The More We Die, The More We Sell?
A Simple Test of the Home-Market Effect[*]

Arnaud Costinot
MIT, CEPR, and NBER

Dave Donaldson
Stanford, CEPR, and NBER

Margaret Kyle
Mines ParisTech and CEPR

Heidi Williams
MIT and NBER

June 2016

Abstract

The home-market effect, first hypothesized by Linder (1961) and later formalized by Krugman (1980), is the idea that countries with larger demand for some products at home tend to have larger sales of the same products abroad. In this paper, we develop a simple test of the home-market effect using detailed drug sales data from the global pharmaceutical industry. The core of our empirical strategy is the observation that a country's exogenous demographic composition can be used a predictor of the diseases that its inhabitants are most likely to die from and, in turn, the drugs that they are most likely to demand. We find that the correlation between predicted home demand and sales abroad is positive and greater than the correlation between predicted home demand and purchases from abroad. In short, countries tend to be net sellers of the drugs that they demand the most, as predicted by Linder (1961) and Krugman (1980).

1 Introduction

Do countries with larger domestic markets for some products tend to sell more of the same products in foreign markets? The idea that local demand may stimulate exports is an old one. First hypothesized by Linder (1961) and later formalized by Krugman (1980), the so-called home-market effect has become a central tenet of the new trade theory (Helpman and Krugman, 1985, 1989) and the new economic geography literature (Fujita, Krugman and Venables, 2001). In terms of policy, it implies that import protection may be used as export promotion, a view often more popular in business communities than among economists (Krugman, 1984).

To establish the empirical validity of the home-market effect, one must overcome a key challenge. While theory predicts that the cross-sectional variation in demand conditions should be causing the pattern of international specialization, observable demand shifters are rarely available in practice. National accounts, for instance, may report how much a country spends on a particular good. But expenditures depend on prices, which themselves depend on supply, not just on demand conditions.

In this paper, we propose a simple test of the home-market effect that uses variation in disease burdens across countries as a way to address this empirical challenge. Our starting point is the observation that countries whose populations, because of exogenous demographic characteristics, are more likely to die from particular diseases are also more likely to demand pharmaceutical treatments that target those diseases. Hence, one can test for the existence of the home-market effect by estimating (*i*) whether higher disease burdens at home tend to increase the sales of domestic drugs abroad (*weak home-market effect*), and if so, (*ii*) whether they tend to increase them by more than the sales of foreign drugs at home (*strong home-market effect*).

To take a concrete example, the drug famotidine (Gaster, Pepcid)—used to treat peptic ulcers and gastroesophageal reflex—was discovered in Japan (Hara, 2003), a country known for particularly high incidence rates of peptic ulcers.[1] Indeed, in our data, individuals in Japan are nearly twice as likely to die from digestive disorders than are individuals in the rest of the world (0.243 deaths per 1,000 population annually in Japan, relative to 0.131 on average in other countries). Our data also suggest that the example of famotidine is not an outlier: sales of Japanese drugs targeting peptic ulcers and gastro-esophageal reflux diseases outside Japan account for 21.46% of world sales, compared to an average of 11.01% for all other disease categories. Our empirical work uses this type of variation

---

[1] For example, Cleave (1962) notes that the age-adjusted death rate from peptic ulcers for Japanese males in 1954 was 34.7 per 100,000, which can be compared to 14.1 per 100,000 in England and Wales. Cleave discusses the potential role of dietary habits in explaining this pattern.

in order to test for the home-market effect in a dataset with near-global coverage of drug sales and disease burdens.

The rest of our paper is organized as follows. After discussing the related literature in Section 2, we present a flexible model of drugs supply and demand in Section 3. For expositional purposes, we first study a perfectly competitive environment. In this context, we introduce a simple test of the weak and strong home-market effects based on a log-linear approximation of our model and characterize the conditions for such effects to arise. We then show that the same test remains valid in a range of imperfectly competitive environments, including the one considered in Krugman (1980). Our theoretical analysis highlights the role of sector-level economies of scale, while clarifying that their particular determinants may be irrelevant.

Section 4 describes our data. Our empirical analysis draws on a linkage between two datasets. The first one documents sales in 56 countries of more than 20,000 molecules by roughly 2,800 firms, which we convert to a dataset of bilateral sales at the disease level, by matching each firm to the country in which it is headquartered and each molecule to the disease that it targets.[2] The second dataset documents the demographic composition of and disease burdens in the previous 56 countries, which we use to compute predicted disease burdens by country and disease.

Section 5 presents our main results. Our simple test focuses on a log-linear specification where bilateral sales of drugs targeting different diseases are allowed to depend on disease burdens in the destination country, i.e., the country where drugs are sold; disease burdens in the origin country, i.e., the country where firms selling those drugs are headquartered; and a full vector of disease indicator variables and destination-and-origin indicator variables. Everything else equal, we document that countries tend to sell relatively more of the drugs for which they have higher disease burdens, in line with the existence of a weak home-market effect. Furthermore, the elasticity of sales towards foreign countries tends to be higher than the elasticity of purchases from foreign countries, in line with the existence of a strong home-market effect.

Section 6 analyzes further the economic determinants of the home-market effect. While the previous results provide empirical support for the notion of a home-market effect in the global pharmaceutical sector, the existence and magnitude of this phenomenon depend, according to our model, both on demand and supply elasticities. Our last results point towards the home-market effect being driven by substantial economies of scale at

---

[2]Our dataset does not contain information about location of production. Thus, we cannot shed light on whether the home-market effect ultimately operates through exports, foreign direct investment, or a mixture of both. We come back to this point when discussing the related literature in Section 2.

the sector-level rather than low elasticities of demand. Quantitatively, though, the sector-level economies of scale that we estimate in the pharmaceutical industry are about 25% smaller than those that Krugman's (1980) monopolistically competitive model predicts.

Finally, section 7 offers some concluding remarks.

2   Related Literature

The literature on the home-market effect is large and varied, in part because different authors use related, but distinct, definitions of "the" home-market effect.

Whereas both Linder's (1961) and Krugman's (1980) original work emphasize the consequences of cross-country differences in demand for the pattern of trade, Helpman and Krugman (1985) focus instead on whether larger countries should tend to specialize in sectors with larger economies of scale.[3] Subsequent work by Davis (1998), Holmes and Stevens (2005), and Behrens et al. (2009) provide additional conditions on the nature of trade costs as well as the number of goods and countries under which the latter pattern may or may not arise. Amiti (1998), in turn, studies whether larger countries should have a comparative advantage in sectors with higher trade costs. Motivated by the theoretical predictions of Helpman and Krugman (1985), Hanson and Xiang (2004) show that high-GDP countries tend to sell disproportionately more in sectors with larger transportation costs and lower elasticities of substitution, a measure of sector-level economies of scale under monopolistic competition. In related work, Feenstra, Markusen and Rose (2001) document that high-GDP countries tend to be net exporters of differentiated goods, which they also interpret as evidence of a home-market effect in such industries.

A number of more recent theoretical papers have extended the work of Krugman (1980) to study the implications of non-homothetic preferences for the pattern of trade and foreign direct investment; see Fajgelbaum, Grossman and Helpman (2011, 2015) and Matsuyama (2015). A key prediction of these models is that in the presence of economies of scale, rich countries that have larger demand for high-quality goods will tend to specialize in those goods. As a result, they will trade more with, or invest more in, other rich countries, as also emphasized by Linder (1961). While not strictly speaking about cross-country differences in demand—exogenous income differences are ultimately causing the pattern of trade—-the underlying mechanism is the same as in Krugman (1980). In line with the previous models, Caron, Fally and Fieler (2015) document that the sectors on which high-GDP countries spend more also tend to be the sectors in which high-GDP countries export more. Dingel (2015) also offer empirical evidence consistent with the

---

[3]Ethier (1982) discusses similar issues in a perfectly competitive model with external economies of scale.

previous mechanism using information about shipment prices from different U.S. cities and the income composition of neighboring cities.

Our analysis is most closely related to the early empirical work of Davis and Weinstein (1996) and later studies by Davis and Weinstein (1999, 2003), Lundback and Torstensson (1998), Head and Ries (2001), Trionfetti (2001), Weder (2003), Crozet and Trionfetti (2008), and Brulhart and Trionfetti (2009). Like ours, the aforementioned papers focus on the impact of differences in demand on the pattern of international specialization. In their review of the literature, Head and Mayer (2004) conclude that this type of empirical evidence on the home-market effect is highly mixed.[4] While empirical specifications and data sources vary across studies, the previous papers all share one key characteristic: data on expenditure shares are used as a proxies for demand differences. As argued earlier, one non-trivial issue with such proxies is that differences in local supply conditions may also be affecting expenditure shares through their effect on local prices. This makes earlier tests of the home-market effect hard to interpret.

Compared to earlier work on the home-market effect, we view the approach in this paper as having both costs and benefits. Since the home-market effect emphasized by Linder (1961) and Krugman (1980) is about the causal effect of demand differences across countries, any test of this effect ultimately requires exogenous demand variation. While we have no silver bullet to deal with endogeneity issues, and we discuss those associated with our approach later in the paper, we believe that using (predicted) disease burdens as observable demand shifters rather than expenditure shares is a significant step forward.

The obvious drawback of our empirical strategy is that its scope is narrower. Another limitation of our dataset is that it does not allow us to distinguish between exports and foreign direct investment. We only observe total sales by firms headquartered in a particular country. Thus, the home-market effect that we identify may operate through both exports and foreign direct investment, not just exports, as emphasized in the previous literature. The previous observation notwithstanding, it is not clear that if the only choice was to study either exports or the sum of exports and sales by foreign affiliates, one should prefer the former to the latter. Indeed, the same economic forces are likely to be at play for both types of sales.

---

[4]Given our focus on the pharmaceutical industry, it is worth nothing that Trionfetti's (2001) sector-level test is rejected for "Chemical Products."

3 Theory

For expositional purposes, we first consider a world economy with perfect competition (Section 3.1) and develop a simple test of the home-market effect in this environment (Section 3.2). We then show that the previous test remains valid in a range of imperfectly competitive environments (Section 3.3).

3.1 Basic Environment

Demand Individuals consume drugs that target multiple diseases, indexed by n, as well as other goods, which we leave unspecified. Empirically, each disease n will correspond to a broad disease class like "cardiovascular diseases." We assume that the aggregate consumption of drugs targeting disease n in country j can be expressed as

$$D_j^n = \theta_j^n D(P_j^n / P_j) D_j, \tag{1}$$

where P_j^n is a price index for drugs targeting disease n, to be described below; D_j and P_j are endogenous country-specific demand shifters that are common to all drugs in country j; and θ_j^n is an exogenous disease-and-country-specific demand shifter, which we will later measure using data on disease burdens.

Within each disease category n, drugs may be purchased from different countries. Any of these countries may be producing different versions of the same molecule (e.g. generic versus non-generic), different molecules targeting the same narrow disease (e.g. angiotensin II receptor blockers and beta blockers, both treatments for high blood pressure, a risk factor for hypertensive heart disease), or different molecules targeting different diseases within the same broad category (e.g. drugs targeting hypertensive heart disease vs. coronary artery disease, within the broad category of cardiovascular diseases). The previous considerations suggest imperfect substitutability between drugs from different countries, which we capture through the following specification,

$$d_{ij}^n = d(p_{ij}^n / P_j^n) D_j^n, \tag{2}$$

where d_{ij}^n denotes country j's consumption of varieties from country i targeting disease n, p_{ij}^n denotes the consumer price for these varieties, and P_j^n is given by the solution to

$$P_j^n = \sum p_{ij}^n d(p_{ij}^n / P_j^n). \tag{3}$$

Given the level of aggregation in our empirical analysis, p_{ij}^n should itself be interpreted

as a price index, aggregating prices across all firms from country i selling drugs targeting disease n in country j. We will make this aggregation explicit in Sections 3.3 and 6.1.[5]

Supply   Firms produce up to the point at which drug prices are equal to marginal costs. For each disease n and country i, this leads to a supply curve,

$$s_i^n = \eta_i^n s(p_i^n), \tag{4}$$

where p_i^n denotes the producer price of drugs targeting disease n in country i and η_i^n is a disease-and-country specific supply shifter, which may capture both technological and regulatory differences. Depending on whether there are external economies of scale or not, $s(\cdot)$ may be upward- or downward-sloping. Trade is subject to iceberg frictions. To sell one unit of a given drug to country $j \neq i$, firms from country i must ship $\tau_{ij}^n \geq 1$ units.[6] Without loss of generality, we set $\tau_{ii}^n = 1$ for all i and n. Non-arbitrage implies

$$p_{ij}^n = \tau_{ij}^n p_i^n. \tag{5}$$

Equilibrium   Supply equals demand for each drug,

$$s_i^n = \sum_j \tau_{ij}^n d_{ij}^n. \tag{6}$$

3.2   A Simple Test of the Home-Market Effect

The home-market effect is the general idea that, everything else being equal, countries tend to sell more abroad in sectors for which they have larger domestic markets. Here, we operationalize this idea in the context of a log-linearized version of our model.

Log-linear Specification   Let $x_{ij}^n \equiv p_{ij}^n d_{ij}^n$ denote the equilibrium sales of drugs targeting disease n by firms from country i in country $j \neq i$. Around a symmetric equilibrium with trade costs, $\tau \geq 1$, and common demand and supply shocks across countries and

---

[5]For the purposes of testing the home-market effect, we do not need the previous demand functions to be consistent with the behavior of a representative agent in country j, an assumption that may be particularly strong in a sector where demand involves physicians, pharmacists, insurers, and patients. We note, however, that equations (1)-(3) are consistent with the common assumption of nested CES utility functions.

[6]Though we abstract from multinational production in our baseline model, equations (4) and (5) would still hold in a world economy with multinational activities à la Ramondo and Rodríguez-Clare (2013). In such an environment, τ_{ij}^n would simply correspond to the minimum cost of accessing country j from country i, either through exports or foreign direct investment. This extension can be found in Appendix A.

diseases, we can express bilateral sales, up to a first-order approximation, as

$$\ln x_{ij}^n = \delta_{ij} + \delta^n + \beta_M \ln \theta_j^n + \beta_X \ln \theta_i^n + \varepsilon_{ij}^n, \tag{7}$$

where δ_{ij} is an origin-destination fixed-effect that captures systematic determinants of bilateral trade flows such as physical distance or whether countries i and j share the same language; δ^n is a disease fixed-effect that captures worldwide variation in demand and supply conditions across drugs targeting different diseases; β_M is the elasticity of trade flows with respect to demand shocks in the importing country; β_X is the elasticity of trade flows with respect to demand shocks the exporting country j; and ε_{ij}^n is a residual that captures idiosyncratic variation in trade costs and supply conditions.

The mapping between the previous coefficients and the structural parameters of Section 3.1 can be found in Appendix B. The economic interpretation of β_M and β_X, which is central to our analysis, is discussed in detail below. At this point, it is worth noting that ε_{ij}^n does not depend on θ_l^n in other countries $l \neq i, j$. Hence, we will not need to impose any restriction on the spatial correlation of demand shocks across countries in order to estimate β_M and β_X in Section 5.

Starting from equation (7), we can then express country i's total exports, $X_i^n \equiv \sum_{j \neq i} x_{ij}^n$, and total imports, $M_i^n \equiv \sum_{j \neq i} x_{ji}^n$, of drugs targeting disease n as

$$X_i^n = \exp(\delta^n) \times \left(\sum_{j \neq i} (\theta_j^n)^{\beta_M} \exp(\delta_{ij} + \varepsilon_{ij}^n) \right) \times (\theta_i^n)^{\beta_X}, \tag{8}$$

$$M_i^n = \exp(\delta^n) \times \left(\sum_{j \neq i} (\theta_j^n)^{\beta_X} \exp(\delta_{ji} + \varepsilon_{ji}^n) \right) \times (\theta_i^n)^{\beta_M}. \tag{9}$$

According to equation (8), after controlling for differences in world exports across diseases, as captured by $\exp(\delta^n)$, differences in the "proximity" to large buyers, as captured by $\sum_{j \neq i} (\theta_j^n)^{\beta_M} \exp(\delta_{ij} + \varepsilon_{ij}^n)$, a country tends to export more of the goods for which it has larger domestic demand if and only if $\beta_X > 0$. And according to equation (9), after also controlling for differences in the "proximity" to large sellers, $\sum_{j \neq i} (\theta_j^n)^{\beta_X} \exp(\delta_{ji} + \varepsilon_{ji}^n)$, a country tends to be a net exporter of the goods for which it has a larger domestic market if and only if $\beta_X > \beta_M$. These two observations motivate the following definition.

Definition. *A given cross-section of bilateral sales $\{x_{ij}^n\}$ satisfies the weak home-market effect if $\beta_X > 0$ and the strong home-market effect if $\beta_X > \beta_M$.*

Our definition, while natural in the context of our model, differs from earlier tests of the home-market effect. Three features of our definition are worth emphasizing.

First, and most importantly, it focuses on elasticities with respect to demand shocks, not expenditure shares. If preferences across sectors are Cobb-Douglas, the two elasticities are equivalent. Away from this empirically knife-edge case, they are not. Assuming that observable demand shocks are available, a case that we make in Section 4, using these shocks alleviates concerns about "false positives"—that is, positive correlations between exports and expenditure shares driven by unobserved supply shocks that are positively correlated with both exports and expenditure shares, absent any variation in demand.

Second, our definition focuses on elasticities with respect to a country's own demand, that is, its home market, not its overall market access. As can be seen from equation (8), the variation in demand from neighboring countries is taken into account in our analysis. However, we are only interested in the elasticity of exports with respect to demand shocks after controlling for such variation. This addresses concerns about a positive test of the home-market effect arising because of a mechanical relationship between exports and foreign demand shocks.

Third, our definition introduces the distinction between the weak home-market effect, which focuses on gross exports, and the strong home-market effect, which focuses on net exports. As we argue next, the weak test, which is unique to our paper, provides a direct way to identify departures from the predictions of neoclassical trade models. The strong test merely puts tighter bounds on the magnitude of these departures, if any.

Economic Interpretation. The economic forces that give rise to weak and home-market effects are best illustrated in a world economy comprising a large number of small open economies in the sense that each country is too small to affect the price of foreign varieties, but large enough to affect the price of its own varieties, as in Gali and Monacelli (2005). In this case, the two elasticities, β_X and β_M, simplify into

$$\beta_X = \frac{\lambda(1 - \epsilon^x)}{\epsilon^s + \epsilon^w}, \tag{10}$$

$$\beta_M = 1 + \frac{\lambda^2(1 - \epsilon^d)(\epsilon^x - \epsilon^D)}{(1 - \lambda\epsilon^d - (1 - \lambda)\epsilon^x)(\epsilon^s + \epsilon^w)}, \tag{11}$$

where λ is the share of expenditure, as well as revenue, on domestic drugs in the symmetric equilibrium; $\epsilon^d \equiv -(d \ln d(z)/d \ln z)_{z=1} > 0$ and $\epsilon^x \equiv -(d \ln d(z)/d \ln z)_{z=\tau} > 0$ are the lower-level elasticity of demand for domestic and foreign varieties, respectively; $\epsilon^D \equiv -(d \ln D(z)/d \ln z)_{z=1} > 0$ is the upper-level elasticity of demand; $\epsilon^w \equiv \lambda\epsilon^d + (1 - \lambda)\epsilon^x - \lambda^2(1 - \epsilon^d)(\epsilon^d - \epsilon^D)/(1 - \lambda\epsilon^d - (1 - \lambda)\epsilon^x) > 0$ is the elasticity of world demand; and $\epsilon^s \equiv (d \ln s(z)/d \ln z)_{z=1}$ is the elasticity of supply, which may be positive or
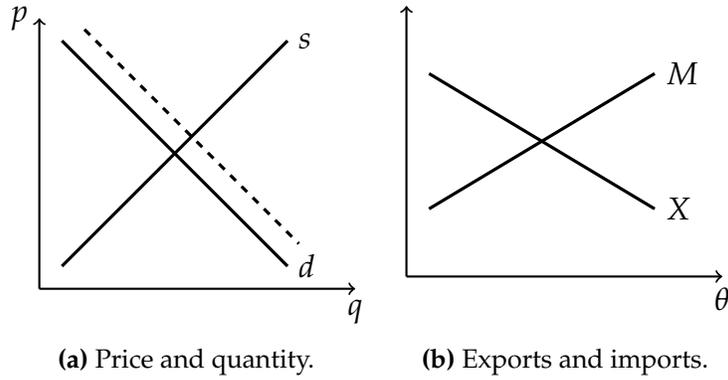
(a) Price and quantity.   **(b)** Exports and imports.

Figure 1: Neoclassical Benchmark.

negative, depending on whether there are economies of scale.[7]

Suppose that $\epsilon^x > 1$ so that countries with lower prices tend to have higher market shares abroad, which will be the empirically relevant case. Then, according to equation (10), there can only be a weak home-market effect in the presence of economies of scale,

$$\epsilon^s < -\epsilon^w < 0.$$

In a neoclassical environment, an increase in domestic demand across sectors, i.e. a positive shift in θ, raises world demand, d, and in turn, producer prices, p, as depicted in Figure 1a. If the price elasticity of exports, ϵ^x, is strictly greater than one, this necessarily lowers the value of exports, X, as depicted in Figure 1b. By lowering the price of goods with larger domestic markets, economies of scale can instead create a positive relationship between exports and domestic demand, as described in Figures 2a and 2b.

Suppose, in addition, that $\epsilon^d > 1$ and $\epsilon^x \geq \epsilon^D$. The second inequality is another mild restriction that requires, for example, French and American drugs targeting cardiovascular diseases to be closer substitutes than drugs targeting cardiovascular and skin diseases. Under this restriction, equations (10) and (11) imply that a strong home-market effect arises if economies of scale are strong enough to dominate the direct effect of domestic demand on imports, namely if

$$- \epsilon^w - \lambda(\epsilon^x - 1 + \lambda(1 - \epsilon^d)(\epsilon^x - \epsilon^D)/(1 - \lambda\epsilon^d - (1 - \lambda)\epsilon^x)) < \epsilon^s < -\epsilon^w. \qquad (12)$$

---

[7]Formally, we obtain the small open economy limit by taking the number of countries in the world economy to infinity and adjusting trade costs, τ, to leave the expenditure share on a country's own good, λ, at a constant and strictly positive level.
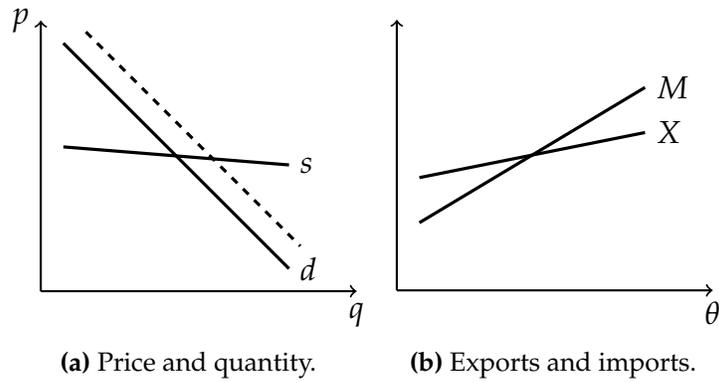
9

(a) Price and quantity. **(b)** Exports and imports.

Figure 2: Weak home-market effect.

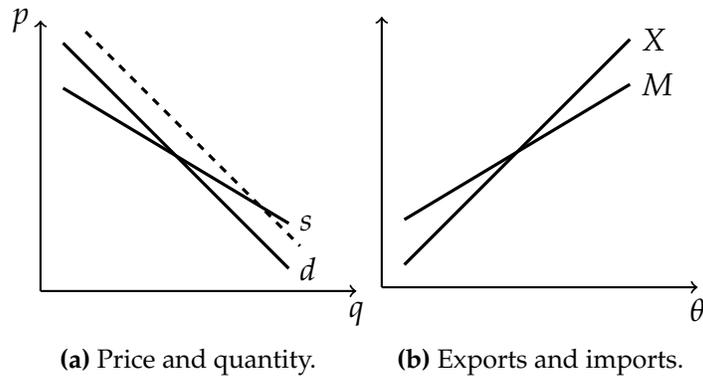**(a)** Price and quantity. **(b)** Exports and imports.

Figure 3: Strong home-market effect.

This situation is depicted in Figure 3.

3.3 Robustness

We have conducted our theoretical analysis in a stylized model with perfect competition. Our empirical analysis will focus on the global pharmaceutical industry, a complex sector in which patents provide an important source of market power. A natural question therefore is the extent to which our simple test, and its interpretation, may carry over to this industry. To shed light on this issue, we provide three examples that illustrate how more complex economic environments, without perfect competition, may reduce to the exact same equilibrium conditions as in Section 3.1. This establishes that the simple test presented in Section 3.2 remains valid in a range of imperfectly competitive environments. For expositional purposes, we only sketch alternative market structures and summarize their main implications. Details can be found in Appendix C.

Monopolistic Competition Consider first an economy where what we have referred to as "country i's variety" in Section 3.1 is itself a composite of multiple differentiated varieties, each produced by monopolistically competitive firms, as in Krugman (1980).

Formally, suppose that country j's consumption of drugs targeting disease n produced by a firm ω from country i is given by

$$d_{ij}^n(\omega) = (p_{ij}^n(\omega)/p_{ij}^n)^{-\sigma} d_{ij}^n, \tag{13}$$

where $p_{ij}^n = (\int (p_{ij}^n(\omega))^{1-\sigma} d\omega)^{1/(1-\sigma)}$ is the Dixit-Stiglitz price index and $\sigma > 1$ is the elasticity of substitution between country i's varieties. All other assumptions on the structure of demand are the same as in Section 3.1. On the supply side, each firm must now pay an overhead fixed cost, $f_i^n > 0$, in order to produce. Once this fixed cost has been paid, firms have a constant marginal cost, $c_i^n > 0$. All firms maximize profits taking their residual demand curves as given and enter up to the point where profits net of the overhead fixed cost are equal to zero.

At the industry-level, the previous assumptions lead to a supply curve similar to (4). Let us define Home's aggregate supply of drug n as the following quantity index,

$$s_i^n = \left(\int (s_i^n(\omega))^{(\sigma-1)/\sigma} d\omega \right)^{\sigma/(\sigma-1)},$$

where $s_i^n(\omega) \equiv \sum_j \tau_{ij}^n d_{ij}^n(\omega)$ is the total quantity supplied by firm ω, regardless of whether it is ultimately sold domestically or exported. Since demand is iso-elastic, monopolistically competitive firms charge constant markups, $\mu \equiv \sigma/(\sigma-1)$, over marginal costs. Together with free entry, this leads to

$$s_i^n = (N_i^n)^{\sigma/(\sigma-1)} f_i^n / ((\mu-1)c_i^n),$$
$$p_i^n = (N_i^n)^{1/(1-\sigma)} \mu c_i^n.$$

where we let $p_i^n \equiv p_{ii}^n$ denote the price index associated with country i's varieties before trade costs have been incurred and we let N_i^n denote the measure of firms producing drugs targeting disease n in country i. The two previous expressions provide a parametric representation of the sector-level supply curve, with the number of firms N_i^n acting as a parameter. In this case, one can eliminate N_i^n to express the supply curve explicitly as

$$s_i^n = \eta_i^n (p_i^n)^{-\sigma},$$

with $\eta_i^n \equiv f_i^n (c_i^n)^{(\sigma-1)} \sigma^\sigma (\sigma-1)^{(1-\sigma)}$. This is the counterpart of the supply equation (4).

11

Finally, since firms charge the same markup μ in all markets, equation (5) must hold for the price indices, p_{ij}^n, of country i's varieties of drug n in any importing country j.

At this point, we have established that equations (1)-(5) continue to hold. By construction of our quantity index, equation (6) must hold as well, as shown in Appendix C. This implies that our simple test remains valid under monopolistic competition. The only distinction between the perfectly competitive model of Section 3.1 and the present one is that monopolistic competition requires sector-level supply curves to be downward-sloping, with an elasticity equal to the opposite of the elasticity of substitution between domestic varieties,

$$\epsilon^s = -\sigma.$$

It is worth pointing out that the magnitude of the overhead fixed cost, f_i^n, is irrelevant for the shape of s and, in turn, irrelevant for the existence of a home-market effect. Though pharmaceutical firms are well-known for having large expenditures on research and development relative to the cost of manufacturing a drug, it does not follow that home-market effects should be particularly strong in that industry. The economic variable of interest for home-market effects is the magnitude of industry-level returns to scale—determined by σ under monopolistic competition—not firm-level returns to scale.

Note also that in the special case considered by Krugman (1980)—with upper-level Cobb-Douglas utility, $\epsilon^D = 1$, and lower-level CES utility, $\epsilon^x = \epsilon^d = \sigma$—the home-market effect is always strong for a small open economy. Indeed, under these parametric restrictions, inequality (12) reduces to

$$-\sigma - \lambda(\sigma - 1) < -\sigma < -\sigma + \lambda^2(\sigma - 1),$$

which must hold for any $\lambda > 0$ and $\sigma > 1$.

Bertrand Oligopoly Consider the same basic environment as in the previous example, but with a finite number of firms, N_i^n, that compete à la Bertrand in each sector. To simplify the analysis, we further assume that $\epsilon^D = \epsilon^x = \epsilon^d$ and that there is an arbitrarily large number of diseases so that firms charge the same markup in all markets.[8] All other assumptions are unchanged.

In equilibrium, firms still maximize their profits taking their residual demand curves as given, albeit internalizing the effect of their decisions on the domestic price index asso-

---

[8]In practice, pricing-to-market may be difficult to sustain in the pharmaceutical industry because of parallel trade, as in the case of the European Union, or because of the use of "international reference pricing" more generally; see Morton and Kyle (2012) for further discussion.

ciated with each disease. This leads to markups that now vary with the number of firms N_i^n. Formally, one can show that country i's aggregate supply of drug n and its price index now satisfy

$$s_i^n = (N_i^n)^{\sigma/(\sigma-1)} f_i^n / ((\mu(N_i^n) - 1)c_i^n),$$
$$p_i^n = (N_i^n)^{1/(1-\sigma)} \mu(N_i^n) c_i^n,$$

with $\mu(N_i^n) \equiv \frac{((1-1/N_i^n)\sigma + \epsilon^d/N_i^n)}{(1-1/N_i^n)\sigma + \epsilon^d/N_i^n - 1}$ the firms' markup under Bertrand competition. Though one can no longer solve explicitly for s_i^n as a function of p_i^n, the two previous expressions still provide a parametric representation of the sector-level supply curve. Since equations (1), (2), and (5) remain unchanged, the existence of such a curve is all we need to apply our simple test.

Locally, the price elasticity of supply is now given by

$$\epsilon^s = -\sigma \times \frac{(\mu - 1)^2 + (1 - 1/\sigma)(d \ln \mu / d \ln N)}{(\mu - 1)^2 (1 - (\sigma - 1)(d \ln \mu / d \ln N))}.$$

Compared to monopolistic competition with constant markups, where $d \ln \mu / d \ln N = 0$, the supply elasticity is lower in absolute value, $|\epsilon^s| < \sigma$, whenever markups are decreasing with the number of firms, $d \ln \mu / d \ln N < 0$. This is what happens for $\sigma > \epsilon^d$. In this case, the larger aggregate output in an industry is, the more firms there are, the lower the markups that they charge, and hence the lower the price that firms are willing to accept to produce a given aggregate quantity. At the sector-level, pro-competitive effects act as an additional source of increasing returns.

Monopoly To conclude, let us consider an economy where countries only produce a single variety of each drug, but unlike in our basic environment, this variety is produced by a monopolist that can invest in R&D, as in Krugman (1984). We follow the same strategy as in the previous example and assume that $\epsilon^D = \epsilon^x = \epsilon^d$ and that there is an arbitrarily large number of drugs so that firms charge the same markup in all markets.

For each disease n, the domestic monopolist in country i takes the residual demand curve in each market as given when simultaneously choosing its prices, p_{ij}^n, and its unit cost of production, c_i^n, in order to maximize its profits,

$$\pi_i^n = \sum_j (p_{ij}^n - \tau_{ij}^n c_{ij}^n) d(p_{ij}^n / P_j^n) D(P_j^n / (\theta_j^n P_j)) D_j - \eta_i^n f(c_i^n),$$

where $\eta_i^n f(c_i^n)$ denotes the amount of R&D required to have unit cost, c_i^n, which we as-

sume to be strictly decreasing and convex. The first-order conditions associated with this maximization problem imply the following version of the supply equation (4),

$$s_i^n = -\eta_i^n f'((\epsilon^d - 1)p_i^n/\epsilon^d).$$

Under the assumption that $f(\cdot)$ is convex, drug-level supply curves are necessarily downward-slopping with local elasticity now given by

$$\epsilon^s = d\ln(-f')/d\ln c.$$

The critical feature of the present model is that the marginal benefit of R&D is increasing with total output, which creates a negative relationship between output and prices.

In the four market structures that we have considered—perfect competition, monopolistic competition, Bertrand competition, and a single monopoly—the nature of economies of scale at the sector-level is very different. Here, it depends on the the elasticity of the marginal returns to R&D; previously, it derived from Marshallian externalities, love of variety, or pro-competitive effects. Nevertheless, equations (1)-(6) always hold. So, the simple economics of the home-market effect described in Section 3.2 remains valid. This motivates an empirical strategy that remains agnostic about such considerations, to which we now turn.

4 Data

Our analysis of the home-market effect rests on the correlation between a country's pattern of sales across drugs in the pharmaceutical sector and its pattern of exogenously-given demand across those drugs. We therefore draw on a linkage between two datasets: one that documents sales by country at the drug level, which we convert to a dataset of bilateral sales as detailed below, and one that describes the demographically-driven burden of each disease in each country. In both cases we use data from 2012—one cross-section of data suffices for testing the home-market effect since its prediction is cross-sectional in nature.

4.1 Pharmaceutical Sales

In order to construct bilateral data on pharmaceutical sales, $\{x_{ij}^n\}$, we draw on the MIDAS dataset produced by the firm IMS Health. IMS is a market research firm that sells MIDAS and other data products to firms in the pharmaceutical and health care industries.

Table 1: Top 10 countries in terms of sales

Country	Share of world sales (%) (1)	Share of world expenditures (%) (2)	Number of firms headquartered (3)
USA	36.67	42.10	361
Switzerland	13.14	0.61	35
Japan	11.62	12.68	53
United Kingdom	10.70	2.67	79
Germany	6.75	4.67	89
France	6.52	4.34	59
India	2.28	1.61	292
China (Mainland)	2.18	3.74	524
Canada	1.40	2.57	48
Italy	1.35	3.36	63

By auditing retail pharmacies, hospitals, and other sales channels, the raw MIDAS data record quarterly revenues and quantities by country at the "package" level, e.g. sales of a bottle of thirty 10mg tablets of the cholesterol-lowering drug Lipitor (atorvastatin). The data record unit sales and revenues (in local currency units), for both private and public purchasers.

Our version of the MIDAS data covers sales in 56 destination countries.[9] Given the comprehensive nature of the data, the vast majority of high revenue drugs globally—over 20,000 unique molecules, both brand-name and generic—are included. Our sample includes sales by roughly 2,800 firms. We observe the name of the firm selling each drug in our data and have used this name to hand-match each firm to the country in which it is headquartered. We refer to this country as the origin country. Given this mapping of firms to origin countries we then use the MIDAS data on sales (for each drug) by firm in each destination country to measure bilateral sales, from origin country to destination country, for each drug. We reiterate that the resulting bilateral sales data do not differentiate between exports and FDI-driven sales; they comprise the sum of all channels through which a firm in origin country i sells its product to consumers in destination country j.

The ten largest firms in our dataset in terms of sales (with origin country in parentheses) are, in descending order, Novartis (Switzerland), Pfizer (US), Merck & Co. (US), Sanofi-Aventis (France), Roche (Switzerland), AstraZeneca (UK), GlaxoSmithKline (UK), Johnson & Johnson (US), Eli Lilly & Co. (US), and Abbvie (US, a spin-off of Abbott Labo-

---

[9]The most recent versions of the MIDAS data cover more than 70 countries.

ratories).[10] While these top ten firms are headquartered in just five countries, firms in our dataset are headquartered in a total of 55 different origin countries. Table 1 reports the distribution of global sales for the ten largest countries in terms of share of world sales, along with the number of firms that are headquartered in each of those countries. There is a clear skewness in both of these variables, so we conduct our tests of the home-market effect in a wide range of subsamples designed to explore potential heterogeneity across large and small countries, as well as countries (such as India and China) where the large number of headquartered firms reflects a relatively large share of generic drug producers.

IMS uses a standard industry classification known as ATC codes, from the Anatomical Therapeutic Chemical Classification System to classify molecules into approximately 600 different therapeutic classes based on the main disease the drug is designed to treat. To link back to the example in our introduction, the ATC code "A02B" corresponds to "drugs for peptic ulcer and gastro-oesophageal reflux disease."

The resulting dataset can be reshaped to describe, within each therapeutic class, the bilateral sales between any origin country and any of 56 destination countries in 2012.

4.2 Disease Burden

We isolate a plausibly exogenous source of demand-side variation for each drug, in each country, by isolating the apparent extent to which drugs have a demographic bias in their relevance, as well as the extent to which countries differ in the demographic composition of their populations. This is the spatial analog of the identification strategy in Acemoglu and Linn (2004), who use changes in the demographic composition of the United States over time to estimate the relationship between market size and innovation in the pharmaceutical industry.

To construct this demand shifter, we draw on two datasets. The first, the World Health Organization (WHO)'s Global Burden of Disease (GBD) dataset, measures the burden of each disease, based on 60 WHO disease codes, in each country and year (where, again, we focus on 2012). Although there may be local variation in the collection of vital statistics that underpin these measures, the WHO ensures that these data are valid for cross-country and cross-disease comparisons. Importantly, these country-year-disease measures of burden are further broken down into six different demographic groups: three age groups (0-14, 15-59 and 60+) for each gender. The provided disease burden measure on which we draw is the number of disability adjusted life-years (DALYs) due to the dis-

---

[10]All comparisons across local currency units in this section use average 2012 exchange rates from the International Monetary Fund. Due to the presence of destination fixed effects, the home-market effect tests in Section 5 and the parameter estimates in Section 6 do not require a conversion across local currency units.

Table 2: Top 10 diseases in terms of sales

Disease class (WHO system)	Share of world sales (%) (1)	Number of origin countries (2)	Average Herfindahl index across destinations (3)
Other infectious diseases	8.62	55	0.08
Hypertensive heart disease	6.56	55	0.10
Cardiovascular diseases	6.30	55	0.13
Ischaemic heart disease	5.99	54	0.14
Other neoplasms	5.80	52	0.12
Diabetes mellitus	4.75	54	0.15
Rheumatoid arthritis	4.55	48	0.23
Genito-urinary diseases	3.97	51	0.14
Obstructive pulmonary disease	3.50	49	0.27
Schizophrenia	3.26	51	0.17

ease, a metric that aims to capture both mortality and morbidity. We have hand-coded a linkage from each of the 600 therapeutic classes in IMS MIDAS to its corresponding WHO disease code. For example, the ATC code "A02B" for "drugs for peptic ulcer and gastro-oesophageal reflux disease" is linked to the WHO code for "peptic ulcer disease."[11] Each of the 60 WHO codes is the empirical counterpart of a disease n in the model of Section 3.

Table 2 describes the top 10 diseases (broken down by WHO codes) in terms of global sales of their corresponding drugs in the MIDAS dataset. For each disease, there are many origin countries participating in the sale of drugs treating that disease. As illustrated in the last column, the typical destination country in our data is served by an extremely unconcentrated set of firms, even within each disease class.

The second input into the construction of our demand shifter is the population of each country in each of the six demographic groups in 2012. We obtain this data from the US Census Bureau's International Database.

Using the data described above, we exploit the twin facts that disease burdens vary by demographic groups, and that countries vary in their demographic composition, to construct a "predicted disease burden", for disease n in country i in year 2012 as:

$$(PDB)_i^n = \sum_{a,g} \left[\text{population}_{iag} \times \left(\frac{\sum_{j \neq i} \text{disease burden}_{jag}^n}{\sum_{j \neq i} \text{population}_{jag}} \right) \right]. \tag{14}$$

---

[11] Around 89% of our ATC4 codes were linked to WHO GBD codes. The main reason for non-matches is that certain ATC4 codes are too broad to be matched to a single GBD disease code.

The ratio $\frac{\sum_{j \neq i} \text{disease burden}_{jag}^n}{\sum_{j \neq i} \text{population}_{jag}}$ measures the average country-level disease burden per population from disease n for gender g and age group a in 2012, calculated excluding the country of interest (that is, summing over all countries j except for country i). This ratio is then weighted by the population for that gender g and age group a, and summed across age and gender groups, for a given country i in 2012.

5   Testing for the Home-Market Effect

5.1   Baseline Results

To test whether bilateral sales in the pharmaceutical industry satisfy the weak and strong home-market effects, we use $(PDB)_i^n$ as an empirical proxy for the demand-shifter θ_i^n in equation (1). That is, we assume that, up to a first-order approximation,

$$\ln \theta_i^n = \gamma \ln (PDB)_i^n + \gamma_i^n, \tag{15}$$

where γ is strictly positive and γ_i^n is an error term that captures other determinants of the demand-shifter θ_i^n for drugs targeting disease n in country i that is uncorrelated with $(PDB)_i^n$. Our results in Table 3 below demonstrate that this proxy is a strong predictor of expenditure. And Table E.1 in Appendix E establishes that the variable $(PDB)_i^n$ is also a strong predictor of the actual burden that any country i is likely to suffer from in disease n. That is, the simple demographic predictor of disease burden in equation (14) is a useful empirical proxy for θ_i^n, despite the myriad other reasons for countries to differ in their demand for drugs targeting any particular disease.

Combining equations (7) and (15), we obtain our baseline estimating equation,

$$\ln x_{ij}^n = \delta_{ij} + \delta^n + \tilde{\beta}_M \ln (PDB)_j^n + \tilde{\beta}_X \ln (PDB)_i^n + \tilde{\varepsilon}_{ij}^n, \tag{16}$$

with $\tilde{\beta}_M \equiv \gamma \beta_M$, $\tilde{\beta}_X \equiv \gamma \beta_X$, and $\tilde{\varepsilon}_{ij}^n \equiv \varepsilon_{ij}^n + \gamma_i^n$. Under the maintained assumption that $\gamma > 0$, a positive test of the weak home-market effect therefore corresponds to $\tilde{\beta}_X > 0$, whereas a positive test of the strong home-market effect corresponds to $\tilde{\beta}_X > \tilde{\beta}_M$.

Several details of the estimation procedure used in this section are important to note. First, we estimate equation (16) on a sample of ij observations for which $i \neq j$, in line with the derivation of equation (7). This ensures that the trivial correlation between home's demand shifter and sales from home to itself does not enter the analysis (though in Table 4 we report the extent to which incorporating this variation changes our findings). Sec-

Table 3: Test of the Home-Market Effect (baseline)

	log(bilateral sales)		
	(1)	(2)	(3)
log(PDB, destination)	0.527		0.561
	(0.098)		(0.109)
log(PDB, origin)		0.927	0.914
		(0.177)	(0.125)
p-value for $H_0 : \tilde{\beta}_X \leq 0$		0.000***	0.000***
p-value for $H_0 : \tilde{\beta}_X \leq \tilde{\beta}_M$			0.028**
Origin $\times$ destination FE	✓	✓	✓
Destination $\times$ disease FE		✓	
Origin $\times$ disease FE	✓		
Disease FE	✓	✓	✓
Adjusted R^2	0.629	0.562	0.539
Observations	18,823	18,966	19,213

Notes: OLS estimates of equation (16). Predicted disease burden (PDB_i^n) is constructed from an interaction between the global (leaving out country i) disease burden by demographic group in disease n, and the size of each demographic group in country i. All regressions omit the bilateral sales observation for home sales (i.e where $i = j$). Standard errors in parentheses are two-way clustered at origin and destination country levels. p-values are based on F-test of H_0. *** p<0.01, ** p<0.05, * p<0.1.

ond, in our baseline estimates we drop observations for which $x_{ij}^n = 0$, but we return this aspect of the variation in Table 8. And finally, because the predicted disease burden regressors vary at the origin and destination levels (but not at the bilateral level) we take a conservative approach to inference and use standard errors that are two-way clustered at both the origin and destination levels throughout.

Table 3 presents OLS estimates of equation (16). We begin in column (1) with a specification designed to estimate $\tilde{\beta}_M$ as accurately as possible. To do so we control for an origin-disease fixed effect (rather than including the origin country's predicted disease burden). While the estimate of $\tilde{\beta}_M > 0$ seen there should not be surprising—that a demand shifter in the destination country is positively correlated with greater purchases by that destination—this can be thought of as a check on the validity and power of demographic variation for predicting drug expenditure. Column (2) proceeds with an analogous specification designed to estimate $\tilde{\beta}_X$ alone, as accurately as possible, while controlling for a destination-disease fixed-effect. The estimated value of $\tilde{\beta}_X$ is clearly positive and statistically significant. This result provides a resounding rejection of the absence

Table 4: Test of the Home-Market Effect (sensitivity analysis I)

	log(bilateral sales)		
	(1)	(2)	(3)
log (PDB, destination)	0.561	0.529	0.536
	(0.109)	(0.104)	(0.105)
log (PDB, origin)	0.914	0.826	0.707
	(0.125)	(0.101)	(0.168)
p-value for $H_0 : \tilde{\beta}_X \leq 0$	0.000***	0.000***	0.000***
p-value for $H_0 : \tilde{\beta}_X \leq \tilde{\beta}_M$	0.028**	0.040**	0.171
Home sales (X_{ii}^n) observations included		✓	
Disease FE × origin GDP per capita			✓
Disease FE × destination GDP per capita			✓
Adjusted R^2	0.539	0.562	0.555
Observations	19,213	21,339	18,954

Notes: OLS estimates of equation (16). All specifications control for origin-destination fixed-effects and disease fixed-effects. See Table 3 for details on construction of variables, sample restrictions, and calculations of standard errors (reported in parentheses) and p-values. *** p<0.01,** p<0.05, * p<0.1.

of a weak home-market effect (or equivalently, means that the probability that the weak home-market effect operates in this context is high).

Finally, column (3) estimates $\tilde{\beta}_M$ and $\tilde{\beta}_X$ simultaneously in the true spirit of equation (7). This is our preferred specification. We first note that the estimates of $\tilde{\beta}_M$ and $\tilde{\beta}_X$ in column (3) are very similar to those in columns (1) and (2), so evidence for the weak home-market effect remains firm. And the p-value on the F-test for $\tilde{\beta}_X \leq \tilde{\beta}_M$ is 0.028, implying that the absence of a strong home-market effect can be rejected at the five, but not at the one, percent level. That is, the balance of probability speaks in favor of the strong home-market effect taking place in the pharmaceutical sector.

5.2 Sensitivity Analysis

Having established these simple facts about the weak and strong home-market effects in the global pharmaceutical sector, we now seek, in the remainder of this section, to explore the robustness of these estimates to various changes in the empirical specification. Throughout, we focus on estimates of equation (7), and hence on tests for both the weak and the strong home-market effect.

We begin in Table 4 by repeating in column (1), for the sake of comparison, our base-

line estimate from Table 3. This baseline specification uses all bilateral sales observations x_{ij}^n that are positive and for which $i \neq j$. Column (2) then assesses the extent to which including home sales observations (those for which $i = j$) affect our estimates of the home-market effect—evidently, this specification decision has little bearing on our estimates. Column (3) tests for the two home-market effects in a demanding specification that also simultaneously controls for interactions between the origin country's per-capita GDP and disease fixed-effects, as well as for the analogous variable on the destination country side. In this specification, the standard errors rise, as expected given the number of regressors that are now being estimated. In turn, while the null of no weak home-market effect can still be rejected at standard confidence level, this is no longer true for the null of strong home-market effect. Reassuringly, the point estimates of $\tilde{\beta}_M$ and $\tilde{\beta}_X$ have not changed much as compared to the estimates in column (1). This suggests that there is not some systematic tendency for rich countries to produce certain drugs that happen to treat the diseases associated with rich country demographics.

In Tables 5, 6, and 7, we continue to explore the robustness of our results. Starting with Table 5, column (1) again repeats our baseline estimate for the purpose of comparison. Column (2) demonstrates that our estimates are not being driven by the largest global drug-selling countries: dropping the five largest such countries does not much change the estimated coefficients. Columns (3) through (5) ask whether the results of Table 3 are being driven by poor origin or destination countries and the answer appears to be resoundingly in the negative: our coefficient estimates are remarkably stable to changing the countries in the sample on the basis of their GDP per capita levels. Admittedly, that this is true should not come as much of a surprise given the results in column (3) of Table 4, documenting insensitivity to a rich set of interactions between the income level of the origin and destination countries. Like in Table 4, we can always reject the null of no weak home-market effect, whereas whether we can reject the null of no strong home-market effect varies across specifications.

Provided that our first-order approximation around a symmetric equilibrium is valid, our test of the market-effect does not require any restriction on the spatial correlation of demand shocks across countries, as mentioned in Section 3.2. Away from this symmetric benchmark, however, demand shocks in third countries may not be absorbed by our disease fixed-effect. Table 6 explores the potential importance of such considerations. Again, column (1) repeats our baseline estimate for the purpose of comparison. Columns (2) and (3) show that restricting sales to a "donut" of destination countries, either located at more than 1,000km or 2,000km from the home market, has little effect on our estimates. The same is true in column (4) when we control for the average disease burdens in third

Table 5: Test of the Home-Market Effect (sensitivity analysis II)

	log(bilateral sales)				
	(1)	(2)	(3)	(4)	(5)
log (PDB, destination)	0.561	0.691	0.541	0.679	0.660
	(0.109)	(0.197)	(0.103)	(0.112)	(0.112)
log (PDB, origin)	0.914	0.670	0.957	0.918	0.841
	(0.125)	(0.162)	(0.232)	(0.162)	(0.299)
p-value for $H_0 : \tilde{\beta}_X \leq 0$	0.000***	0.000***	0.000***	0.000***	0.004***
p-value for $H_0 : \tilde{\beta}_X \leq \tilde{\beta}_M$	0.028**	0.531	0.056*	0.076*	0.268
Drop largest 5 origins		✓			
Drop poorest 1/3 origins			✓		✓
Drop poorest 1/3 dest.				✓	✓
Adjusted R^2	0.539	0.460	0.529	0.545	0.540
Observations	19,213	9,200	16,864	12,580	11,243

Notes: OLS estimates of equation (16). All specifications control for origin-destination fixed-effects and disease fixed-effects. See Table 3 for details on construction of variables, sample restrictions, and calculations of standard errors (reported in parentheses) and p-values. *** p<0.01,** p<0.05, * p<0.1.

countries, weighted by their distance to the origin and destination country.

Table 7 explores our baseline specification in a sample which only includes generic drugs: the coefficient on predicted disease burden in the origin country is economically smaller and statistically distinguishable in the generics-only sample (column 2) relative to our baseline estimate in column (1). Perhaps as expected, this suggests that sector-level economies of scale, and in turn weak and strong home-market effects, may be less important for generic drugs than for branded drugs.

Finally, in Table 8 we estimate the home-market effect along the extensive margin of whether a foreign market is penetrated at all. Given that our previous results (in Tables 3-5 above) used the log of bilateral sales (x_{ij}^n) as the dependent variable, any country pair-disease observations with zero bilateral sales were omitted from the estimation sample. Therefore, for completeness, we present here analogous specifications where the dependent variable is simply an indicator variable for whether $x_{ij}^n > 0$ or not. For simplicity, we estimate this as a linear probability model. There is robust support in Table 8 for the idea that home demand shocks also lead to more exports abroad along the extensive margin.

Table 6: Test of the Home-Market Effect (sensitivity analysis III)

	log(bilateral sales)			
	(1)	(2)	(3)	(4)
log (PDB, destination)	0.561	0.565	0.620	0.557
	(0.109)	(0.114)	(0.089)	(0.108)
log (PDB, origin)	0.914	0.926	0.812	0.913
	(0.125)	(0.148)	(0.170)	(0.129)
p-value for $H_0 : \tilde{\beta}_X \leq 0$	0.000***	0.000***	0.000***	0.000***
p-value for $H_0 : \tilde{\beta}_X \leq \tilde{\beta}_M$	0.028**	0.049**	0.181	0.029**
Distance cutoff	–	1,000 km	2,000 km	–
$\sum_{k \neq i} \ln PDB_k^n \cdot dist_{ik}^{-1}$				✓
$\sum_{k \neq j} \ln PDB_k^n \cdot dist_{kj}^{-1}$				✓
Adjusted R^2	0.539	0.562	0.549	0.539
Observations	19,213	16,652	13,440	19,213

Notes: OLS estimates of equation (16). All specifications control for origin-destination fixed-effects and disease fixed-effects. "Distance cutoff" refers to specifications in which ij country pairs for which $dist_{ij}$ less than cut-off are dropped. See Table 3 for details on construction of variables, sample restrictions, and calculations of standard errors (reported in parentheses) and p-values. *** p<0.01,** p<0.05, * p<0.1.

Table 7: Test of the Home-Market Effect (sensitivity analysis IV)

	log(bilateral sales)	
	(1)	(2)
log (PDB, destination)	0.561	0.450
	(0.109)	(0.207)
log (PDB, origin)	0.914	0.245
	(0.125)	(0.140)
p-value for $H_0 : \tilde{\beta}_X \leq 0$	0.000***	0.003***
p-value for $H_0 : \tilde{\beta}_X \leq \tilde{\beta}_M$	0.028**	0.702
Generic drugs only		✓
Adjusted R^2	0.539	0.524
Observations	19,213	10,152

Notes: OLS estimates of equation (16). All specifications control for origin-destination fixed-effects and disease fixed-effcts. See Table 3 for details on construction of variables, sample restrictions, and calculations of standard errors (reported in parentheses) and p-values. *** p<0.01,** p<0.05, * p<0.1.

Table 8: Test of the Home-Market Effect (extensive margin)

	𝟙(bilateral sales>0)			
	(1)	(2)	(3)	(4)
log (PDB, destination)	0.008	0.009	0.009	0.012
	(0.004)	(0.004)	(0.004)	(0.006)
log (PDB, origin)	0.055	0.063	0.033	0.087
	(0.013)	(0.013)	(0.010)	(0.017)
p-value for $H_0 : \tilde{\beta}_X \leq 0$	0.000***	0.000***	0.001***	0.000***
p-value for $H_0 : \tilde{\beta}_X \leq \tilde{\beta}_M$	0.000***	0.000***	0.004***	0.000***
Origin. × destination FE	✓	✓	✓	✓
Disease FE	✓	✓	✓	✓
Disease FE × origin GDP/capita		✓		
Disease FE × destination GDP/capita		✓		
Drop largest 5 origins			✓	
Drop poorest 1/3 origins				✓
Drop poorest 1/3 destinations				✓
Adjusted R^2	0.488	0.502	0.341	0.511
Observations	178,640	172,260	162,690	77,256

Notes: Linear probability model estimates based on equation (16). All specifications control for origin-destination fixed-effects and disease fixed-effcets. See Table 3 for details on construction of variables, sample restrictions, and calculations of standard errors (reported in parentheses) and p-values. *** p<0.01,** p<0.05, * p<0.1.

6 Disentangling Demand and Supply Elasticities

The results of Section 5 provide indelible support for the notion of a home-market effect in the global pharmaceutical sector. But, as discussed in Section 3, weak and strong home-market effects depend both on demand and supply elasticities. Thus, the structural interpretation of the previous effect remains open. Does the home-market effect reflect the presence of economies of scale at the sector-level or merely a low elasticity of demand? And if it does indeed reflect economies of scale, how large are they? We now address these questions.

6.1 Estimating the Elasticity of Demand

Around a symmetric equilibrium, bilateral sales can be expressed, up to a first-order approximation, as

$$\ln x_{ij}^n = \delta_j^n + (1 - \epsilon^x) \ln p_i^n + (1 - \epsilon^x) \ln \tau_{ij}^n, \tag{17}$$

where δ_j^n is a destination-disease fixed effect.[12] We aim here to estimate the price elasticity of exports, ϵ^x. Our approach proceeds in two steps. First, we estimate the elasticity of trade costs with respect to distance. And second, we estimate the elasticity of exports with respect to distance. Then, by equation (17), $1 - \epsilon^x$ is given by the ratio of the second elasticity over the first.

Distance elasticity of trade costs

Suppose that, up to a first-order approximation, trade costs τ_{ij}^n can be expressed as

$$\ln \tau_{ij}^n = \alpha \ln dist_{ij} + v_{ij}^n, \tag{18}$$

where $dist_{ij}$ is the physical distance between country i and country j and v_{ij}^n is the component of trade costs not explained by distance. We use micro-data on prices of individual drugs in order to identify α. For any individual variety of a drug ω within the class of drugs that treat disease n, suppose that prices satisfy

$$p_{ij}^n(\omega) = \tau_{ij}^n p_i^n(\omega). \tag{19}$$

---

[12]Details can be found in Appendix D. The expression in equation (17) is valid, up to a first-order approximation, for any demand system that can be written in the form of equations (1)-(3). But this expression becomes valid (globally and without approximation) in the commonly applied case where the function $d(\cdot)$ in equation(2) is CES.

This condition mandates that the price at which the same variety ω sells, after adjusting for trade costs, is the same across destination markets. Substituting (18) into (19), we have

$$\ln p_{ij}^n(\omega) = \alpha \ln dist_{ij} + \ln p_i^n(\omega) + v_{ij}^n.$$

Thus we can obtain an unbiased estimate of α using OLS to estimate the following specification

$$\ln p_{ij}^n(\omega) = \alpha \ln dist_{ij} + \delta_i^n(\omega) + \delta_{ij}^n(\omega), \tag{20}$$

where $\delta_i^n(\omega)$ is a product fixed-effect and $\delta_{ij}^n(\omega)$ is an error term.[13] The basic idea here is that if a given product sells in many destination countries, then the extent to which the prices of that disease vary across destinations j that are different distances $dist_{ij}$ from the producer's origin country i identifies α.

The results from estimating equation (20) are reported in Table 9 (column 1).[14] Distance is evidently a shifter of costs at distant destination locations and is hence positively correlated ($p = 0.07$) with the retail price (for the same product, sold from the same origin), despite the manifold reasons for retail prices to vary across consumer markets in the pharmaceutical sector. As might be expected, given the relatively high weight-to-value of pharmaceutical products, the effect of distance on prices (captured by the parameter α) is low relative to analogous estimates (for all traded merchandise sectors) in the literature. For example, Head and Mayer (2013) report a preferred distance elasticity of -0.89 and a preferred trade elasticity of -5.03. Together, these estimates imply $\alpha = 0.18$, which is three times larger than our estimate for the pharmaceutical sector.

Distance elasticity of sales

Substituting for trade costs in (17) using (18), we obtain

$$\ln x_{ij}^n = \delta_j^n + \delta_i^n + \rho \ln dist_{ij} + \chi_{ij}^n \tag{21}$$

with $\rho \equiv (1 - \epsilon^x)\alpha$, $\chi_{ij}^n \equiv (1 - \epsilon^x)v_{ij}^n$, and δ_i^n representing an origin-drug fixed-effect. As before, we estimate this equation with OLS. Our estimate of ρ is reported in Table 9, column (2). As is commonly found in estimates of the gravity equation (21), bilateral

---

[13]By "product" we refer to the finest level of product identification available in the IMS data—a level that is analogous to the barcode/UPC classification system in the consumer goods sector. In practice, product fixed-effects refer to the permutation of physiologically active molecules (since some drugs contain more then one active molecule), interacted with the dosage size and method (e.g. capsule vs. pill), interacted with the disease for which the drug is intended to treat (since, in rare cases, the same molecule can be marketed in separate therapeutical classes), and interacted with the firm selling the drug.

[14]Data on bilateral country pair distance is from the CEPII Gravity dataset; see Head and Mayer (2010).

Table 9: Demand elasticity estimates

	log(price)	log(bilateral sales)
	(1)	(2)
log (bilateral distance)	0.064	-0.301
	(0.035)	(0.073)
Product FE	✓	
Origin × disease FE		✓
Destination × disease FE		✓
Adjusted R^2	0.881	0.573
Observations	64,618	18,703

Notes: Column (1) reports OLS estimates of equation (20); product fixed-effects control for interactions between all combinations of active molecules, drug dosage sizes/methods, and disease classes; standard errors (in parentheses) two-way clustered by firm and destination. Column (2) reports OLS estimates of equation (21). Standard errors in parentheses are two-way clustered at origin and destination country levels. All regressions omit the bilateral sales observation for home sales (i.e where $i = j$).

distance has a negative and statistically significant impact on impeding bilateral drug sales in this setting. But again, as with our discussion of α above, the estimated effect of distance on trade is about three times lower than typical estimates from trade data in other sectors.

Elasticity of demand

The identity $\rho \equiv (1 - \epsilon^x)\alpha$ implies that $\epsilon^x = 1 - \rho/\alpha$. Given the estimates in Table 9, this implies that $\epsilon^x = 5.70$, a point estimate that implies that demand is elastic in the present setting. As per the discussion in Section 3.2, this then implies that, at least for a small open economy, the tests for the weak and strong home-market effects reported in Section 5.1 provide bounds on economies of scale. For example: we know that the evidence for the weak home-market effect reported in Table 1 implies that industry-level (positive) economies of scale are at work in this setting. Naturally, such a bound is of only limited use for quantitative policy questions so we turn now to a method that uses the demand elasticity estimate here in order to obtain a point estimate of the elasticity of supply.

6.2   Estimating the Elasticity of Supply

Let $r_i^n \equiv p_i^n s_i^n$ denote the total sales of drugs targeting disease n by firms from country i. Around a symmetric equilibrium, up to a first-order approximation, the supply relation

in equation (5) can be written as

$$\ln r_i^n = (1 + \epsilon^s) \ln p_i^n + \ln \eta_i^n.$$

Using the previous expression to substitute for p_i^n in equation (17), we obtain

$$\ln x_{ij}^n = \delta_j^n + \delta_{ij} + \left(\frac{1 - \epsilon^x}{1 + \epsilon^s} \right) \ln r_i^n + \phi_{ij}^n, \tag{22}$$

with δ_{ij} representing an origin-destination fixed-effect and $\phi_{ij}^n \equiv \chi_{ij}^n - \bar{\chi}_{ij}^n - \left(\frac{1 - \epsilon^x}{1 + \epsilon^s} \right) \ln \eta_i^n$. OLS estimates of equation (22) would be biased because both the supply shock η_i^n and unobserved trade costs χ_{ij}^n in the error term ϕ_{ij}^n contributes to total sales r_i^n. But for all destination observations $j \neq i$, an exogenous shifter of demand at the origin country i (such as the predicted disease burden variable PDB_i^n introduced in equation 14) can be used as a valid instrumental variable for r_i^n. Such an IV estimation procedure identifies $\left(\frac{1 - \epsilon^x}{1 + \epsilon^s} \right)$, which, combined with the estimate of ϵ^x from above, identifies the slope of the supply curve given by ϵ^s.

Table 10 reports estimates from specification (22). We begin in column (1) by reporting the first-stage regression of $\ln r_i^n$ on $\ln PDB_i^n$, conditional on origin-destination and destination-disease fixed-effects. That predicted disease burden is strongly correlated with total sales (the F-statistic on this excluded instrument is equal to 148.84, the square of the t-statistic reported in column 1) should come as no surprise given the results in Table 3. Column (2) then reports the OLS estimate of equation (22) and column (3) the corresponding IV estimate.[15] This (statistically significant) IV estimate implies that $\left(\frac{1 - \epsilon^x}{1 + \epsilon^s} \right) = 0.746$. Given our estimate of $\epsilon^x = 5.70$ from above, this implies that $\epsilon^s = -7.30$. As expected, given the bounds implied by the weak home-market effect discussed above, the estimated industry-level supply curve in this setting is downward-sloping, indicating returns to scale.

How does this estimate of ϵ^s compare to those in prior work? Both empirical and theoretical findings offer points of reference. From the empirical literature, one strand aims to estimate industry-level economies of scale directly, via industry-level production functions. A prominent estimate (pooled among all U.S. manufacturing sectors, so unfortunately not available for the pharmaceutical sector alone) from Basu and Fernald (1997) estimates industry-level economies of scale that generate an industry-level supply curve

---

[15]The fact that the OLS estimate in column (2) is smaller than the IV estimate in column (3) is consistent with downward-sloping supply curves since when $\epsilon^s < -1$ (and given elastic demand, $\epsilon^x > 1$) the error term ϕ_{ij}^n in equation (22) depends negatively on the supply shock η_i^n.

Table 10: Supply elasticity estimates

	log(total sales)	log(bilateral sales)	
	OLS	OLS	IV
	(1)	(2)	(3)
log (PDB)	1.236		
	(0.101)		
log (total sales)		0.684	0.746
		(0.055)	(0.127)
p-value for $H_0 : \left(\frac{1-\epsilon^x}{1+\epsilon^s}\right) = 1$			0.046**
Adjusted R^2	0.798	0.627	0.626
Observations	16,881	16,881	16,881

Notes: Column (2) reports the OLS estimate, and column (3) the IV estimate, of equation (22). Column (1) reports the corresponding first-stage specification. The instrumental variable is log (predicted disease burden) in the origin country (and the corresponding first-stage regression is reported in column 1). All regressions omit the bilateral sales observation for home sales (i.e where $i = j$) and control for orign-disease and destination-disease fixed-effects. Standard errors in parentheses are two-way clustered at origin and destination country levels. p-value is based on F-test of H_0. *** $p<0.01$,** $p<0.05$, * $p<0.1$.

with $\epsilon^s = -4.45$. A second strand, initiated by Antweiler and Trefler, 2002, uses patterns of comparative advantage revealed in international trade data to infer relative costs for each country-industry, and then estimates the extent to which those inferred costs depend on scale. For the pharmaceutical sector Antweiler and Trefler's (2002) estimates imply $\epsilon^s = -4.27$. Since lower supply elasticities in absolute value imply larger effects of quantity on producer prices, both of these estimates imply somewhat stronger economies of scale than found in our estimate of $\epsilon^s = -7.30$. That said, neither of these estimates is based on an empirical strategy that isolates demand-side variation and is powerful enough to circumvent weak instrument concerns.

The influential model of Krugman (1980) also provides a clear benchmark. As discussed in Section 3.3, this model is a special case in which there is a particularly stark connection between industry-level supply and demand elasticities: $\epsilon^s = -\epsilon^x$. This implies that $\left(\frac{1-\epsilon^x}{1+\epsilon^s}\right)$, the coefficient reported in column (3) of Table 8, should be one. Instead our IV estimate is equal to 0.746, about 25% smaller. While the reported p-value demonstrates that the particular parameter value assumed in Krugman (1980) is rejected, our estimate is certainly closer to this benchmark value than to the constant-returns extreme in which $\epsilon^s = \infty$ (and hence the coefficient in column 3 would be equal to zero).

7 Concluding Remarks

Since the home-market effect hypothesized by Linder (1961) and formalized by Krugman (1980) is about the causal effect of cross-country differences in demand on the pattern of international specialization, any empirical test of this phenomenon requires exogenous demand variation. In this paper, we have focused on the global pharmaceutical industry as a way to obtain such variation. Our empirical strategy builds on the basic observation that countries whose populations, because of exogenous demographic characteristics, are more likely to suffer from particular diseases are also more likely to have high demand for drugs targeting those diseases.

We have conducted tests of two different notions of the home-market effect. The first test, which is based on what we have referred to as the weak home-market effect, investigates whether countries tend to sell more abroad in sectors for which they have larger domestic markets. In the present context, this boils down to estimating whether the elasticity of a country's foreign sales with respect to its demographically predicted disease burden is positive. In line with the work of Linder (1961), the answer is a resounding yes. In short, the more we die (at home), the more we sell (abroad).

Our second test, defined by what we have referred to as the strong home-market effect, explores whether the previous effect can be important enough to turn countries with larger demand for some products into net sellers of those products, a stronger implication of Krugman's (1980) monopolistically competitive model. Our baseline results speak in favor of the strong home-market effect taking place in the pharmaceutical sector, though, in comparison with the weak home-market effect, we are not able to reject the null of no strong home-market effect in some of our specifications.

To delve further into the economic determinants of the home-market effect, we have concluded our analysis by estimating demand and supply elasticities in the pharmaceutical industry. Our estimates point towards the home-market effect being driven by substantial economies of scale at the sector-level rather than a low elasticity of demand. Quantitatively, we have estimated a supply elasticity that is about three-quarters the size of what a monopolistically competitive model, like Krugman (1980), would predict. Recent quantitative work on international trade and economic geography has typically assumed, without attempting to estimate, economies of scale that are either zero, as in Eaton and Kortum (2002), or of the Krugman's (1980) magnitude. In our context, both extremes are rejected by the data. Our analysis, however, demonstrates how a single supply-side parameter can nest these two cases, and how a plausibly exogenous demand shifter can let the data speak freely to this parameter's value.

References

Acemoglu, Daron and Joshua Linn, "Market size in innovation: Theory and evidence from the pharmaceutical industry," *Quarterly Journal of Economics*, 2004, *119* (3), 1049–1090.

Amiti, Mary, "Inter-industry trade in manufactures: does country size matter?," *Journal of Intermational Economics*, 1998, *44*, 231–255.

Antweiler, Werner and Daniel Trefler, "Increasing Returns and All That: A View from Trade," *American Economic Review*, 2002, *92* (1), 93–119.

Basu, Susanto and John G. Fernald, "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 1997, *105* (2), 249–283.

Behrens, Kristian, Andrea R. Lamorgese, Gianmarco Ottaviano, and Takatoshi Tabuchi, ""Beyond the home market effect: Market size and specialization in a multi-country world," *Journal of International Economics*, 2009, *79* (2), 259–265.

Brulhart, Marius and Federico Trionfetti, "A test of trade theories when expenditure is home biased," *European Economic Review*, 2009, *53* (7), 830–845.

Caron, Justin, Thibault Fally, and Ana Cecilia Fieler, "Home-market effects on innovation," *mimeo UC Berkeley*, 2015.

Cleave, TL, *Peptic Ulcer: A New Approach to its Causation, Prevention, and Arrest, Based on Human Evolution*, Bristol: Great Britain: John Wright and Sons Ltd, 1962.

Crozet, Mathieu and Federico Trionfetti, "Trade costs and the home market effect," *Journal of Intermational Economics*, 2008, *76*, 309–321.

Davis, Donald R., "The home market, trade and industrial structure," *American Economic Review*, 1998, *88* (5), 1264–1276.

_ **and David Weinstein**, "Does Economic Geography Matter for International Specialization?," *NBER Working Paper 5706*, 1996.

_ **and** _ , "Economic geography and regional production structure: an empirical investigation," *European Economic Review*, 1999, *43*, 379–407.

_ **and** _ , "Market access, economic geography and comparative advantage: an empirical test," *Journal of Intermational Economics*, 2003, *59*, 1–23.

Dingel, Jonathan I., "The Determinants of Quality Specialization," 2015. mimeo University of Chicago.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Ethier, Wilfred J., "Decreasing Costs in International Trade and Frank Graham's Argument for Protection," *Econometrica*, 1982, *50* (5), 1243–1268.

Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman, "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 2011, *119* (4), 721–765.

_ , _ , **and** _ , "A Linder Hypothesis for Foreign Direct Investment," *Review of Economic Studies*, 2015, *82*, 83–121.

Feenstra, Robert C., James R. Markusen, and Andrew K. Rose, "Using the Gravity Equation to Differentiate among Alternative Theories of Trade," *Canadian Journal of Economics*, 2001, *34* (2), 430–447.

Fujita, Masahisa, Paul R. Krugman, and Anthony Venables, *The Spatial Economy: Cities, Regions, and International Trade*, MIT press, 2001.

Gali, Jordi and Tommaso Monacelli, "Monetary Policy and Exchange Rate Volatility in a Small Open Economy," *Review of Economic Studies*, 2005, *72*, 707–734.

Hanson, Gordon H. and Chong Xiang, "The home market effect and bilateral trade patterns," *American Economic Review*, 2004, *94* (4), 1108–1129.

Hara, Takuji, *Innovation in the Pharmaceutical Industry: The Process of Drug Discovery and Development*, Cheltenham: UK: Edward Elgar, 2003.

Head, Keith and John Ries, "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of U.S.-Canada Trade," *American Economic Review*, 2001, *91* (4), 858–876.

_ **and Thierry Mayer**, *The Empirics of Agglomeration and Trade*, Vol. 4 of *Handbook of Regional and Urban Economic*, Amsterdam: Holland: Elsevier, 2004.

_ **and** _ , "The erosion of colonial trade linkages after independence," *Journal of International Economics*, 2010, *81* (1), 1–14.

__ **and** __ , *Gravity Equations: Toolkit, Cookbook, Workhorse*, Vol. 4 of *Handbook of International Economics*, Elsevier, 2013.

Helpman, Elhanan and Paul Krugman, *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Cambridge, Massachusetts: MIT Press, 1985.

__ **and Paul R. Krugman**, *Trade Policy and Market Structure*, Cambridge, Massachussetts: MIT Press, 1989.

Holmes, Thomas J. and John J. Stevens, "Does home market size matter for the pattern of trade?," *Journal of International Economics*, 2005, *65*, 489–505.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *The American Economic Review*, 1980, *70* (5), 950–959.

Krugman, Paul R., *Import protection as export promotion: international competition in the presence of oligopoly and economies of scale* Monopolistic competition in international trade, Oxford University Press, 1984.

Linder, Staffan B., *An Essay on Trade and Transformation*, Uppsala: Almqvist and Wiksells, 1961.

Lundback, Erik J. and Johan Torstensson, "Demand, comparative advantage and economic geography in international trade: Evidence from the OECD," *Review of World Economics*, 1998, *134* (2), 230–249.

Matsuyama, Kiminori, "The Home Market Effect and the Pattern of Trade Between Rich and Poor Countries," *mimeo Northwestern University*, 2015.

Morton, Fiona Scott and Margaret Kyle, *Markets for Pharmaceutical Products*, Vol. 2 of *Handbook of Health Economics*, Elsevier, 2012.

Ramondo, Natalia and Andres Rodríguez-Clare, "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, 2013, *121* (2), 273–322.

Trionfetti, Federico, "Using Home-Biased Demand to Test Trade Theories," *Review of World Economics*, 2001, *137* (3), 404–426.

Weder, Rolf, "Comparative home-market advantage: An empirical analysis of British and American exports," *Review of World Economics*, 2003, *139* (2), 220–247.

A   Multinational Enterprises (Section 3.1)

In this appendix, we illustrate how to incorporate multinational production into our basic environment. Following Ramondo and Rodríguez-Clare (2013), suppose that each firm headquartered in country i that sells drugs targeting disease n in country $j \neq i$ can choose the country l in which its production takes place. If $l = i$, then the firm exports, if $l = j$, it engages in horizontal FDI, and if $l \neq i, j$, it engages in platform FDI. At each production site l, we assume that the marginal cost of a firm from country i serving country j is constant and given by

$$c_{ij}^n(l) = c(\eta_i^n s_i^n)\kappa_{ij}^n(l), \tag{A.1}$$

where $c(s_i^n / \eta_i^n)$ captures the extent of economies of scale, which depends on the total supply of drugs targeting disease n in country i, and $\kappa_{ij}^n(l)$ captures the costs of local inputs in country l, the potential frictions associated with replicating Home's technology in country l, as well as the trade costs associated with shipping goods from country l to country j. In equilibrium, all firms from country i will serve country j from the location l that minimizes $c_{ij}^n(l)$ across all l and charge an equilibrium price,

$$p_{ij}^n = \min_l \{c_{ij}^n(l)\}. \tag{A.2}$$

Now set $p_i^n \equiv c(s_i^n / \eta_i^n)$ and $\tau_{ij}^n = \min_l \{\kappa_{ij}^n(l)\}$. By construction, equation (4) holds with $s(\cdot) = c^{-1}(\cdot)$, whereas equation (5) derives from equations (A.1) and (A.2).

B   Log-Linearization (Section 3.2)

In this appendix, we derive equation (7) by log-linearizing our model around a symmetric equilibrium with $\theta_j^n = 1$ for all j and n; $\eta_i^n = 1$ for all i and n; $\tau_{ii}^n = 1$ for all i and n and $\tau_{ij}^n = \tau$ for all n and $i \neq j$. We let D, d, and x denote aggregate drug consumption, the per disease expenditure on domestic drugs, and the per disease expenditure on drugs form any other country, respectively, in the symmetric equilibrium. In turn, we $\lambda = d/(d + (I-1)x)$ denote the share of expenditure on domestic drugs, with I the total number of countries. In the symmetric equilibrium, λ is also equal to the share of revenue on the domestic market. Finally, without loss of generality, we normalize all prices, and hence all price indices, to one in the symmetric equilibrium.

Up to a first-order approximation, for all n and $i \neq j$, equations (1), (2), and (5) imply

$$\ln x_{ij}^n = \ln x + \ln \theta_j^n - \epsilon^x (\ln p_i^n - \ln P_j^n + \ln \tau_{ij}^n - \ln \tau)$$
$$- \epsilon^D (\ln P_j^n - \ln P_j) + \ln D_j - \ln D + \ln p_i^n + \ln \tau_{ij}^n - \ln \tau.$$

with $\epsilon^x \equiv -(d \ln d(z)/d \ln z)_{z=\tau}$ and $\epsilon^D \equiv -(d \ln D(z)/d \ln z)_{z=1}$. By equation (3), we also know

34

that, up to a first-order approximation,

$$\ln P_j^n = \frac{\lambda(1-\epsilon^d)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}\ln p_j^n + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}\sum_{l\neq j}(\ln p_l^n + \ln \tau_{lj}^n - \ln \tau),$$

with $\epsilon^d \equiv -(d\ln d(z)/d\ln z)_{z=1}$. Combining the two previous expressions, we obtain

$$
\begin{aligned}
\ln x_{ij}^n =& \kappa_j + \ln \theta_j^n + (1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x})\ln p_i^n + \frac{\lambda(1-\epsilon^d)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}\ln p_j^n \quad \text{(B.1)} \\
&+ \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}\sum_{l\neq i,j}(\ln p_l^n + \ln \tau_{lj}^n) + (1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x})\ln \tau_{ij}^n,
\end{aligned}
$$

with

$$\kappa_j \equiv \ln x + (\epsilon^x - \frac{(1-\lambda)(1-\epsilon^x)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}(\epsilon^x - \epsilon^D) - 1)\ln \tau + \epsilon^D \ln P_j + \ln D_j - \ln D.$$

For $i = j$, the same logic implies

$$
\begin{aligned}
\ln x_{ii}^n =& \tilde{\kappa}_i + \ln \theta_i^n + (1 - \epsilon^d + \frac{\lambda(1-\epsilon^d)(\epsilon^d - \epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x})\ln p_i^n \quad \text{(B.2)} \\
&+ \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^d - \epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}\sum_{l\neq i}(\ln p_l^n + \ln \tau_{li}^n),
\end{aligned}
$$

with

$$\tilde{\kappa}_i \equiv \ln d - \frac{(1-\lambda)(1-\epsilon^x)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x}(\epsilon^d - \epsilon^D)\ln \tau + \epsilon^D \ln P_j + \ln D_j - \ln D.$$

Next, let us compute producer prices around a symmetric equilibrium. Up to a first-order approximation, for all i and n, equations (4), (5), and (6) imply

$$(1+\epsilon^s)\ln p_i^n + \ln \eta_i^n = \lambda(\ln x_{ii}^n - \ln c) + \frac{(1-\lambda)}{I-1}\sum_{j\neq i}(\ln x_{ij}^n - \ln x),$$

with $\epsilon^s \equiv (d\ln s(z)/d\ln z)_{z=1}$. Together with equations (B.1) and (B.2), this implies

$$(\epsilon^s + \epsilon^w)\ln p_i^n - \frac{(\epsilon^w - \epsilon^D)}{I-1}\sum_{l\neq i}\ln p_l^n = \lambda \ln \theta_i^n + \frac{(1-\lambda)}{I-1}\sum_{j\neq i}\ln \theta_j^n + \xi_i^n,$$

with

$$\epsilon^w \equiv \lambda\epsilon^d + (1-\lambda)\epsilon^x - \frac{\lambda^2(1-\epsilon^d)(\epsilon^d - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x} - \frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x}\frac{(1-\lambda)^2}{I-1},$$

$$\xi_i^n \equiv \lambda(\tilde{\kappa}_i - \ln d) + \frac{(1-\lambda)}{I-1}\sum_{j\neq i}(\kappa_j - \ln x) - \ln\eta_i^n$$

$$+ \lambda\frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^d - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x}\sum_{l\neq i}\ln\tau_{li}^n + \frac{(1-\lambda)^2}{(I-1)^2}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x}\sum_{j\neq i}\sum_{l\neq i,j}\ln\tau_{lj}^n$$

$$+ \frac{(1-\lambda)}{I-1}(1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x})\sum_{j\neq i}\ln\tau_{ij}^n.$$

The solution to the previous system is given by

$$\ln p_i^n = \frac{(\lambda - \frac{1-\lambda}{I-1})}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}}\ln\theta_i^n + \frac{\xi_i^n}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}} + \zeta^n, \tag{B.3}$$

with

$$\zeta^n \equiv \frac{(\frac{1-\lambda}{I-1} + \frac{\epsilon^w - \epsilon^D}{(\epsilon^s + \epsilon^D)(I-1)})\sum_j\ln\theta_j^n + \frac{1}{(\epsilon^s + \epsilon^D)}\frac{(\epsilon^w - \epsilon^D)}{I-1}\sum_l\xi_l^n}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}}.$$

Combining equations (B.1) and (B.3), we obtain, for any n and $i \neq j$,

$$\ln x_{ij}^n = \delta_{ij} + \delta^n + \beta_M\ln\theta_j^n + \beta_X\ln\theta_i^n + \varepsilon_{ij}^n,$$

36

with

$$\delta_{ij} \equiv \frac{(1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x})\bar{\xi}_i + \frac{(1-\epsilon^d)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}\lambda\bar{\xi}_j + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}\sum_{l \neq i,j}\bar{\xi}_l}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}}$$

$$+ (1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x})(\overline{\ln \tau_{ij}}) + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}\sum_{l \neq i,j}(\overline{\ln \tau_{lj}}) + \kappa_j,$$

$$\delta^n \equiv \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}(\lambda - \frac{1-\lambda}{I-1})(\sum_l \ln \theta_l^n)/(\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}) + (1 - \epsilon^D)\zeta^n,$$

$$\beta_M \equiv 1 + \frac{(\epsilon^x - \epsilon^D)(\frac{\lambda(1-\epsilon^d)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x} - \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x})(\lambda - \frac{1-\lambda}{I-1})}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}},$$

$$\beta_X \equiv \frac{(1 - \epsilon^x)(\lambda - \frac{1-\lambda}{I-1})}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}},$$

$$\epsilon_{ij}^n \equiv \frac{(1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x})(\xi_i^n - \bar{\xi}_i) + \frac{(1-\epsilon^d)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}\lambda(\xi_j^n - \bar{\xi}_j) + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}\sum_{l \neq i,j}(\xi_l^n - \bar{\xi}_l)}{\epsilon^s + \epsilon^w + \frac{\epsilon^w - \epsilon^D}{I-1}}$$

$$+ \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x}\sum_{l \neq i,j}(\ln \tau_{lj}^n - (\overline{\ln \tau_{lj}})) + (1 - \epsilon^x + \frac{1-\lambda}{I-1}\frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1-\lambda\epsilon^d - (1-\lambda)\epsilon^x})(\ln \tau_{ij}^n - (\overline{\ln \tau_{ij}})),$$

where $\bar{z} \equiv \frac{1}{\#diseases}\sum_n z^n$ denotes the arithmetic average of a given variable z across all diseases.

C   Robustness (Section 3.3)

C.1   Monopolistic Competition

For each disease n, profit-maximization by a firm ω from country i selling in country j requires

$$\frac{p_{ij}^n(\omega) - \tau_{ij}^n c_i^n}{p_{ij}^n(\omega)} = \frac{1}{\sigma}. \tag{C.1}$$

Free entry requires

$$\sum_j (p_{ij}^n(\omega) - \tau_{ij}^n c_i^n)d_{ij}^n(\omega) = f_i^n,$$

which can be rearranged as

$$s_i^n(\omega) = \frac{f_i^n}{p_i^n(\omega) - c_i^n}, \tag{C.2}$$

with $s_i^n(\omega) \equiv \sum_j \tau_{ij}^n d_{ij}^n(\omega)$. By definition, we also know that

$$s_i^n = \left(\int (s_i^n(\omega))^{(\sigma-1)/\sigma} d\omega \right)^{\sigma/(\sigma-1)}, \tag{C.3}$$

$$p_i^n = \left(\int (p_i^n(\omega))^{(1-\sigma)} d\omega \right)^{1/(1-\sigma)}. \tag{C.4}$$

Equations (C.1), (C.2) and (C.3) imply

$$s_i^n = (N_i^n)^{\sigma/(\sigma-1)} f_i^n / ((\mu - 1) c_i^n), \tag{C.5}$$

whereas equations (C.1) and (C.4) imply

$$p_i^n = (N_i^n)^{1/(1-\sigma)} \mu c_i^n. \tag{C.6}$$

Finally, note that equations (C.2) and (C.5) imply

$$s_i^n = (N_i^n)^{\sigma/(\sigma-1)} \sum_j \tau_{ij}^n d_{ij}^n(\omega),$$

for any firm ω, whereas equations (13), (C.1), and (C.6) imply

$$d_{ij}^n(\omega) = (N_i^n)^{-\sigma/(\sigma-1)} d_{ij}^n.$$

Equation (6) follows from the two previous expressions.

C.2   Bertrand Oligopoly

The demand for varieties produced by an individual firm, $d_{ij}^n(\omega)$, is given by equations (1), (2), and (13). Under the assumption of an arbitrarily large number of sectors, firms' decisions in any given sector have no effect on the country-specific demand shifters, P_j and D_j. Hence, the elasticity of demand with respect to a firm's own price is such that

$$\frac{d \ln d_{ij}^n(\omega)}{d \ln p_{ij}^n(\omega)} = \begin{cases} -\sigma + \frac{(\sigma - \epsilon^x)}{N_i^n} + \frac{(\epsilon^x - \epsilon^D)}{N_i^n} \frac{(1-\epsilon^x)}{1 - \lambda \epsilon^d - (1-\lambda)\epsilon^x} \frac{(1-\lambda)}{I-1} & \text{, if } i \neq j, \\ -\sigma + \frac{(\sigma - \epsilon^d)}{N_i^n} + \frac{(\epsilon^d - \epsilon^D)}{N_i^n} \frac{(1-\epsilon^d)}{1 - \lambda \epsilon^d - (1-\lambda)\epsilon^x} \lambda & \text{, if } i = j. \end{cases}$$

Under the additional assumption that $\epsilon^D = \epsilon^x = \epsilon^d$, the elasticity of demand simplifies into

$$\frac{d \ln d_{ij}^n(\omega)}{d \ln p_{ij}^n(\omega)} = -\sigma + \frac{(\sigma - \epsilon^d)}{N_i^n}.$$

For each disease n, profit-maximization by a firm ω from country i selling in country j therefore now requires

$$\frac{p_{ij}^n(\omega) - \tau_{ij}^n c_i^n}{p_{ij}^n(\omega)} = \frac{1}{\sigma - (\sigma - \epsilon^d)/N_i^n}. \qquad (C.7)$$

The free entry condition (C.2) remains unchanged, whereas equations (C.3) and (C.4) still hold with $\sum$ rather than $\int$.

Equations (C.2), (C.3), and (C.7) now imply

$$s_i^n = (N_i^n)^{\sigma/(\sigma-1)} f_i^n / ((\mu(N_i^n) - 1)c_i^n),$$

with $\mu(N_i^n) \equiv \frac{((1-1/N_i^n)\sigma + \epsilon^d/N_i^n)}{(1-1/N_i^n)\sigma + \epsilon^d/N_i^n - 1}$, whereas equations (C.4) and (C.7) imply

$$p_i^n = (N_i^n)^{1/(1-\sigma)} \mu(N_i^n) c_i^n.$$

C.3 Monopoly

The first-order conditions associated with profit maximization imply

$$\frac{p_{ij}^n - \tau_{ij}^n c_i^n}{p_{ij}^n} = \frac{1}{\epsilon^d},$$

$$s_i^n = -\eta_i^n f'(c_i^n),$$

with $s_i^n = \sum_j \tau_{ij}^n d_{ij}^n$ the total quantity produced by the monopolist. Combining the two previous expressions, we immediately obtain

$$s_i^n = -\eta_i^n f'((\epsilon^d - 1)p_i^n/\epsilon^d).$$

D Bilateral Sales (Section 6.1)

In Appendix B, we have already shown that

$$\ln x_{ij}^n = \alpha_j + \ln \theta_j^n + (1 - \epsilon^x + \frac{1-\lambda}{I-1} \frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x}) \ln p_i^n + \frac{\lambda(1-\epsilon^d)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x} \ln p_j^n$$

$$+ \frac{1-\lambda}{I-1} \frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x} \sum_{l \neq i,j} (\ln p_l^n + \ln \tau_{lj}^n) + (1 - \epsilon^x + \frac{1-\lambda}{I-1} \frac{(1-\epsilon^x)(\epsilon^x - \epsilon^D)}{1 - \lambda\epsilon^d - (1-\lambda)\epsilon^x}) \ln \tau_{ij}^n.$$

This can be rearranged as

$$\ln x_{ij}^n = \alpha_j^n + (1 - \epsilon^x) \ln p_i^n + (1 - \epsilon^x) \ln \tau_{ij}^n,$$

with $\alpha_j^n \equiv \alpha_j + \ln \theta_j^n + \frac{\lambda(1-\epsilon^d)(\epsilon^x-\epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x} \ln p_j^n + \frac{1-\lambda}{I-1} \frac{(1-\epsilon^x)(\epsilon^x-\epsilon^D)}{1-\lambda\epsilon^d-(1-\lambda)\epsilon^x} \sum_{l\neq j}(\ln p_l^n + \ln \tau_{lj}^n)$.

E   Additional empirical analysis

Table E.1 establishes that predicted disease burden is indeed a strong predictor of a country's actual disease burden, even conditional on country and disease fixed-effects (which we condition on whenever we use the predicted disease burden in our tests of the home-market effect). Column (1) shows that the predictive power of a country's demographic composition, interacted with the demographic disease pattern of a disease, is substantial within a sample of countries that purchase at least some drug (that is, countries j with $\sum_{i,n} x_{ij}^n > 0$). And column (2) establishes the same feature in a sample of selling countries (those countries i with $\sum_{j,n} x_{ij}^n > 0$).

Table E.1: Predicting disease burden using demographic variation

	log(disease burden)	
	(1)	(2)
log(predicted disease burden)	1.769	1.618
	(0.370)	(0.356)
Sample of destination countries (j such that $\sum_{i,n} X_{ij}^n > 0$)	✓	
Sample of origin countries (i for which $\sum_{j,n} X_{ij}^n > 0$)		✓
Adjusted R^2	0.886	0.900
Observations	2,311	2,930

Notes: For details on construction of variables, sample restrictions and standard error (reported in parentheses) calculations see notes to Table 3. All specifications control for country and disease fixed-effects.