exporting to country 1, τ_{i1}, and assume that $\tau_{11} < ..... < \tau_{41}$. Finally, we assume that $u(\tilde{z}_{ij}) = \log(\tilde{z}_{ij})$ and that the log productivities are drawn independently from a normal distribution with mean zero and standard deviation σ. We draw a vector of productivities for the four countries one hundred thousand times (corresponding to one hundred thousand products).[12]

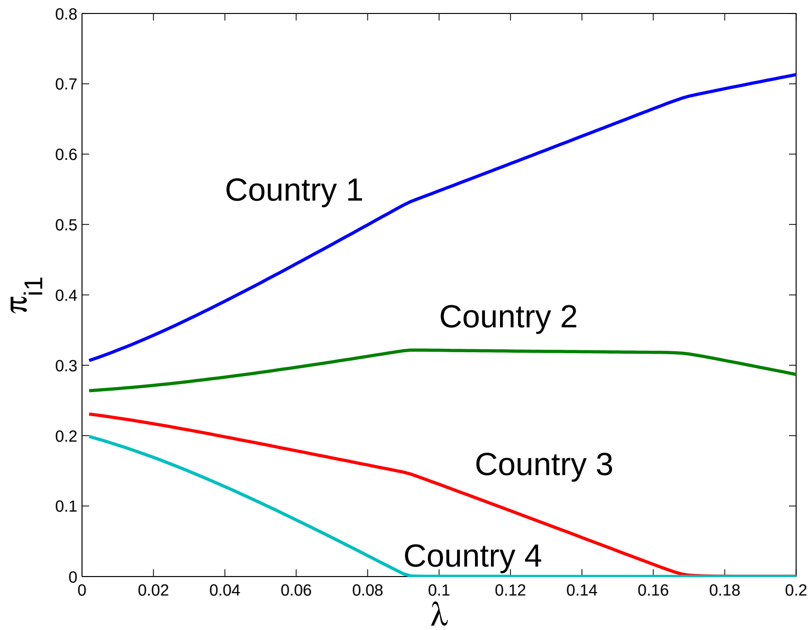


Figure 1: How π_{i1} varies with information costs

Figure 1 shows the prior probabilities π_{i1} ($i = 1, ..., 4$) for different levels of information processing cost λ. As Figure 1 illustrates, in the presence of information processing costs, as traditional trade costs increase, trade declines much more than what would be predicted in a full information world. In a full information model ($\lambda \to 0$) such as Eaton and Kortum [2002], when traditional trade costs increase, the expected price of country i products increases. In our model when $\lambda > 0$, there is an additional effect. The rationally inattentive importer in country j compares the expected marginal benefit of processing information about country i's productivity with the marginal cost of information. As the probability of getting the lowest price in country i declines, so does the information processed by country j importers about country i. Consequently, π_{ij} drops even more – the presence of information costs creates a *magnification effect*.

Figure 1 also sheds light on two properties of π_{ij} that highlight novel insights from rational inattention theory – *asymmetry* and *non-monotonicity*. A change in the information cost has an

[12] In particular, we assume that $\tau_{11} = 1.000, \tau_{21} = 1.005, \tau_{31} = 1.010, \tau_{41} = 1.015$ and $\sigma = 0.05$.

asymmetric effect on these probabilities. If λ increases, π_{i1} from countries other than 1 do not necessarily decline. Rather, when λ is small, an increase in λ actually leads to an increase in π_{21}. Intuitively, when information costs increase, importers in country 1 reallocate attention to countries with lower expected costs, to the detriment of other countries. Thus, for our chosen parameter values, an increase in λ leads to an increase in the attention allocated to countries 1 and 2, but a reduction of attention to countries 3 and 4, resulting in an increase in π_{11} and π_{21}, and a decrease in π_{31} and π_{41}. It is *as if* country 1 imposed differential import tariffs on goods imported from the other countries, with the tariff being higher for the country that is farther away. Hence, information costs have asymmetric effects on bilateral trade flows as they may increase the share of imports from countries with low expected costs and decrease the share of imports from countries with high expected costs. Standard trade models do not share this prediction. Rather, in these models, an exogenous change in a trade cost, which is applied uniformly across source countries, affects all import flows in exactly the same way. In a richer model with a more general trade cost function, however, the effect of a change in a "global" trade cost on trade flows could be asymmetric, playing a similar role as information costs in our setup. We sketch out such a model in Appendix B.[13]

Figure 1 also shows that the probability of country 1 importers buying a product from country 2 displays a hump-shaped behaviour with respect to information costs. This contrasts with the response of import shares to a change in standard trade costs, as stated in Proposition 3, where increases in input costs c_i and trade costs τ_{ij} have monotonic effects. As discussed above, when there is an increase in information costs starting from low levels, importers in country 1 reduce the total amount of information processed and substitute their attention from countries 3 and 4 (countries with high trade costs) to countries 1 and 2. But for high enough information costs, country 1 importers re-allocate attention from country 2 to 1, resulting in a decline in imports from country 2. Hence, the effect of information costs on trade shares from country 2 is non-monotonic.

Figure 2 plots the trade elasticity against the information processing cost.[14] It shows that the trade elasticity is increasing in information costs, as suggested by Figure 1. When information costs are high, importers in country 1 optimally allocate more attention to countries with lower trade costs, resulting in disproportionately more trade with those countries. Small trade costs

[13]We would like to thank a referee for pointing out this similarity between our model and a model with a more general trade cost function.

[14]The trade elasticity in this picture was computed using import flows into country 1. For a given λ, we computed $(\ln \pi_{i1} - \ln \pi 11)/(\ln \tau_{i1} - \ln 1)$ for every trade partner i and then took an average across i. This is a partial equilibrium exercise where we are focussing on imports only into country 1.

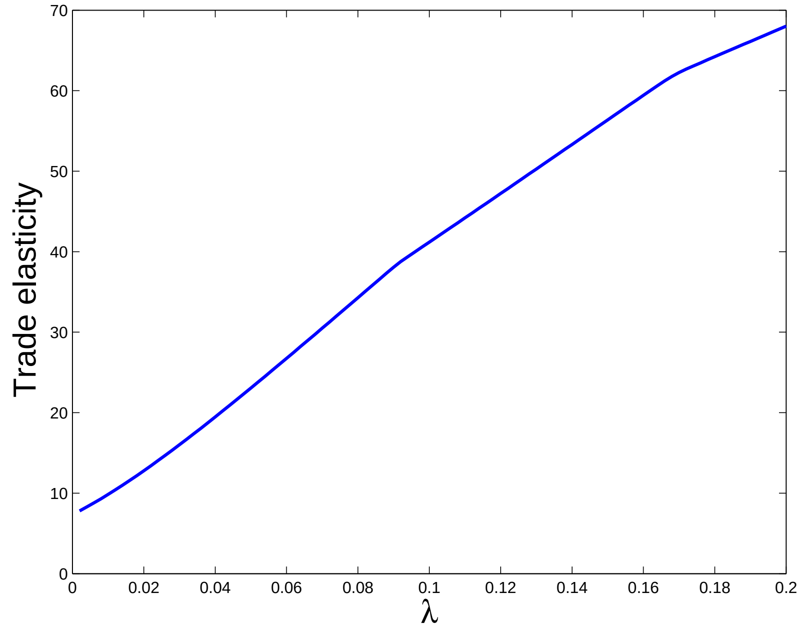


Figure 2: How trade elasticity varies with information costs

impose heavy penalties on countries that are *ex-ante* not very attractive sources for a product.

Figure 2 suggests that if our model had different types of products with product-specific information processing costs, then products with high λ would have a higher trade elasticity than those with low λ. If we assume that differentiated products have higher λ than reference-priced products, then our model is consistent with the findings by Rauch [1999], where he showed that the elasticity of trade with respect to distance is higher for differentiated goods relative to reference-priced goods.[15] Rauch conjectured that the cost of learning about differentiated products is higher relative to reference-priced products as the former have multiple attributes and might require search and matching. In other words, the cost of processing information about differentiated products might be higher.

Our model also provides a possible explanation for the distance elasticity puzzle. This puzzle refers to the issue that the elasticity of trade costs with respect to distance is much smaller than what is needed to explain trade data using traditional models. Grossman [1998] was one of the first researchers to point out that freight costs are not enough to account for the effect of distance on trade. In fact, Grossman suggested that distance could be a proxy for other barriers such as information frictions. Although our model may qualitatively resolve the distance elasticity

[15]Reference-priced products are those that are not transacted in centralized exchanges, but whose prices are published in trade journals.

puzzle, the bigger question remains: how much can it actually explain? In Section 5, we attempt to provide an answer.[16]

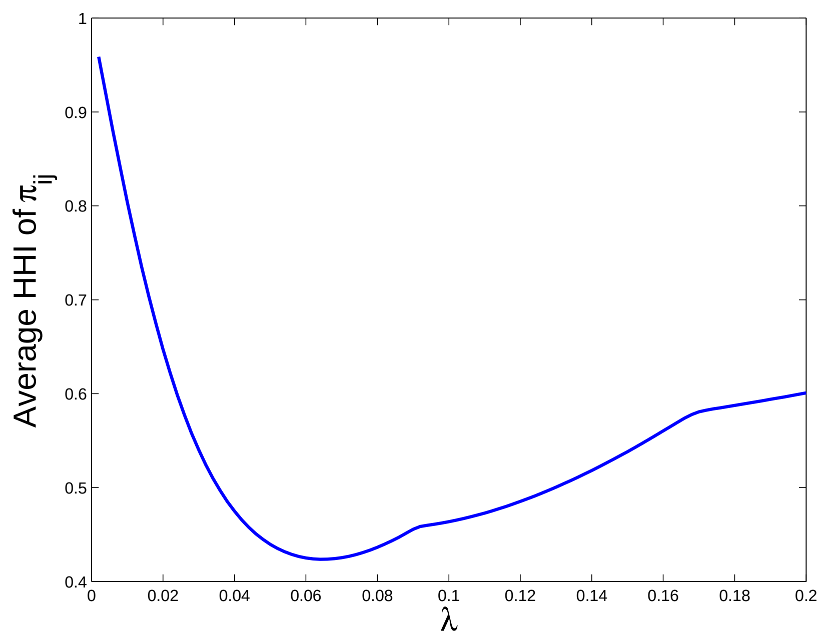


Figure 3: Import concentration as a function of information costs

Recall that in deriving a theoretical relation between the information processing cost λ and the concentration of the import distribution, we could characterize only the end-points ($\lambda = 0$ and $\lambda = \infty$). Numerically, we can characterize the import concentration over the entire range of λ. Figure 3 shows the average import concentration for country 1, measured by the Herfindahl-Hirschman index (HHI), plotted against λ.[17] The figure shows that as λ increases, the import distribution of a product tends to become less concentrated initially, before its concentration starts to rise.

To summarize, a model with rationally inattentive importers magnifies the effect of traditional trade barriers on trade. A change in a trade barrier such as transport cost not only has a direct effect on import probabilities, but by changing how importers process information, has an indirect effect as well. Furthermore, a change in the information processing cost has asymmetric and non-monotonic effects on import probabilities. In the next section, we try to provide evidence that is supportive of our model.

[16] An alternative explanation for why the distance elasticity is large is provided by Krautheim [2012].

[17] For a variable taking T distinct values with the corresponding shares being s_t, ($\sum_{t=1}^{T} s_t = 1$), the Herfindahl-Hirschman index is given by $HHI = \sum_{t=1}^{T} s_t^2$. The HHI lies between 0 and 1, with 1 corresponding to a de-generate (completely concentrated) distribution.

4 Evidence

Our model of inattention generates the following prediction: for a given product and importing country, the concentration of purchase probabilities, π_{ij}, is initially decreasing and then increasing in information costs (See Figure 3). In this section, we test this prediction. Since we do not observe these probabilities, we state a lemma that establishes a useful property of π_{ij}:

Lemma 1. *Under Cobb-Douglas preferences, π_{ij} equals country j's share of expenditure on goods imported from country i.*

The above lemma establishes the equivalence between the purchase probabilities and import shares under Cobb-Douglas preferences. Hence, all the results involving π_{ij}-s apply to the corresponding import shares as well. To examine the prediction about non-monotonicity of the import share concentration, we carry out a cross-country analysis, and conjecture that the information cost varies across countries. A caveat is in order. All we uncover in this section is a correlation. We do not make any attempt to infer causality. Nevertheless, the prediction relating information costs with the concentration of imports is novel and we think of our exercise as a preliminary attempt to examine whether the data is indeed consistent with this prediction.

Data: From the NBER-UN database, we construct a sample of 770 4-digit Standard International Trade Classification (SITC, Revision 2) product categories and 110 importing countries for the year 1999. Of course, not every country imports every product. Table 1 summarizes the distribution of products in terms of the countries they are imported from. Observe that 75 percent of products are imported by at least half the countries in the sample, while the top 10 percent of products are imported by more than 100 countries. We measure the concentration of an import distribution for *each product-country pair* using the Herfindahl-Hirschman index (HHI).

	Share of products				
Percentile	10	25	50	75	90
# of countries per product	35	55	77	94	103

Source: NBER-UN World Trade Flows.

Table 1: Distribution of importing countries per SITC code

Capturing the true cost of processing information is challenging. The measure that we use in this paper is the inverse of "international bandwidth", which is defined as the maximum rate of data transmission from a country to the rest of the world. It is typically measured in megabits per second or gigabits per second.[18] The data for international bandwidth is collected by the International Telecommunications Union, a United Nations specialized agency for information and communication technologies. It is calculated by adding up the capacity of all international data lines connecting a country with all the other countries.

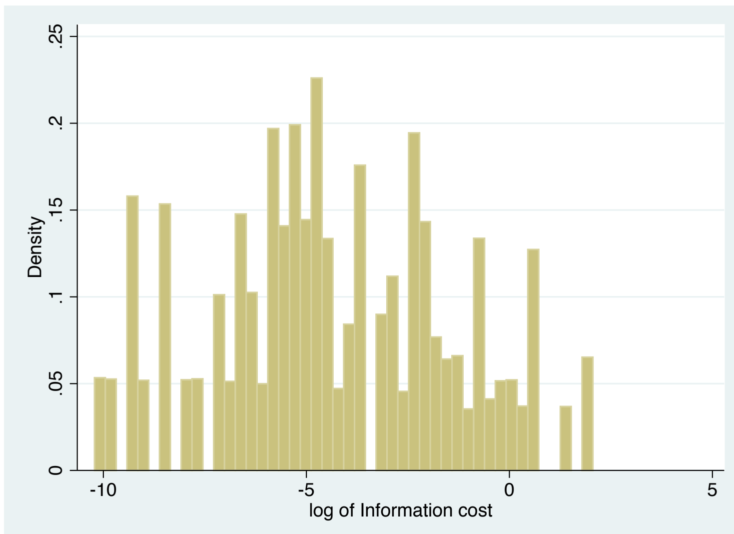


Figure 4: Histogram of the information cost across countries

The international bandwidth is used primarily to carry internet traffic. Our choice of international bandwidth as an inverse measure of the cost of processing information is based on the assumption that faster internet speeds allow users to process more information, effectively lowering the cost of processing each bit of information. There is a lot of variation in the information cost for our sample of countries. This is displayed in Figure 4.

In order to investigate the relationship between the concentration of imports and information costs, we would like to impose as little structure as possible. This is because, our theory does not predict a particular shape for the relationship except that it is non-monotonic. Consequently, we choose a nonparametric approach to uncover the shape of the relationship. In particular, we divide the range for the information cost into intervals and then run the following regression separately within each interval:

$$\ln HHI_j^h = \alpha^h + \beta \ln I_j + \gamma' X + \epsilon_j^h,$$

[18]Other proxies for information frictions used in the literature include the number of web hosts located in a country [Freund and Weinhold, 2004] and bilateral telecommunication prices Fink et al. [2005].

where HHI_j^h is a measure of concentration of imports for product h in importing country j, I_j is the inverse of international bandwidth, our preferred measure of the information cost in country j and X is a vector of other importer-specific and product-importer specific regressors. The coefficient α^h captures product fixed effects. Finally, ϵ_j^h is an error term orthogonal to the regressors. If we find that β is negative for low values of the information cost but positive for high values, that would lend support to our hypothesis.[19]

An issue with regressing HHI simply on the information cost is that an entirely different mechanism might be driving the relationship between these two variables. To see this, observe that if country j is equally likely to import from N_j different countries, the HHI of its import distribution reduces to $1/N_j$. Consequently, the more sources a country imports from, the lower is its HHI. Now, if countries with a lower information cost also import from more sources (probably because these are also the richer countries), a positive value of β could be explained without using our model of rational inattention.[20] Therefore, we include the number of source countries for a product as an additional regressor.

The results without product fixed effects are displayed in Table 2. We create five intervals for the information cost, using the 20th, 40th, 60th and 80th percentiles of cost. Each column in the table corresponds to the regression in one of the five intervals. Two results stand out. First, as hypothesized above, the import distribution for a product seems to be less concentrated when the number of countries sourcing that product is higher. Second, β is negative in the low-cost intervals, increases as we move to higher-cost intervals and becomes positive in the highest-cost interval. Controlling for the number of countries a product is sourced from, the import distribution initially becomes less concentrated as the cost of information increases, but becomes more concentrated for further increases.

In Table 3, we add two other regressors. It is quite possible that the relationship between the information cost and the import concentration is driven by selection. Suppose that poor countries import only from other poor countries and it is much harder to find out what the true prices are in those countries. At the same time, rich countries import primarily from other rich countries, with the information about producers in those countries being much easier to obtain. In this case, one could argue that importers in poor countries face higher uncertainty about prices and would diversify their purchases across trading partners, resulting in less concentrated import distribu-

[19]This nonparametric approach has an additional advantage: unlike polynomial methods, this approach produces coefficients that are locally robust. This means that the relationship between HHI and the information cost at, say, high levels of cost is unaffected by observations corresponding to low levels of cost [Imbs and Wacziarg, 2003].

[20]Although absent in our model, fixed costs of exporting/importing could generate systematic predictions about the extensive margin of trade.

Table 2: Country-product level regressions

	Dependant variable: HHI				
	(1)	(2)	(3)	(4)	(5)
Information cost Percentile	0-20	20-40	40-60	60-80	80-100
Information cost	-0.08***	-0.08***	-0.06***	-0.04***	0.01***
	(0.005)	(0.011)	(0.009)	(0.005)	(0.003)
Number of exporters per product	-0.53***	-0.56***	-0.55***	-0.62***	-0.65***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.005)
Constant	-0.69***	-0.54***	-0.39***	-0.15***	-0.02***
	(0.047)	(0.066)	(0.039)	(0.014)	(0.002)
Product FEs	No	No	No	No	No
Observations	7346	6888	8750	9621	6387
R^2	0.52	0.66	0.69	0.78	0.84

Note: Clustered (at the country level) standard errors in paranthesis. *, ** and *** refer to significance at the 10%, 5% and 1% levels respectively. All variables are in logs. Dependant variable is the Herfindahl-Hirschman Index for a given importer-product. The information cost is measured as the inverse of international bandwidth.
Source: NBER-UN World Trade Flows for trade, International Telecommunications Union (UTI) for international bandwidth.

tions, while importers in rich countries would tend to purchase larger shares from countries with lower prices, resulting in more concentrated import distributions. To check this possibility, we compute the average income (per capita GDP) of the exporting countries for each product that a country imports. Inclusion of this variable does not alter the results. Furthermore, the coefficient on this variable is negative in most of the cost intervals, which is the opposite of what we had conjectured.

Another possibility is that rich countries happen to be closer to countries with high productivity while poor countries are not. For example, Canada is much closer to a high productivity country such as the U.S. than it is to a low productivity country such as Peru. This would cause Canadian importers to purchase much more from the U.S. relative to Peru, resulting in a highly concentrated import distribution. On the other hand, Honduras, which is roughly equidistant from both the U.S. and Peru, would tend to buy less from the U.S. and more from Peru, both relative to Canada, resulting in a less concentrated import distribution.

Table 3: Country-product level regressions

	Dependant variable: HHI				
	(1)	(2)	(3)	(4)	(5)
Information cost Percentile	0-20	20-40	40-60	60-80	80-100
Information cost	-0.05***	-0.08***	-0.07***	-0.04***	0.01***
	(0.006)	(0.011)	(0.010)	(0.005)	(0.003)
Number of exporters per product	-0.53***	-0.56***	-0.55***	-0.62***	-0.65***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.005)
Average exporter income	-0.07***	-0.06***	-0.05***	-0.02***	-0.01***
	(0.017)	(0.008)	(0.007)	(0.003)	(0.002)
HHI of exporter productivity	-0.18***	-0.03***	-0.01*	-0.00	0.03***
	(0.014)	(0.009)	(0.005)	(0.007)	(0.010)
Constant	-2.64***	0.76***	0.80***	0.23***	0.25***
	(0.418)	(0.210)	(0.171)	(0.083)	(0.054)
Product FEs	No	No	No	No	No
Observations	7346	6888	8750	9621	6387
R^2	0.53	0.67	0.69	0.78	0.84

Note: Clustered (at the country level) standard errors in paranthesis. *, ** and *** refer to significance at the 10%, 5% and 1% levels respectively. All variables are in logs. Dependant variable is the Herfindahl-Hirschman Index for a given importer-product. The information cost is measured as the inverse of international bandwidth.
Source: NBER-UN World Trade Flows for trade, International Telecommunications Union (UTI) for international bandwidth, Penn World Tables for per capita GDP and total factor productivity.

To examine this possibility, we define a measure of concentration of distance-adjusted productivity for country j, the HHI of exporter productivity:

$$\Phi_j = \sum_i \hat{\phi}_{ij}^2,$$

where

$$\hat{\phi}_{ij} = \frac{\phi_i / dist_{ij}}{\sum_l \phi_l / dist_{lj}}.$$

and ϕ_i is average productivity of country i. *Ceteris paribus* $\hat{\phi}_{ij}$ is higher, the more productive is

Table 4: Country-product level regressions

	Dependant variable: HHI				
	(1)	(2)	(3)	(4)	(5)
Information cost Percentile	0-20	20-40	40-60	60-80	80-100
Information cost	-0.10**	-0.07**	-0.07**	-0.04***	0.01*
	(0.039)	(0.032)	(0.030)	(0.017)	(0.004)
Number of exporters per product	-0.61***	-0.55***	-0.56***	-0.66***	-0.67***
	(0.068)	(0.017)	(0.007)	(0.014)	(0.015)
Constant	-0.65**	-0.55**	-0.39**	-0.15***	-0.01
	(0.247)	(0.203)	(0.138)	(0.048)	(0.006)
Product FEs	Yes	Yes	Yes	Yes	Yes
Observations	7346	6888	8750	9621	6387
R^2	0.63	0.74	0.74	0.81	0.86

Note: Clustered (at the country level) standard errors in paranthesis. *, ** and *** refer to significance at the 10%, 5% and 1% levels respectively. All variables are in logs. Dependant variable is the Herfindahl-Hirschman Index for a given importer-product. The information cost is measured as the inverse of international bandwidth.
Source: NBER-UN World Trade Flows for trade, International Telecommunications Union (UTI) for international bandwidth.

country i, or the smaller the distance between i and j. Because the $\hat{\phi}_{ij}$-s are shares, a large $\hat{\phi}_{ij}$ for some i implies that Φ_j is close to one. In the example above, Φ_j for Canada would be larger than Φ_j for Peru. Inclusion of this variable in the above regression generates a coefficient that is mostly negative.[21] This suggests that countries which are close to a few productive sources (high Φ_j) actually have less concentrated import distributions for individual products. Hence, this observation cannot explain our findings either.[22]

In Tables 2 and 3, we run pooled OLS regressions. A concern is that the observed relationship between the import concentration and the information cost is being driven by the fact that different countries import different products and there are technological differences across products, resulting in more or less concentrated import distributions. Accordingly, we run the above specifications with product fixed effects. Results are displayed in Tables 4 and 5. The significance of

[21]For average productivity, we use TFP measures from the Penn World Tables.

[22]We also included per capita GDP, but the coefficient on this variable turns out to be insignificant owing to the very high correlation with the information cost.

Table 5: Country-product level regressions

	Dependant variable: HHI				
	(1)	(2)	(3)	(4)	(5)
Information cost Percentile	0-20	20-40	40-60	60-80	80-100
Information cost	-0.07	-0.09***	-0.07**	-0.03**	0.01*
	(0.037)	(0.024)	(0.032)	(0.020)	(0.004)
Number of exporters per product	-0.64***	-0.56***	-0.56***	-0.66***	-0.67***
	(0.069)	(0.016)	(0.008)	(0.015)	(0.015)
Average exporter income	-0.12	-0.06***	-0.06***	-0.02	-0.00
	(0.076)	(0.015)	(0.017)	(0.010)	(0.003)
HHI of exporter productivity	-0.18	-0.03	-0.01	-0.01	0.02
	(0.136)	(0.022)	(0.014)	(0.025)	(0.021)
Constant	2.01	0.58	1.04**	0.19	0.17*
	(1.738)	(0.412)	(0.360)	(0.225)	(0.098)
Product FEs	Yes	Yes	Yes	Yes	Yes
Observations	7346	6888	8750	9621	6387
R^2	0.64	0.74	0.74	0.81	0.86

Note: Clustered (at the country level) standard errors in paranthesis. *, ** and *** refer to significance at the 10%, 5% and 1% levels respectively. All variables are in logs. Dependant variable is the Herfindahl-Hirschman Index for a given importer-product. The information cost is measured as the inverse of international bandwidth.
Source: NBER-UN World Trade Flows for trade, International Telecommunications Union (UTI) for international bandwidth, Penn World Tables for per capita GDP and total factor productivity.

some of the coefficients falls. For example, in Table 5, the coefficient on the information cost in the first interval (0-20th percentile) is not significant any more. Nevertheless, the pattern established earlier continues to hold: β is negative when the information cost is low, becomes smaller in absolute terms as the cost rises, and eventually becomes positive.

To summarize, the import concentration of a product has a non-monotone relationship with the cost of processing information. Furthermore, our finding cannot be explained by importing countries varying systematically in terms of whether they import a product from primarily rich or primarily poor countries, and whether they happen to be close to a few productive source countries. We should point out that none of the standard models of trade have any systematic prediction regarding the distribution of imports for a particular product on the one hand, and characteristics of the importing country on the other.[23] Of course, despite all the controls, we cannot fully rule out the possibility that the variations in the import concentration and the information cost are being driven by some third factor – the relationship that we establish is a correlation, and not causal. Next, we calibrate a multi-country model to evaluate the importance of information costs in facilitating cross-country trade.

5 Do information costs matter quantitatively?

One of the key insights that emerges from our analysis of inattentive importers is that in the presence of information processing costs, small differences in trade costs get magnified into large differences in trade flows (see Figure 1). This observation is relevant if we are trying to understand how trade flows respond to changes in policy barriers. In this section, we consider two counterfactual exercises that shed light on how a model with inattention could diverge from a full information model. To do that, first we need to specify how costs are determined.

We assume that each differentiated product is manufactured using a Cobb-Douglas technology that combines labour and an intermediate input, where the latter is a composite of all the available products. This leads to the following cost function for a product in country i:

$$c_i = w_i^{\alpha} P_i^{1-\alpha},$$

where w_i is the nominal wage in country i and P_i, the input price index, is given by

$$log(P_i) = \int \Big(\sum_{j=1}^{N} f_{ji}(Z) p_{ji}(Z) \Big) dZ. \tag{9}$$

Because each product is purchased from multiple countries, the term within the parenthesis in the above equation is the expected price for a product conditional on the productivity draw Z. The (logarithm of) aggregate price index is the integral of the expected prices across products. Because labour is the only factor of production, $w_i L_i$ is the aggregate income of country i.

The solution strategy for this model is as follows: First, for a given vector of P_i and π_i, use

[23] A wide class of models generate the following demand in country j for goods produced in country i: $\pi_{ij} = (p_{ij}/P_j)^{-\chi}$, where P_j is an aggregate price index for goods available in country j and χ, the elasticity of trade cost, is a structural parameter whose interpretation varies across models. In this case, the relative imports from two sources, a and b, is given by $(p_{aj}/p_{bj})^{-\chi}$, a term independent of country j's characteristics.

(8) to solve for the f_{ij}-s. Second, use (8) to check whether the resulting π_{ij} is the same as the initial guess and iterate until it is. Third, use (9) to check whether the resulting P_i is the same as the initial guess and iterate until it is. Therefore, solving for the equilibrium involves solving a series of fixed-point problems.

Because of the high-dimensionality of the model, computational time increases exponentially with the number of countries and products. In what follows, we choose $N = 25$ and we draw a vector of productivities for each country one hundred thousand times (corresponding to one hundred thousand products). We use 25 countries that accounted for more than 80 percent of all imports in 1990.[24] The products are assumed to be symmetric.

5.1 Calibration

To quantify the model, we need to choose a distribution of productivities $G(Z)$ and parameter values for α (the share of value-added in production), τ_{ij} (trade costs), w_j (nominal wages) and λ_j (information cost).[25] Following Eaton and Kortum [2002], we assume that $G(Z)$ follows a Fréchet distribution, with a shape parameter θ. Note that our full information model then becomes equivalent to the Eaton and Kortum (EK) model. We also use a parsimonious specification for observable trade costs τ_{ij}:

$$\tau_{ij} = tar^{1-FTA_{ij}} d_{ij}^{\rho},$$

where d_{ij} denotes the great circle distance between capital cities of countries i and j, FTA_{ij} takes a value of one if countries i and j are part of a free-trade area, and zero otherwise, and tar is an average tariff that applies to trade flows between any two countries that are not part of a free-trade area. We assume that tar equals 1.03, i.e., average import tariff is 3 percent. This is in line with the current weighted average tariffs for the world as a whole (United Nations Conference on Trade and Development). The two free-trade areas in our sample are NAFTA and E.U. We ignore all other geographic barriers to trade commonly used in the gravity literature such as borders and contiguity, common language and common colonial history.

To evaluate whether information frictions matter or not, we perform the counterfactuals with respect to two models – one with full information and the other with inattention. We assume that

[24] The countries in the sample include 18 OECD countries – Australia, Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, France, United Kingdom, Ireland, Italy, Japan, South Korea, Netherlands, Singapore, Sweden, United States – and 7 non-OECD countries – Brazil, China, India, Mexico, Malaysia, Russia and Thailand.

[25] Our decision to fix nominal wages was dictated by technical considerations. The solving of the π_{ij}s makes the problem computationally intensive. Each fixed-point problem increases computation time exponentially. Keeping nominal wages fixed allows us to carry out various counterfactuals within a reasonable amount of time. Fixed nominal wages also allow us to analyze the effect of trade shocks on trade balance, an object of independent interest.

the scale parameter for the Fréchet distribution is country-specific and use total factor productivity (TFP), obtained from the Penn-World Tables, as a proxy for these parameters. We also assume that nominal wages are proportional to TFP. Finally, we assume that $\lambda_j = \lambda$, i.e., countries face the same cost of processing information. Following Waugh [2010], we set α equal to 0.3. We calibrate the remaining parameters separately for each model, so that they match the *same set* of moments. The parameters are θ and ρ for both models, and the additional parameter λ for the model with inattention.

For the full information model, we proceed as follows. Simonovska and Waugh [2014] recently argued that in the EK model, $\theta = 4$. For the full information model, we use their preferred value for θ. To calibrate ρ, we make use of a well established result in the literature. In most models that generate a gravity equation, the elasticity of trade with respect to distance, δ, is the product of two elasticities: the elasticity of trade with respect to trade costs (or simply, the trade elasticity), ϵ, and the elasticity of trade costs with respect to distance, ρ. In the EK model, ϵ happens to equal θ.[26] Conducting a meta-analysis, Disdier and Head [2008] find that δ takes a value, on average, of around -1. This elasticity has been surprisingly stable over time, even increasing slightly under some specifications [Coe et al., 2007, Berthelon and Freund, 2008]. This implies that ρ must take a value of 0.25 to be consistent with $\delta = -1$. This combination of ρ and θ generate an average own import share (the share of expenditure a country spends on its own products) of 0.35. The average import share can be considered to be a measure of a country's openness. As a comparison, in 1999, this share in the data was around 0.5.

For the model with inattention, we allow the value of θ to be different from the full information model. Hence, we have three parameters, θ, ρ and λ which we calibrate to match three moments. The first moment is the same as the one used in the full information model – the distance elasticity of trade $\delta = -1$. When importers are inattentive, the distance elasticity is no longer a simple function of θ and ρ, but is some complex function involving those two parameters and λ. We also match an average own import share of 0.35, which is the own import share obtained in the model with full information. Finally, the additional moment that we use for the model with inattention is κ, the information processed by importers. The literature on rational inattention has found a number of estimates for information processing capacity, ranging between 0.08 and 3.[27] For our

[26] In Armington type national-product-differentiation models or monopolistic competition models of trade with homogenous firms, ϵ is the elasticity of substitution, while in monopolistic competition models of trade with heterogenous firms, ϵ is the Pareto shape parameter.

[27] See Luo [2008], Luo and Young [2009], Mackowiak and Wiederholt [2009] and Pasten and Schoenle [2016] for calibrated values of κ between 0.08 and 3.

benchmark specification, we target a κ in the middle of this interval, 1.5.[28] The values of the calibrated parameters are shown in Table 6.

Table 6: Calibrated values of parameters

	θ	ρ	λ
Full information	4	0.25	-
Inattention	5.5	0.08	0.6

The parameter θ is higher in the model with inattention than in the model with full information, while the parameter ρ is lower . Most studies that estimate ρ using measures of freight costs have found a value close to 0.025 [Limao and Venables, 2001, Hummels, 2007], which is an order of magnitude smaller than the 0.25 that is normally used in the full information model. Under inattention, the calibrated value of ρ is closer to its micro-estimates. A smaller value of ρ under incomplete information is not surprising. Faced with a cost of processing information, importers pay less attention to distant sources and, as a result, small differences in trade costs can lead to large differences in trade flows. Hence, to generate a similar amount of trade as in the full information model, we need a lower elasticity of trade cost with respect to distance.

A direct consequence of incomplete information is that for most products, importers no longer pay the minimum available price. Recall that inattentive importers import the same product from multiple countries, and hence pay different prices for the same product. Figure 5 shows the distribution of the ratio of the (average) price paid by U.S. importers for a product over its minimum price. As the figure suggests, there is a large variation in the (average) price paid for a product relative to the minimum, with the median (average) price being around 20 percent higher than the minimum price. One implication is that cross-country price differences may not always reflect the presence of conventional trade barriers.

We proceed to perform two counterfactuals – an increase of tariffs within NAFTA member countries and an increase of tariffs between the U.K. and the rest of E.U. following Brexit.

[28]The unit of κ is nats. If instead of using $\ln$ in the mutual information we had used log_2, then κ would be measured in bits. A $\kappa = 1.5$ nats is equivalent to $\kappa = 2.16$ bits. A signal with one bit of information allows the agent one question with a binary answer. Hence, in our calibrated model, importers, on average, are allowed to ask slightly more than 2 questions. An example of a question could be “Is country X the lowest cost producer of a product Y?” where the possible answers would be yes or no.

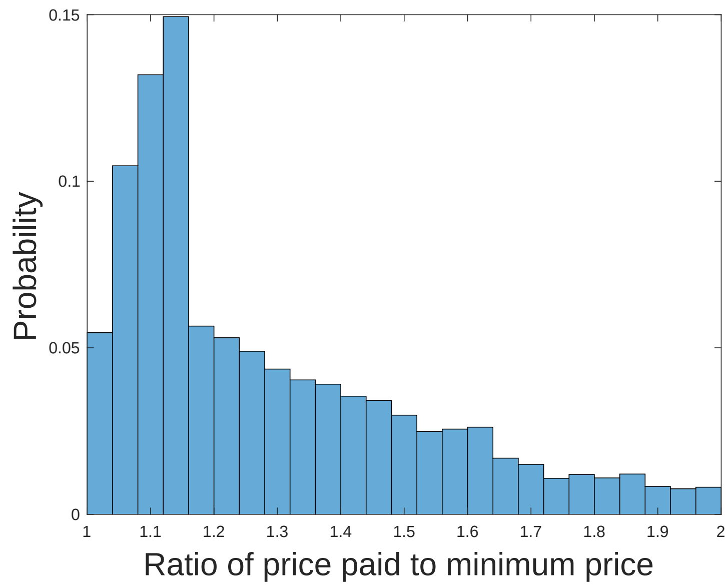


Figure 5: Distribution of prices paid by U.S. importers

5.2 Counterfactual 1: NAFTA termination

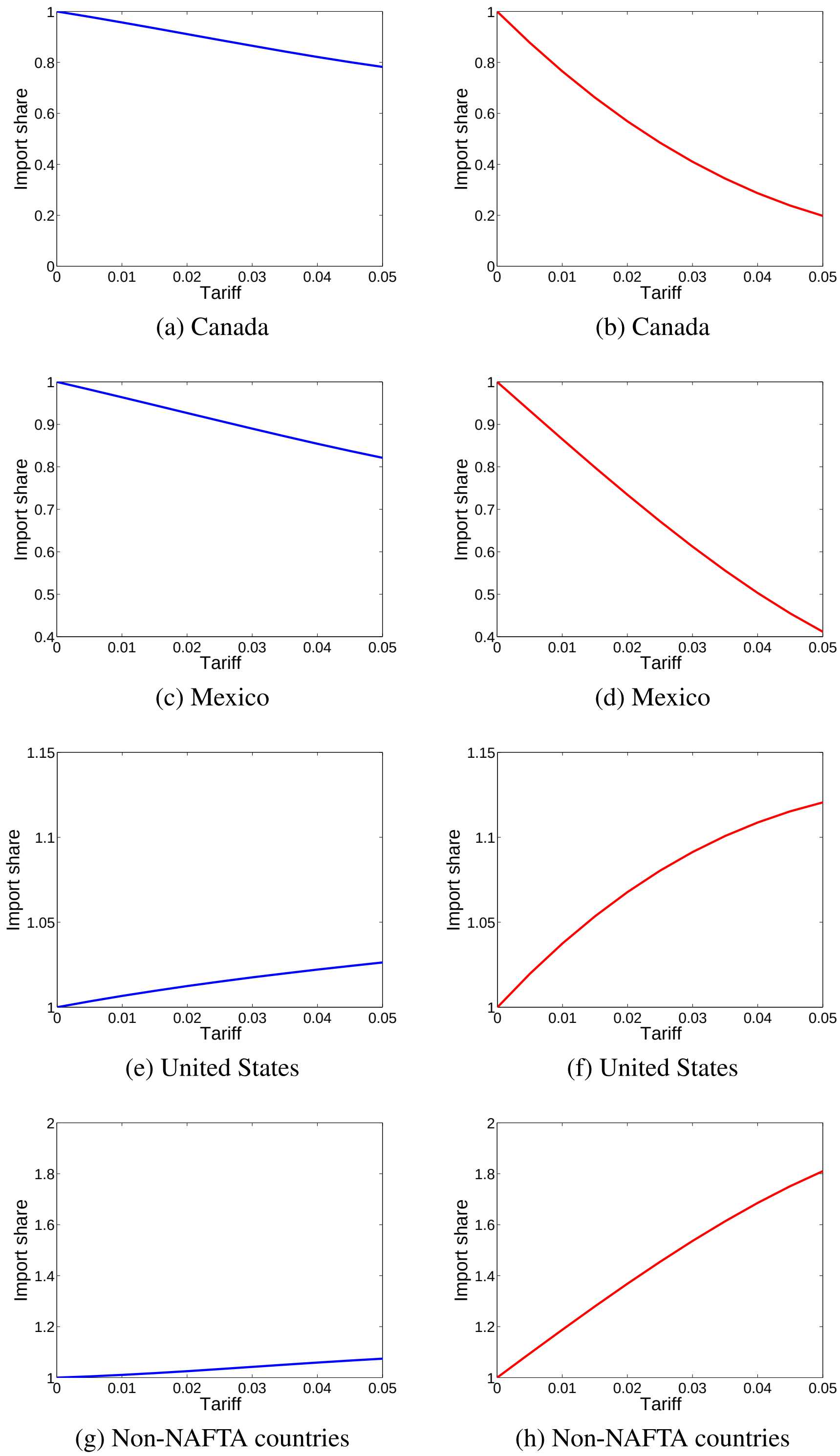
Since the election of Donald Trump as the president of the United States in November 2016, the fate of NAFTA has become uncertain. Calling NAFTA the "worst trade deal maybe ever signed anywhere" during the campaign trail, Trump vowed to terminate it if he were elected. Although, the Trump administration has, since then, decided to re-negotiate the agreement, there is ample skepticism that the negotiations will succeed.[29]

What happens to cross border tariffs if NAFTA is terminated? The answer depends on whether the U.S. government is able to negotiate separate agreements with Canada, and possibly Mexico, along the lines of NAFTA's predecessor, the U.S. - Canada Free Trade Agreement. In the absence of any agreements, U.S. import tariffs could jump on a range of products, from softwood lumber and aircrafts coming from Canada to cars and textiles coming from Mexico.[30,31] We consider a scenario where import tariffs within the free trade area rise uniformly across countries and products. In particular, we consider a gradual increase of tariffs from 0 to 5 percent. Given that

[29] See "Trump threat hovers over NAFTA as Ottawa talks end with no major progress", *CTV News*, September 27, 2017.

[30] See "U.S. imposing 20-per-cent tariff on Canadian softwood", *The Globe and Mail*, April 24, 2017, and "U.S. Slaps Duties on Canadian Jet, Raising Trade Tensions", *The New York Times*, September 26, 2017.

[31] See "NAFTA Talks Will Centre On Mexico's Dominance Of Auto Industry, Insiders Say", *Huffington Post*, August 3, 2017, and "U.S. demands risk scuttling NAFTA talks", *Bloomberg*, September 28, 2017.



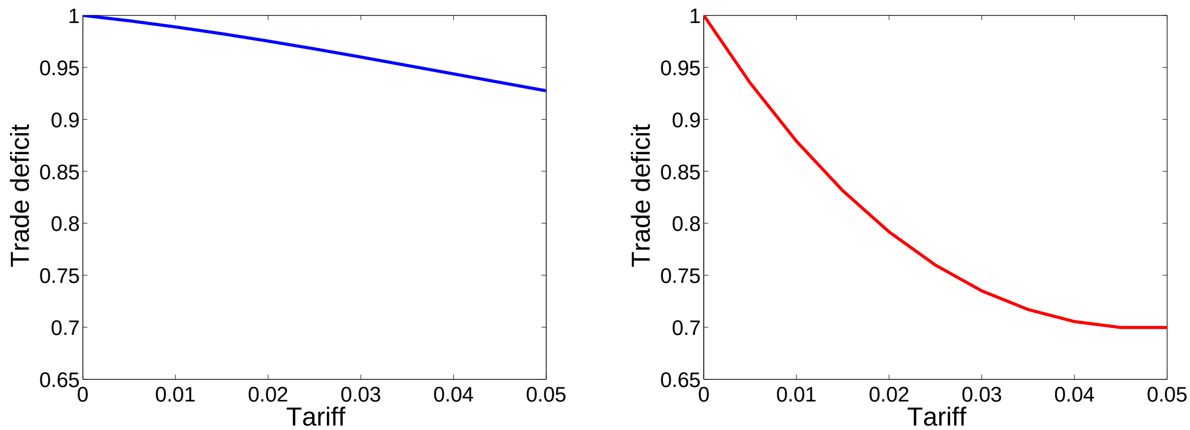
(a) Canada (b) Canada

(c) Mexico (d) Mexico

(e) United States (f) United States

(g) Non-NAFTA countries (h) Non-NAFTA countries

Figure 6: U.S. imports (Left panel – full information, Right panel – rational inattention)



(a) United States (b) United States

Figure 7: U.S. trade deficit (Left panel – full information, Right panel – rational inattention)

the current most-favoured nation (or MFN) tariff for the U.S. is around 3.7 percent, this is a useful counterfactual.[32]

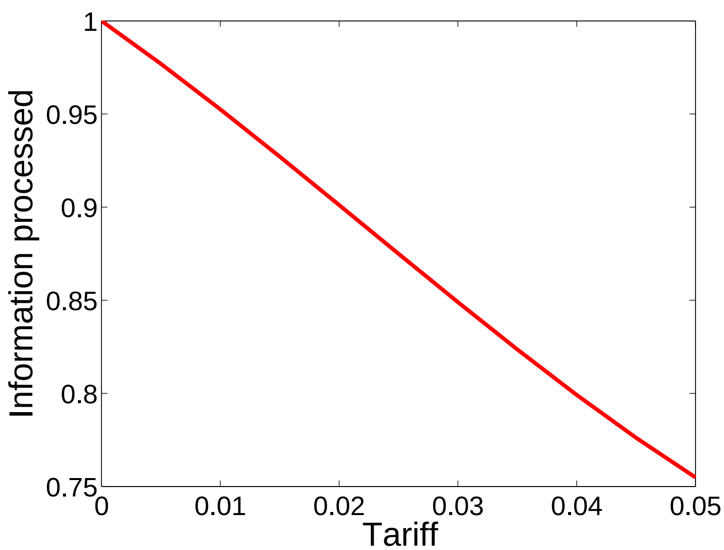


Figure 8: κ for United States

The effect on trade flows is shown in Figure 6. In the first three rows, we plot the percentage change in the shares of U.S. imports coming from Canada, Mexico and the U.S. for different values of the tariff. The left panel shows the results for the full information model, while the right panel shows the results for the model with inattention. Recall that the parameters in the two

[32]United Nations Conference on Trade and Development (UNCTAD).

models are chosen to match the same set of moments. For each plot, we have normalized the shares corresponding to zero tariff at 1, so that the observations give us percentage changes from the no-tariff case. For an increase in tariff by 2 percent, the import share from Canada declines by around 10 percent in the full information model, while it declines by more than 40 percent in the model with inattention. Similarly, the import share from Mexico declines by 10 percent in the full information model, while it declines by more than 25 percent in the model with inattention. Finally, the share of goods that U.S. importers purchase from the U.S. itself rises by 1.5 percent and 6 percent in the two models respectively. The last row of Figure 6 shows that the large declines in imports from Canada and Mexico is accompanied by large increases in imports from the rest of the world (non-NAFTA countries). The bottom line is that there are large differences in the response of trade shares to an increase in tariffs among NAFTA member countries across different models.

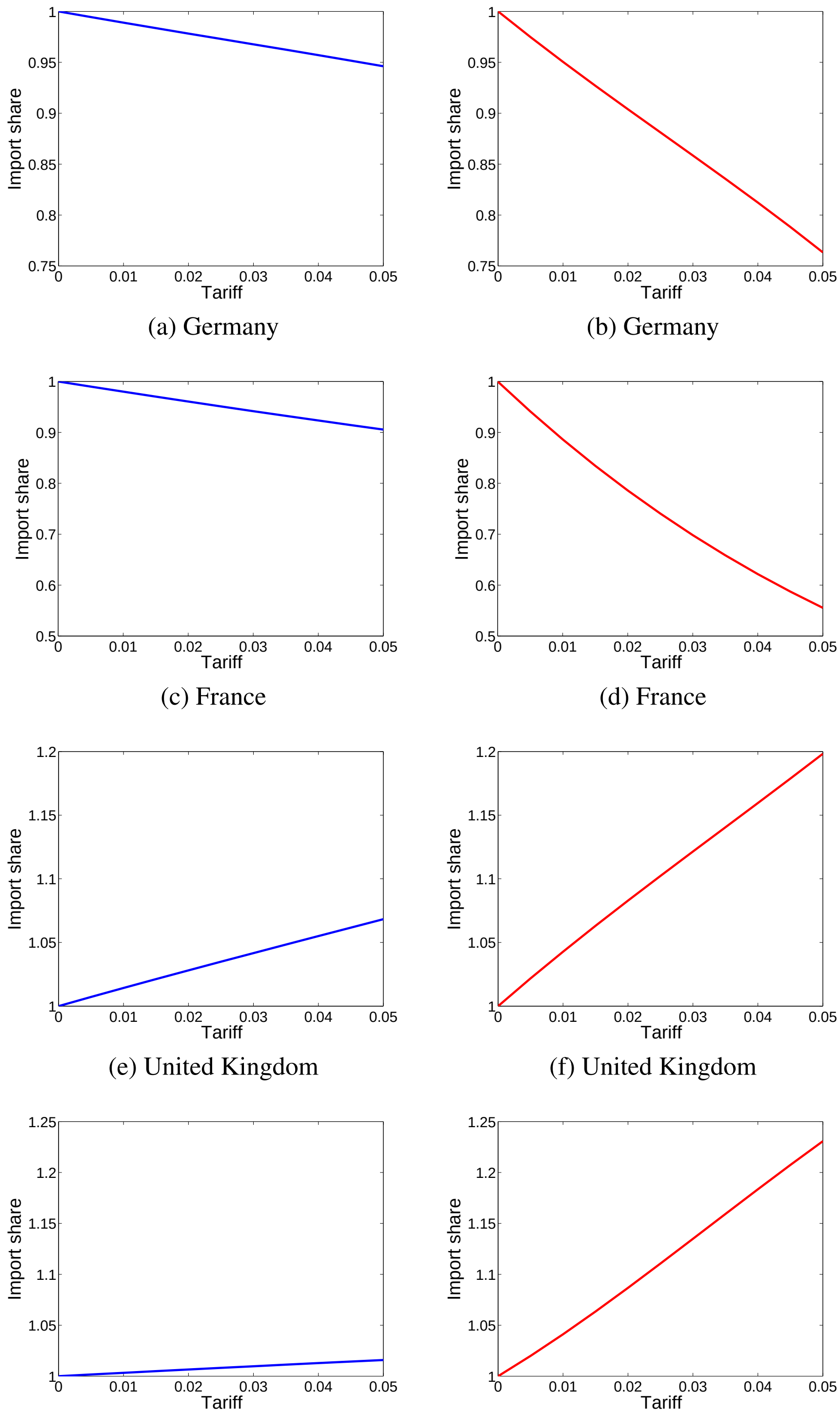
One of the stated objectives of the increasing protectionist position of the current U.S. administration is to reduce the large trade deficit.[33] Because wages are fixed, most countries in our model are either running trade surpluses or trade deficits. We can look at how the U.S. trade balance evolves with the termination of NAFTA in the two models. This is shown in Figure 7, where the trade balance with no tariffs is normalized to 1. In equilibrium, the U.S. is running a trade deficit with rest of the world. An increase in tariffs reduces the trade deficit in both models. But the reduction is significantly larger under inattention. This is not surprising given the large reduction in import shares from Canada and Mexico and the corresponding increase in domestic purchases. Although import shares from the rest of the world increase sharply under inattention, the absolute level of imports from the rest of the world is small and has little effect on the trade deficit.

Figure 8 displays how the information processed, κ, by U.S. importers changes with the tariff. A rise in tariff leads to a steady decline in κ. As the expected price of purchasing goods from Mexico and Canada rise, U.S. importers optimally choose to pay less attention to those countries and instead purchase more from the U.S. As a result, an increase in trade costs leads to a large decline in trade flows – the magnification effect we mentioned in the Section 3.

5.3 Counterfactual 2: Brexit

On June 23, 2016, the United Kingdom voted to leave the European Union, an event known as Brexit. Currently, the British government and their E.U. counterparts are engaged in negotiating the terms of exit. One of the issues is whether U.K. will continue to have the same access to

[33] In 2016, the total U.S. trade deficit was $502 billion.



(a) Germany (b) Germany

(c) France (d) France

(e) United Kingdom (f) United Kingdom

(g) Non-EU countries (h) Non-EU countries

Figure 9: U.K. imports (Left panel – full information, Right panel – rational inattention)

the E.U. common market as it had pre-Brexit.[34] In the absence of a deal, the U.K. would have to operate under World Trade Organization rules, which would mean not only tariffs, but also customs checks as well as the possibility of costly trade disputes. To explore the impact of Brexit on trade, we consider a scenario where tariffs between the U.K. and the rest of the E.U. member countries in our sample rise uniformly across products from 0 to 5 percent. Given that the current MFN tariff in the U.K. is 4.4, we consider this to be a useful exercise.[35] As before, we compare the response of trade flows to tariffs in two models – with full information and with inattention.

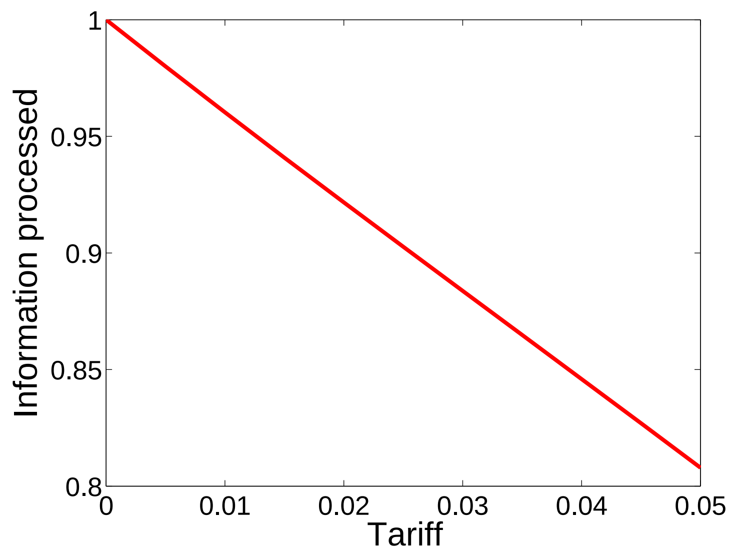


Figure 10: κ for United Kingdom

Figure 9 displays the results. The left-hand side panel plots the response of the tariff on imports into the U.K. from Germany, France and the U.K. itself under full information. The right-hand side plots the corresponding responses under inattention. An across-the-board increase in U.K. tariffs by 5 percent causes import shares from Germany to fall by about 5 percent in the full information model. Under inattention, the decline is almost 25 percent, five times larger than in the model with full information. The results are similar for France, a 5 percent increase in tariffs generates a reduction of import shares by 9 percent and 45 percent in the model with and without full information respectively. Finally, the share of goods that U.K. imports from itself rises by 20 percent under inattention, compared to around 7 percent under full information. As in the case of U.S. in the previous exercise, as tariffs increase, importers in the U.K. process less information and reduce the attention allocated to their major trade partners such as Germany and France. This

[34]"No middle way on trade for Brexit Britain", *Financial Times*, September 26, 2017.

[35]United Nations Conference on Trade and Development (UNCTAD)

is displayed in Figure 10.

6 Conclusion

In this paper, we make three main contributions. First, we establish a formal link between trade flows and the costs of processing information. We show how endogenous processing of information by importers generates novel comparative static results involving the probability of importing from a source country and information costs. Second, we provide evidence that the import distribution for a product is, on average, less concentrated in countries that have intermediate costs of processing information. While none of the standard full-information models predict a systematic variation in the import concentration across countries, our model does. And third, calibrating the information cost parameter in the model, we show quantitatively how small changes in trade costs get translated into large changes in trade flows.

In their survey on the resistance to globalization, Head and Mayer [2013] point out that at most 30 percent of the variation in trade flows can be explained by observable freight costs – the remaining 70 percent of the variation is a "dark" trade cost. We believe that in order to shed light on these "dark" costs, we need a better understanding of the role of information in facilitating trade. Borrowing from the theory of rational inattention, we have developed a framework that allows us to do just that. Much needs to be done, however.

In a recent paper, Dickstein and Morales [2015] ask a related but slightly different question: what is it that exporters know? Dickstein and Morales show that exporters typically have information on a very limited set of variables – distance to a destination, aggregate exports to that destination in the previous year and own productivity in the previous year. Accordingly, their expectations of future profits and consequently, entry decisions, are based on information sets that are far from full. In the context of our framework, the finding of Dickstein and Morales raise interesting questions: if exporters are rationally inattentive, then what is the optimal signal? Is focussing on the above-mentioned variables optimal? Or can exporters do better, given their information processing constraints? We leave the answers for future work.

References

Treb Allen. Information frictions in trade. *Econometrica*, 82(6):2041–2083, 2014.

James E. Anderson and Eric van Wincoop. Trade costs. *Journal of Economic Literature*, 42: 691–751, 2004.

Costas Arkolakis, Jonathan Eaton, and Samuel Kortum. Staggered adjustment and trade dynamics. mimeo, Yale University, 2012.

Matias Berthelon and Caroline Freund. On the conservation of distance in international trade. *Journal of International Economics*, 75(2):310–320, 2008.

N. Scott Cardell. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, 13(02):185–213, 1997.

Jeffrey Chan. Soldiers and the structure of trade: Evidence from 50 years of U.S. military deployment. University of Toronto, mimeo, 2016.

Thomas Chaney. The network structure of international trade. *The American Economic Review*, 104(11):3600–3634, 2014.

David Coe, Arvind Subramanian, and Natalia Tamirisa. The missing globalization puzzle: Evidence of the declining importance of distance. *IMF Staff Papers*, 54:34–58, 2007.

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

Michael J Dickstein and Eduardo Morales. What do exporters know? Technical report, National Bureau of Economic Research, 2015.

Anne-Célia Disdier and Keith Head. The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics*, 90(1):37–48, 2008.

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

Carsten Fink, Aaditya Mattoo, and Ileana Cristina Neagu. Assessing the impact of communication costs on international trade. *Journal of International Economics*, 67(2):428–445, 2005.

Caroline L Freund and Diana Weinhold. The effect of the internet on international trade. *Journal of International Economics*, 62(1):171–189, 2004.

David M Gould. Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. *The Review of Economics and Statistics*, 76(2):302–16, 1994.

Gene Grossman. Comment on Deardorff. In J.A. Frankel, editor, *The Regionalization of the World Economy*, pages 33–57. U. Chicago Press, 1998.

Keith Head and Thierry Mayer. What separates us? Sources of resistance to globalization. *Canadian Journal of Economics*, 46(4):1196–1231, 2013.

Keith Head and John Ries. Immigration and trade creation: Econometric evidence from Canada. *Canadian Journal of Economics*, 31(1):47–62, 1998.

David Hummels. Transportation costs and international trade in the second era of globalization. *The Journal of Economic Perspectives*, 21(3):131–154, 2007.

Jean Imbs and Romain Wacziarg. Stages of diversification. *The American Economic Review*, 93 (1):63–86, 2003.

Daniel Kahneman. *Attention and effort*. Prentice-Hall series in experimental psychology. Prentice-Hall, 1973.

P Klenow and A Rodriguez-Clare. Quantifying variety gains from trade liberalization. mimeo, University of Chicago, 1997.

Sebastian Krautheim. Heterogeneous firms, exporter networks and the effect of distance on international trade. *Journal of International Economics*, 87(1):27–35, 2012.

Nuno Limao and Anthony J. Venables. Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review*, 15(3):pp. 451–479, 2001.

Yulei Luo. Consumption dynamics under information processing constraints. *Review of Economic Dynamics*, 11(2):366–385, 2008.

Yulei Luo and Eric R Young. Rational inattention and aggregate fluctuations. *The BE Journal of Macroeconomics*, 9(1), 2009.

Bartosz Mackowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *The American Economic Review*, 99(3):769–803, 2009.

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, 105(1):272–298, 2014.

Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

Ernesto Pasten and Raphael Schoenle. Rational inattention, multi-product firms and the neutrality of money. *Journal of Monetary Economics*, 80:1 – 16, 2016.

James E. Rauch. Networks versus markets in international trade. *Journal of International Economics*, 48(1):7 – 35, 1999.

James E. Rauch and Vitor Trindade. Ethnic Chinese networks in international trade. *Review of Economics and Statistics*, 84(1):116–130, 2002.

Claude Elwood Shannon. A mathematical theory of communication. *Bell Technical Journal*, 3 (27):379–423, 1948.

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

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50 (3):665 – 690, 2003.

Christopher A. Sims. Rational inattention: Beyond the linear-quadratic case. *The American Economic Review*, 96(2):158–163, 2006.

Claudia Steinwender. Information frictions and the Law of One Price: When the States and the Kingdom became United. Forthcoming, *The American Economic Review*.

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

Appendix A

Proof of Proposition 1. This proof follows in the steps of the proof of Theorem 1 in Matejka and McKay [2014]. If $\lambda_j > 0$, then the Lagrangian of importers in country j is given by

$$\mathcal{L} = \sum_{i=1}^{N} \int_Z u(\tilde{z}_{ij}) f_{ij}(Z)\, dG(Z) + \\ - \lambda_j \left[-\sum_{i=1}^{N} \pi_{ij} \ln \pi_{ij} + \int_Z \left(\sum_{i=1}^{N} f_{ij}(Z) \ln f_{ij}(Z) \right) dG(Z) \right] + \\ + \int_Z \xi_{ij}(Z) f_{ij}(Z)\, dG(Z) - \int_Z \mu(Z) \left(\sum_{i=1}^{N} f_{ij}(Z) - 1 \right) dG(Z)$$

where $\xi_{ij}(Z) \geq 0$ and $\mu(Z) \geq 0$ are the Lagrange multipliers of equations (6) and (7) respectively. If $\pi_{ij} > 0$, the first order condition with respect to $f_{ij}(Z)$ is given by

$$u(\tilde{z}_{ij}) + \xi_{ij}(Z) - \mu(Z) + \lambda_j \left(\ln \pi_{ij} + 1 - \ln f_{ij}(Z) - 1 \right) = 0.$$

As (6) does not bind, then the first order condition can be re-arranged to yield

$$f_{ij}(Z) = \pi_{ij} e^{(u(\tilde{z}_{ij}) - \mu(Z))/\lambda_j} \tag{10}$$

Plugging (10) into (7), we obtain

$$e^{\mu(Z)/\lambda_j} = \sum_{i=1}^{N} \pi_{ij} e^{u(\tilde{z}_{ij})/\lambda_j}$$

If we plug this expression back into (10), we get (8). Equation (8) holds even for $\pi_{ij} = 0$, as otherwise equation (4) would not hold. □

Proof of Proposition 2. Part (i) is trivial. For part (ii), divide (8) by $e^{u(\tilde{z}_{ij})/\lambda_j}$ to obtain

$$f_{ij}(Z) = \frac{\pi_{ij}}{\pi_{ij} + \sum_{k \neq j} \pi_{kj} e^{\frac{1}{\lambda_j}\left[u(\tilde{z}_{kj}) - u(\tilde{z}_{ij}) \right]}}.$$

Suppose $\tilde{z}_{ij} = max \tilde{z}_{kj} \forall k$. Then, as $\lambda_j \to 0$, $\forall k\ \frac{1}{\lambda_j}\left[u(\tilde{z}_{kj}) - u(\tilde{z}_{ij}) \right] \to -\infty$. It follows that in this case,

$$f_{ij}(Z) \to \frac{\pi_{ij}}{\pi_{ij}} = 1.$$

If, on the other hand, suppose $\exists h$ such that $\tilde{z}_{hj} > \tilde{z}_{ij}$. Then, as $\lambda_j \to 0$, $\frac{1}{\lambda_j}\left[log(\tilde{z}_{hj}) - u(\tilde{z}_{ij})\right] \to \infty$. In this case,

$$f_{ij}(Z) \to \frac{\pi_{ij}}{\infty} = 0.$$

□

Proof of Lemma 1. Because there is a measure one of symmetric varieties, the average expenditure on product ω is simply

$$E[X_{ij}(\omega)] = X_{ij}. \tag{11}$$

Let the import share for product ω (in value terms) be denoted by $s_{ij}(\omega)$. Then we can write,

$$X_{ij}(\omega) = s_{ij}(\omega)X_j(\omega),$$

where $X_j(\omega)$ is total expenditure by j on product ω. Now under *trade separability*, the expenditure of j on ω is independent of the allocation of this expenditure across different source countries. Hence, we can write

$$E[X_{ij}(\omega)] = E[s_{ij}(\omega)]E[X_j(\omega)]. \tag{12}$$

Now,

$$E[s_{ij}(\omega)] = E[s_{ij}(\omega)|_{\text{importer in } j \text{ buys } \omega \text{ from } i}] \cdot w_{ij} + E[s_{ij}(\omega)|_{\text{importer in } j \text{ does not buy } \omega \text{ from } i}] \cdot (1 - w_{ij}),$$

where w_{ij} is the fraction of importers in j who buy ω from i. Using a Law of Large Numbers, we have

$$\begin{aligned} w_{ij} &= \int f_{ij}(Z)dZ, \\ &= \pi_{ij}. \end{aligned}$$

Replacing this relation above, we have

$$E[s_{ij}(\omega)] = \pi_{ij}.$$

Furthermore, under Cobb-Douglas preference (a form of trade separable utility function), the expenditure shares are constant, i.e.,

$$E[X_j(\omega)] = X_j(\omega) = X_j,$$

where X_j is aggregate expenditure by j.[36] Therefore (12) can be re-written as

$$E[X_{ij}(\omega)] = \pi_{ij} X_j.$$

Replacing this in (11) and re-arranging, we have

$$\frac{X_{ij}}{X_j} = \pi_{ij}.$$

□

Proof of Proposition 3. This proof follows in the steps of the proof of Lemma 2 and Proposition 3 in Matejka and McKay [2014]. Note that the optimization problem of consumers in country j can be equivalently formulated as a maximization over the unconditional probabilities, $\{\pi_{ij}\}_{i=1}^{N}$:

$$\max_{[\pi_{ij}]_{i=1}^{N}} \int_{\tilde{Z}} \lambda_j \ln \left(\sum_{i=1}^{N} \pi_{ij} e^{u(\tilde{z}_{ij})/\lambda_j} \right) dG(Z) \tag{13}$$

subject to (6) and (7). To see this, substitute equation (5) into the objective function to get

$$\sum_{i=1}^{N} \int_{Z} u(\tilde{z}_{ij}) f_{ij}(Z)\, dG(Z) + \lambda_j \left[\sum_{i=1}^{N} \pi_{ij} \ln \pi_{ij} - \int_{Z} \left(\sum_{i=1}^{N} f_{ij}(Z) \ln f_{ij}(Zv) \right) dG(Z) \right]$$

Rearranging this expression and using (8), we obtain

$$\begin{aligned}
&= \int_{Z} \sum_{i=1}^{N} f_{ij}(Z) \left[u(\tilde{z}_{ij}) - \lambda_j \ln \left(\frac{\pi_{ij} e^{u(\tilde{z}_{ij})/\lambda_j}}{\sum_{k=1}^{N} \pi_{kj} e^{u(\tilde{z}_{kj})/\lambda_j}} \right) \right] dG(Z) + \lambda_j \sum_{i=1}^{N} \pi_{ij} \ln \pi_{ij} \\
&= \int_{Z} \sum_{i=1}^{N} f_{ij}(Z) \lambda_j \left[-\ln \pi_{ij} + \ln \left(\sum_{k=1}^{N} \pi_{kj} e^{u(\tilde{z}_{kj})/\lambda_j} \right) \right] dG(Z) + \lambda_j \sum_{i=1}^{N} \pi_{ij} \ln \pi_{ij} \\
&= \int_{Z} \sum_{i=1}^{N} f_{ij}(Z) \lambda_j \ln \left(\sum_{k=1}^{N} \pi_{kj} e^{u(\tilde{z}_{kj})/\lambda_j} \right) dG(Z) \\
&= \int_{Z} \lambda_j \ln \left(\sum_{k=1}^{N} \pi_{kj} e^{u(\tilde{z}_{kj})/\lambda_j} \right) dG(Z)
\end{aligned}$$

When we include the constraint (7) into the objective function (13), the optimization problem

[36] $X_j = \int X_j(\omega) d\omega = X_j(\omega)$ due to symmetry and measure one of varieties.

of consumers in country j can be rewritten as

$$\max_{[\pi_{ij}]_{i=1}^{N}} \int_Z \lambda_j \ln \left[\sum_{i=1}^{N-1} \pi_{ij} e^{u(\tilde{z}_{ij})/\lambda_j} + \left(1 - \sum_{i=1}^{N-1} \pi_{ij}\right) e^{u(z_{Nj})/\lambda_j} \right] dG(Z),$$

subject to (7). Let us focus on the case where the constraint (6) is not binding because that is a non-trivial case. The gradient of the objective function with respect to π_{1j} is given by

$$\Delta_1 \equiv \lambda_j \int_Z \frac{e^{u(\tilde{z}_{1j})/\lambda_j} - e^{u(\tilde{z}_{Nj})/\lambda_j}}{\sum_{i=1}^{N} \pi_{ij} e^{u(\tilde{z}_{ij})/\lambda_j}} dG(Z),$$

where $\pi_{Nj} = 1 - \sum_{i=1}^{N-1} \pi_{ij}$.

1. Differentiating with respect to either c_1 or τ_{1j} leads to $\frac{\partial \Delta_1}{\partial c_1} < 0$ or $\frac{\partial \Delta_1}{\partial \tau_{1j}} < 0$ respectively. This establishes that at the original optimum, an increase in either c_1 or τ_{1j} leads to a decrease of the gradient of the objective function with respect to the probability of the first option. Thus, consumers in country j will decrease π_{1j}.

2. When $\lambda_j \to \infty$, importers in j process no information and decisions are based on *ex-ante* expectations. Given that country i has lowest expected price, *ex-ante* expected $u(\tilde{z}_{ij})$ is the highest and $\pi_{ij} \to 1$.

3. If countries are *ex-ante* identical, $c_i = c; \tau_{ij} = \tau$ for all i, then $G(Z)$ is invariant to permutations of its arguments. Therefore, as showed by Matejka and McKay [2014], the solution for unconditional probabilities is unique and given by $\pi_{ij} = 1/N$ for all i. □

Appendix B

A general cost formulation: In this appendix, we show that a more general formulation for trade costs can generate the prediction that an increase in costs can cause asymmetric changes in import shares.

An importer can use resources (labour, capital, time, effort) to reduce the cost of imports. One can think of an importer using resources to improve the search process for foreign suppliers, improve screening mechanisms for product quality, set up more efficient distribution networks, etc. Any such measure reduces the marginal cost of the imported good. The important assumption is that importers can choose how much resources to assign to products from *each* country, subject to a resource constraint.

Let us assume that products are differentiated by their country of origin (the Armington assumption). Importers in country j have CES preference over products exported by every country i. The aggregate price index in j is then given by

$$P_j^{1-\sigma} = \int_i p_{ij}^{1-\sigma} di,$$

where p_{ij} is the price of country i's product in country j and $\sigma > 1$ is the elasticity of substitution. Let the price of a product be given by

$$p_{ij} = f(\hat{p}_{ij}, \gamma_{ij}),$$

where γ_{ij} is the amount of resources used by an importer in j to reduce the cost of importing a product from country i and $\hat{p}_{ij}$ is the price that the importer pays if he devotes an arbitrarily large amount of resources towards country i, i.e., $p_{ij} \to \hat{p}_{ij}$ as $\gamma_{ij} \to \infty$. We assume that $\frac{\partial f}{\partial \gamma_{ij}} < 0$. The importer's problem then reduces to

$$\max_{\gamma_{ij}} \int_i f(\hat{p}_{ij}, \gamma_{ij})^{1-\sigma} di$$

subject to a resource constraint:

$$\int_i \gamma_{ij} di = \bar{\gamma}_j.$$

Observe that given the preferences, the relative expenditure on products from two countries will depend on the relative prices. Next, we show that if there is an increase in the cost of resources, which we interpret as a reduction in $\bar{\gamma}_j$, the relative prices might change, causing the relative expenditure to change as well. To this end, we make a functional form assumption about f. In particular, we assume that $p_{ij} = \hat{p}_{ij} + \frac{1}{\gamma_{ij}}$. For simplicity, let us also assume that country j is importing from two countries: 1 and 2. The first-order condition for cost minimization then yields

$$\frac{1}{\gamma_{1j}^2}\Big(\hat{p}_{1j} + \frac{1}{\gamma_{1j}}\Big)^{-\sigma} = \frac{1}{\gamma_{2j}^2}\Big(\hat{p}_{2j} + \frac{1}{\gamma_{2j}}\Big)^{-\sigma}.$$

Combining the previous expression with the resource constraint, $\gamma_{1j} + \gamma_{1j} = \bar{\gamma}_j$, we have

$$\frac{\gamma_{1j}}{\gamma_{2j}} = \frac{(\bar{\gamma}_j - \gamma_{1j})\hat{p}_{2j} + 1}{\hat{p}_{1j} + 1}.$$

It can be shown that as long as $\hat{p}_{1j} \neq \hat{p}_{2j}$, a reduction in $\bar{\gamma}_j$ will cause the right-hand side of the above equation to change. The first-order condition implies that

$$\frac{p_{1j}}{p_{2j}} = \frac{\hat{p}_{1j} + \frac{1}{\gamma_{1j}}}{\hat{p}_{2j} + \frac{1}{\gamma_{2j}}} = \left(\frac{\gamma_{2j}}{\gamma_{2j}}\right)^{\frac{1}{\sigma}}.$$

Hence, in this setting, as the resource constraint changes, the importer in country j will change the relative allocation of resources to each country, thereby altering relative sales. Note that this is similar to what happens in our model of inattention as the cost of processing information changes. The information cost implicitly defines a capacity constraint. As the information cost rises, and the capacity to process information declines, the importer re-allocates attention across source countries, causing relative import shares to change.

Appendix C

In this section, we make additional assumptions to derive a closed-form solution for π_{ij}. First, we set $u(\tilde{z}_{ij}) = \ln(\tilde{z}_{ij})$. Second, we impose $\lambda_j = \lambda$ for all j. Third, we use a specific form for $G(\tilde{Z})$. Following Cardell [1997], we define a distribution $C(\beta)$.

Definition. For $0 < \beta < 1$, $C(\beta)$ is a distribution with a probability density function given by

$$g_\beta(z) = \frac{1}{\beta} \sum_{n=0}^{\infty} \left[\frac{(-1)^n e^{-nz}}{n!\Gamma(-\beta n)}\right].$$

The main property of the $C(\beta)$ distribution is that if a random variable ϵ is drawn from a Type I extreme value (Gumbel) distribution and another random variable ν is drawn from $C(\beta)$, then $\nu + \beta\epsilon$ is a random variable distributed as Type I extreme value. The relation between $C(\beta)$ and a Gumbel distribution is shown in Figure 11.[37] It is clear that qualitatively, the two distributions are very similar. The next result shows that when the logarithm of the random productivities is distributed as $C(\beta)$ and under a very restrictive parameter constraint, there exists a closed-form solution for π_{ij}.

RESULT 1 If $\beta = \lambda$ and $\ln(\tilde{z}_i)$ is drawn independently from a cumulative distribution $C(\lambda)$

[37] For this plot, we choose $\beta = 0.5$.

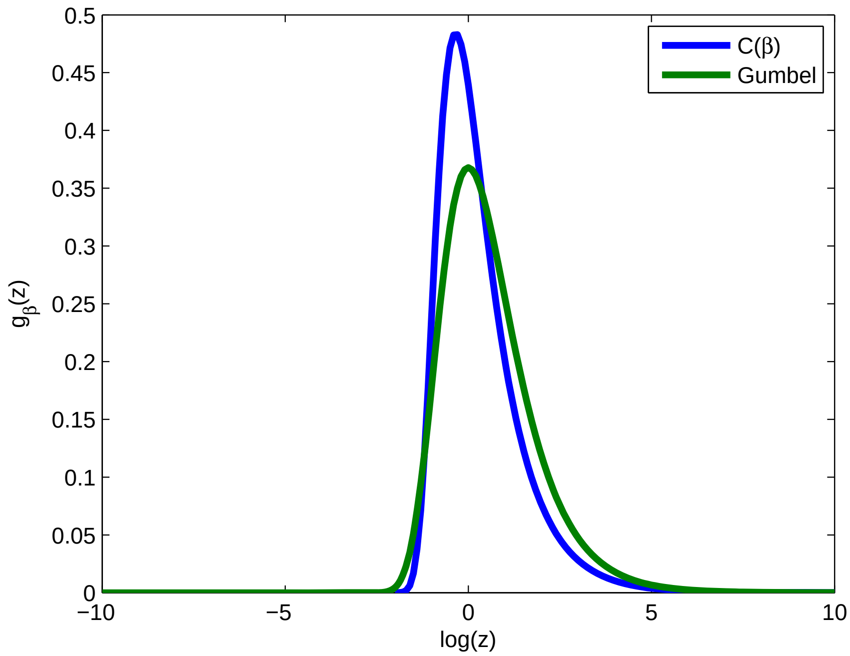


Figure 11: Distribution $C(\beta)$

where $0 < \lambda < 1$, then π_{ij} is given by

$$\pi_{ij} = \frac{(c_i \tau_{ij})^{\frac{-1}{1-\lambda}}}{\sum_{h=1}^{N} (c_h \tau_{hj})^{\frac{-1}{1-\lambda}}}.$$

The *unconditional* probability that country j buys the manufactured good from country i has a form that also bears resemblance to a multinomial logit. The next result describes some properties of π_{ij}. Without loss of generality, we assume that for a given country j, the exporting countries are ordered with respect to $\frac{1}{c_i \tau_{ij}}$ with $\frac{1}{c_1 \tau_{1j}}$ being the largest and $\frac{1}{c_N \tau_{Nj}}$ being the smallest.

RESULT 2 The closed-form solution of π_{ij} has the following additional properties:

1. $\frac{\partial \ln \pi_{ij}}{\partial \ln \tau_{ij}}$ (the trade elasticity) is increasing in λ.
2. π_{1j} is monotone increasing while π_{Nj} is monotone decreasing in λ. For any other i, π_{ij} has a hump-shape as a function of λ.

The parameter λ has a dual role in the closed-form solution of π_{ij} as it governs both the cost of processing information and the shape of the productivity distribution in each country. Therefore, one needs to be careful in interpreting the comparative static exercises involving λ as the results are driven by both changes in the shape of the productivity distribution and changes in information costs. The theoretical results, however, confirm the numerical results that we have presented before, where we varied the information cost, holding productivity dispersion constant.

Appendix D

Introducing dynamics to a model with information choices is a complicated task. Under a specific set of assumptions, we could solve a dynamic model. We need two assumptions: (i) z_{it} follows an autoregressive process such that $z_{it} = \phi z_{it-1} + \epsilon_{it}$ and (ii) the vector with the productivity draws Z_t of all countries at time t is observable at the end of the period. These assumptions imply that at each time t, importers in country j solve the following optimization problem:

$$\max_{F, i \in C} E\big[u(\tilde{z}_{ijt}) - \lambda_j \kappa_j\big]$$

where $\tilde{z}_{ijt} = \frac{\phi z_{it-1}}{c_i \tau_{ij}} + \frac{\epsilon_{it}}{c_i \tau_{ij}}$, $\kappa_j = I(Z_t; S_t, Z_{t-1}) = H(Z_t) - E_{S_t}\big[H\big(Z_t|S_t, Z_{t-1}\big)\big]$ and the expectation is taken with respect to the distribution over (Z_t, S_t) induced by the prior G. The model and its solution would be identical as the static model except that $\tilde{z}_{ijt}$ would now have two components. The first component $\frac{\phi z_{it-1}}{c_i \tau_{ij}}$ would be a constant that depends on the previous period's productivity realizations. This component does not appear in the static version of the model and it would help realize *ex-post* mistakes done in $t-1$ when choosing the source country to import a product from. Since z_{it-1} is observed at time t, importers would be able to learn from their mistakes and the model would move towards an Eaton-Kortum (EK) model. The second component $\frac{\epsilon_{it}}{c_i \tau_{ij}}$, however, is equivalent to the $\tilde{z}_{ij}$ in the static model. If $\phi = 0$, then importers would face a static optimization problem at each t given by this second component. This second component of $\tilde{z}_{ijt}$ implies that importers always make mistakes in the dynamic version and it prevents importers to fully learn which country is offering the lowest price. Therefore, in a dynamic model with these two additional assumptions, we would converge to the EK model, but we would never reach full convergence.

How close the dynamic model is to the EK model or the model with inattention will depend on parameters. Two key sets of parameters are ϕ, and the parameters governing the distribution of ϵ_{it}. A large ϕ will give previous realizations of productivity a large weight. In this case, the first component of $\tilde{z}_{ijt}$ will be important and the dynamic model will be closer to the EK model. On the other hand, parameters that lead to a large dispersion in ϵ_{it} will provide a larger role to information frictions. In this case, the second component of $\tilde{z}_{ijt}$ will be important and the dynamic model will be closer to the model with inattention.