International Diffusion of Technology: Accounting for Heterogeneous Learning Abilities

Diana Van Patten†

UCLA

Preliminary and Incomplete

Abstract

An important question in the field of economic growth and development is how developing countries learn to adopt and use new technologies. This paper studies how countries learn from each other through international trade. First, I build a panel of bilateral trade flows between industries in different countries. Matching this panel with data on industry-level productivity, I document how productivity grows systematically faster for countries that trade with partners with better technologies, but that this is reducing the gap between local and foreign productivity. Second, I build a model in which knowledge transfers can occur through imported technology, leading to productivity growth. In my framework, agents have heterogeneous learning abilities: The probability of a producer adopting a technology slightly better than hers is larger than the probability of adopting a much more sophisticated one —the trade-off being that conditional on adoption, more sophisticated technologies lead to higher productivity. I document how the model matches the empirical dependence of productivity growth on productivity gaps across trading partners, and the firm size distribution. The model also highlights how ignoring differences in learning abilities can overestimate the impact of exposure to high-TFP trading partners, leading to suboptimal trade policies. I conclude that developing countries should direct relatively more trade to mid-productive countries —as opposed to very productive ones— to maximize technology transfers and increase growth.

Keywords: International Trade, Learning, Economic Growth
JEL Codes: D83, F43, O11.

†dvpr@ucla.edu. I thank Andrew Atkeson, Manuel Amador, Francisco Buera, Ariel Burstein, Pablo Fajgelbaum, Doireann Fitzgerald, Hugo Hopenhayn, Timothy Kehoe, Lee Ohanian, Edward Prescott, Víctor Ríos-Rull, Esteban Rossi-Hansberg, Edouard Schaal, Vladimir Smirnyagin, Vitaly Titov, Jonathan Vogel, and seminar participants at UCLA, Minnesota and the Workshop on Macroeconomic Dynamics in Vigo for their helpful comments and discussion.

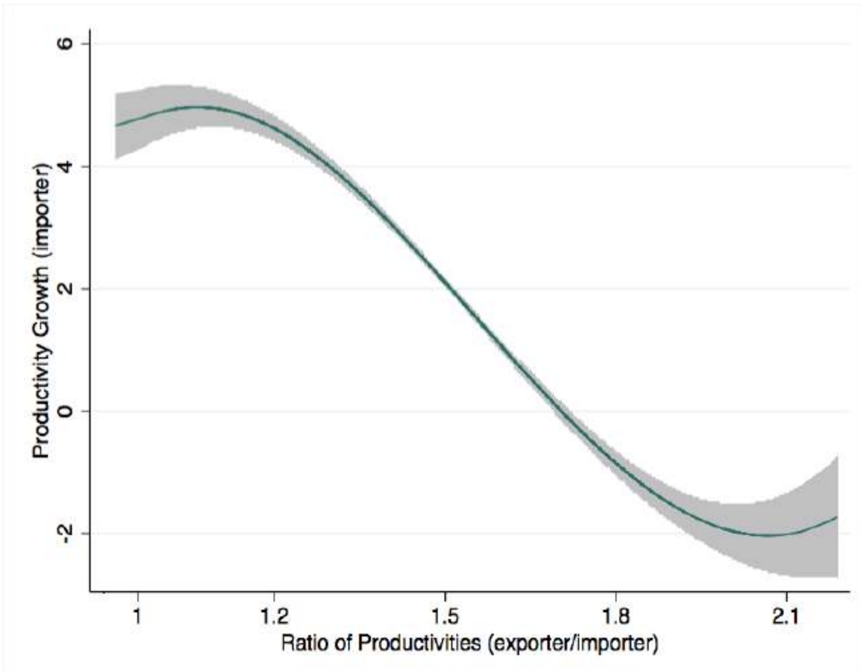
1 Introduction

There are well-known gains from trade for developing countries when trading with developed countries, such as lower prices and less competition with locally produced goods. However, these channels have proven to be insufficient in delivering sizable gains from trade. A recent body of literature has focused on dynamic gains from trade as a key element to understand how trade can be an engine of significant growth (Alvarez et al. (2013); Buera and Oberfield (2016); Perla et al. (2015); Sampson (2014)). This paper proposes how, in terms of dynamic gains from trade, countries are better off trading with partners who have higher —but similar— levels of development.

In general, papers that study the mechanism through which international trade interacts with economic growth develop models with a wide range of learning processes through which diffusion takes place. However, little is known about this type of knowledge transfer across countries. A natural question is whether the model is a good representation of the true learning process behind knowledge transfers across countries, and this paper explores the question empirically. Who learns from whom? Is the magnitude of these technology transfers economically and statistically significant? Does this depend on the distance between the levels of knowledge of the parties involved? To answer these questions, I build a panel that spans over two decades, consisting of a network of industry-country pairs, the trade flows between each of them, and data on industry-country productivity per year.

Using this data, I find, first, that the growth rate of productivity of importers is increasing in their trading partners' productivity, which is in line with the mechanism of Buera and Oberfield (2016). Second, there is a negative relationship between the importer's productivity growth rate and the ratio of productivities of the trading partners: If the ratio is very small, increasing it leads to more productivity growth. This is intuitive: If both parties know roughly the same, there are little gains from interacting. However, if the ratio is relatively large, increasing it leads to less productivity growth. The latter is

Figure 1: Relationship between self-productivity growth and distance to trading partners' productivities.



Notes: Figure 1 shows a 5-degree polynomial obtained after running a panel non-parametric regression with fixed effects, with 95% confidence intervals shown in gray. The dependent variable is productivity growth of an importer from industry k in country i from period t to period $t+1$. The independent variable shown is the ratio between the productivity of each exporter (industry-country specific) to the productivity of importer from *the same industry* in country i in period t, from 1995-2011. The regression also includes the share of imports, the gap of productivities, and fixed effects, as described in detail in Section **??**.

also intuitive; if you do not know how to multiply, attending a quantum physics class might not be useful. This relationship is shown in Figure 1. Through the lens of the literature on diffusion of knowledge through trade, this is informative regarding the bounds to learning that intervene in this diffusion process. Third,the data suggest little congestion and no magnification effect in learning, as the knowledge transfer depends on the interaction between partners, but not on the share of goods being traded. I document how these effects are present over time, and both across and within industries. I also use the China shock, as well as an instrumental variable based on Autor et al. (2013) and revisit my regressions, finding little quantitative and no qualitative change in my results.

Guided by this evidence, I extend the benchmark model of Buera and Oberfield (2016) and develop a model in which producers can learn from their trading partners, and this learning is heterogeneous: Given a match, the probability of learning depends on the gap in knowledge of the parties that interact. This mechanism will allows for countries to learn easier from partners who are "close" to them, in terms of productivity. To validate the model, I rerun all my initial regressions using model-generated data, and find that learning heterogeneity allows the model to match the empirical coefficients, while shutting down this heterogeneity leads to largely counterfactual results.

Using the model, I find sizable differences in the predicted gains from trade when shutting down the heterogeneity in learning abilities, and in particular when comparing the contribution of trade to TFP dynamics for growth miracles. Moreover, in the extended model, when opening to trade, highly productive agents are learning the most. This is consistent with Steinwender (2015), who finds evidence that there are productivity increases after expanding export markets, but only for firms that were ex-ante the more productive ones. Learning heterogeneity is also crucial to match the firms size distribution and be consistent with Gibrat's law for firms.

Policy implications differ between models: For developing countries, the extended model implies inducing larger optimal trade shares with mid-developed countries, as opposed to trading more with very productive and developed ones. Moreover, this model allows —theoretically— for a divergence in welfare between low- and high-income countries.

The model and its calibration strategy rely on standard, but relatively strong, assumptions. However, they introduce heterogeneity in learning abilities in the general Buera-Oberfield benchmark model, without losing its tractability. This force could be introduced in a similar way in a wide range of models, including Alvarez et al. (2013), Perla et al. (2015), and Sampson (2014). As such, I view the structure imposed, based on the empirical results I obtained, as a natural way to study the general effects this important, but usually ignored, force can have. My empirical strategy using industry-country shares and productivities and my proposed model can be used to study a variety of phenomena that might affect learning across countries at an aggregate level, dynamically, and differentially across sectors.

This paper belongs to a strand of literature studying the interaction between growth, trade and knowledge. This paper contributes to the literature on diffusion of ideas through trade, by providing a foundation as to why learning is heterogeneous depending on the relative level of development of the trading partners. The earliest models studying endogenous growth and the role of knowledge generally assumed knowledge was a public good, as in the learning-by-doing models of Arrow (1962), or Grossman and Helpman (1990) where searching for new technologies was rewarded with patents. Later on, Romer (1990) published a seminal paper on the role of ideas and knowledge in sustained growth.

In models of diffusion, producers improve their productivity by learning from others, but interpretations of the specific learning process have varied considerably. This paper proposes a new intuitive approach to modeling learning: Embed heterogeneous learning probabilities in the workhorse model of Buera and Oberfield (2016). In my model, conditional on learning, there are more gains from meeting better producers or running into better ideas, such as in Jovanovic and Rob (1989), Kortum (2008), Eaton and Kortum (1999), Alvarez et al. (2013) and Buera and Oberfield (2016). However, in my model there is a trade-off: the probability of learning from a meeting is decreasing in the gap between a manager's current level of knowledge and the level of the manager she meets.

The paper also relates to the literature on diffusion and development, as it captures the limited extent to which technologies can spread across countries when there is a large gap between the levels of development of the trading partners. The former is crucial when analyzing the key aspects in which less developed countries should invest in order to increase their local productivity, in particular through learning from trading partners with more advanced technologies. In this sense, my work also relates to the literature on absorptive capacity. The notion of absorptive capacity dates back to Cohen and Levinthal (1989), who argue that a firm's innovative capabilities depend crucially on its ability to identify external information, assimilate it, and apply it to commercial ends.

Several other researchers have emphasized the difficulties in adapting advanced technologies, especially to the needs of less developed countries, such as Findlay (1978). Evenson and Westphal (1995) study how convergence between countries is slowed down as new technologies need a significant amount of "tacit knowledge", while Atkinson and Stiglitz (1969) and Stewart (1977) study the relevance of the "appropriateness" of technology. More recently, Acemoglu and Zilbotti (1999) considers the differences in relative supplies of skills across countries in a model in which technologies designed for skilled workers cannot be perfectly adopted by countries with mostly low-skilled labor.

The rest of the paper is organized as follows. Section 2 describes the data and the reduced-form evidence. Section 3 includes the theoretical model and an extended version of it that allows for international trade. A quantitative analysis using the model is presented in Section 4, which also includes my calibration strategy, the model's validation, an analysis of TFP dynamics, and a description of the main implications of the model. Section 5 concludes.

2 Reduced-Form Evidence

Guided by the theoretical model, this section aims to study the impact of trade interactions on productivty growth. My goal is to provide empirical evidence on the true learning process behind the diffusion of ideas through trade. Moreover, this will allow me to assess the relevance of productivity gaps in this process.

To do so, I begin by showing reduced-form evidence of the key mechanism in the model: learning being easier for trading partners who have similar TFP. I also find evidence of the learning in Buera and Oberfield (2016): productivity growth is significantly increasing in the productivity level of the trading partner. After establishing these correlations, I then turn to an instrumentation strategy for changes in exposure to trading partners, discuss the data used in the analysis, and present of my empirical findings.

I then use these reduced-form regressions to validate my model, and show how my mechanism is key to match both the sign and magnitude of the relations found empirically, by comparing it with the case of homogeneous learning.

2.1 Regression Framework

In this section, I use panel data on trade flows and productivity growth to investigate the relationship between trade-partner productivity and self-productivity growth. This will be informative about the true learning process behind the international diffusion of ideas, and about the relevance of productivity gaps in this process.

To do so, I begin in a reduced-form fashion relating the productivity growth of each industry-country pair to the productivity of its trading partners, and the ratio between self-productivity and each partner's productivity, as a measure of the "distance" between both technologies, using various flexible reduced-form specifications.

I am particularly interested in two questions: first, does productivity increase more when my trading partners have high productivity levels? Second, if so, how does this depend on the ratio between my productivity and his? Further, I present various robustness checks, with a particular focus on ruling out high non-linearities in this relationship that could flip the result, mean reversion, and a spurious relation not related to trade.

While none of the reduced-form specifications is tightly grounded in my theory, I nonetheless argue that the resulting picture is useful in understanding the true learning process behind learning from trade; a topic where the literature offers a menu of models with different mechanisms and implications, but little empirical grounds. Moreover, the regressions are useful as means of model validations in Section 4.3.

I begin with the following baseline specification:

$$\Delta\% z^{v}_{it+1} = \beta_1 + \beta_2 log\left(\frac{z^{u}_{jt}}{z^{v}_{it}}\right) + \beta_3 log\left(z^{u}_{jt}\right) + \beta_4 log\left(\frac{z^{u}_{jt}}{z^{v}_{it}} \times z^{u}_{jt}\right) + \beta_5 log\left(\pi^{uv}_{ij}\right) + \Psi + \epsilon_t, \quad (1)$$

where z^{v}_{it} is the log productivity of industry v in country i[1], and z^{u}_{jt} is the productivity a trading partner from industry u in country j. π^{uv}_{ij} denotes shares of imports (as producers learn from sellers) of each industry-country pair. Finally, Ψ captures industry, country, and year fixed effects.

Using productivity growth through Equation 3, provides a strategy for deriving a version of this reduced-form equation from the model. The details of this derivation can be found in Appendix C. However, to address issues suppressed in the theory but likely to matter in the estimation, I will not focus on structural regressions in the analytical model, but on reduced-form regressions that are motivated by the model but that do not identify structural parameters. Instead, in the quantitative analysis of Section 4.3, I will use these regressions for model validation, by running the same non-structural regressions in data generated by the extended model.

Equation 1 describes crucial relations from the model. It allows me to estimate whether the gap between the productivity of the learner (the importer) and the productivity of its trading partners (the exporters) induce on average an increase or a decrease in productivity growth, thereby allowing me to test if the mechanism behind my model is relevant empirically, and shed light on the true learning process behind learning from trade. A negative β_2 would support the idea of absorptive capacity constraints: the larger the ratio, the harder it is to learn given interactions take place.

The coefficient β_3 tests forces that should be present in this model as well as in the

[1] All productivities are demeaned using the country-industry mean across the sample (1995-2010), to make their magnitudes comparable across sectors

original Buera Oberfield model: that larger partners should lead to larger gains conditional on the productivity gap. I include the coefficient β_4 to allow the effect of the ratio to vary with the partner's productivity level. Finally, β_5 is included to allow for the share of imports to play a role. Note that, in the model, this share is irrelevant (and empirically it proves to be insignificant), however I include it as an effort to make the learning process more flexible in this empirical exploration.

2.2 Data

In my baseline analysis, I study changes in productivity between 1995 and 2010. In the sensitivity analysis, I use 2007 as an alternative end year to exclude the great recession. The data used to construct the import shares comes from the World Input-Output Database (WIOD)[2]. Data used to obtain the productivities per sector and per country comes from the GGDC database and world KLEMS data, which include data on real GDP per hours worked for a panel of 33 countries[3] dividing economic activity into 10 sectors[4]. Using the industry codes, I can map the sectors from the WIOD to these 10 aggregate sectors. this mapping is included in Appendix D.

For my dependent variable, I use the the log change in productivity for an industry within a country shown in 1; for a panel composed by each of the 10 industries in the 31 countries studied. For my independent variables, I measure the log of the ratio of the productivity of the synthetic trading partner, constructed as explained in Section **??**.

[2]The WIOD traces the flow of goods and services across 35 industries, 40 coutnries, and a constructed rest of the world (Timmer et al., 2015b).

[3]The countries included in this dataset are, from Africa: Botswana, Ethiopia, Ghana, Kenya, Malawi, Mauritius, Nigeria, Senegal, South Africa, Tanzania, and Zambia; from Asia: China, Hong Kong, India, Indonesia, Japan, South Korea, Malaysia, Philippines, Singapore, Taiwan, and Thailand; from Latin America: Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Mexico, Peru, and Venezuela.

[4]These sectors are: agriculture, hunting, forestry and fishing; mining and quarrying; manufacturing; electricity, gas and water supply; construction; wholesale and retail trade, hotels and restaurants; transport, storage, and communication; finance, insurance, real estate and business services; government services; community, social and personal services. Data is avalailable from 1950 - 2010; details on the construction of these sectors can be found in citetggdc.

As explained before, these productivities are demeaned using the average productivity of each industry within each country across the years of the study (1995-2010), to make them comparable across industries.

2.3 Empirical Results

The results from running 1 are shown in Table 1. In column (1), I exclude both the trading partner's productivity and the interaction term. This naive regression shows a correlation that is negative and large (-0.56) between the change in the productivity of a country and the gap between its productivity and that of its trading partners. Namely, a 1% increase in the ratio (trading partner's z/own z) leads to a decrease in productivity growth of -0.56 percentage points(pp); a sizable and significant effect.

Column (2) captures a force that is central in the work of Buera and Oberfield (2016): that more productive trading partners (sellers, in particular) lead to more productivity growth. The coefficients are sizable hand have the expected signs: (a) given a ratio, a more productive partner leads to better ideas, namely, a 1% increase in the trading partner's productivity leads to a 0.94 pp increase in productivity growth; and (b) a larger ratio limits learning with a 1% increase in the ratio leading to a -1.46 pp decrease in productivity growth. Moreover, it highlights the importance of the ratio: given the magnitudes of the log variables, Column (2) shows that the role of the ratio can be even larger than the one of the partner's productivity level.

In Column (3), I include the interaction term and allow the ratio to have differential effects on productive and improductive trading partners; while Column (4) incorporates the trade share. First, Column (4) shows that on average the effect of a large ratio is smaller for very productive trading partners, although the magnitude of this effect is small (-0.002). This result is consistent with the theoretical model: a large ratio implies a low probablility of adoption, however, if adoption occurs, the larger the productivity of the

Table 1: Trade and Changes in Productivity: Industry-Country Pairs

Dependent variable: change in productivity of each country's industry $\left(\Delta z^k_{it+1}\right)$.

Log Variables	(1)	(2)	(3)	(4)
Ratio $\left(\frac{z^u_{jt}}{z^v_{it}}\right)$	-0.56^{***}	-1.46^{***}	-1.55^{***}	
	(0.014)	(0.037)	(0.038)	
Trading partner's productivity $\left(z^u_{jt}\right)$		0.94^{***}	0.83^{***}	
		(0.036)	(0.037)	
$\left(\frac{z^u_{jt}}{z^v_{it}}\right) \times z^u_{jt}$			-0.002^{***}	
			(0.000)	
Trade share $\left(\pi^{uv}_{ij}\right)$				
				(0.006)
FE	✓	✓	✓	✓
Adj R^2				

Notes: Table 1 reports the regression results when the dependent variable is change in productivity of each country's industry $\left(\Delta z^k_{it+1}\right)$. All productivities are demeaned using the mean of a country's industry during the period studied (1995-2010). Independent variables are in logs. In all cases, the number of observations equals 1,889. Constants not reported. Robust standard errors are in parentheses. Significance levels: $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

trading partner the more an agent would be able to learn.

These results highlight a new important element in the interaction of trade and productivity growth. The combination of a low productivity (e.g., manufacturing in Tanzania) and having a very large gap with your trading partners (e.g., machinery in Switzerland[5]) significantly decreases the dynamic gains from trade that come from ideas' diffusion. In Section 4, I will use my quantitative framework to show, in a generalized model, how this new channel can contribute to understanding the growth we see in countries like growth miracles through an examination of the TFP dynamics in these countries.

Within-Industry Results In line with the model presented in Section 3, the previous section ran Regression 1 pooling across industries. This allows for producers to learn from importers from any industry, not only their own (for instance, they could learn about managerial practices that are applicable to a range of industries, or about how to apply technologies from other industries in their own). Although gains across fields are not hard to imagine, in this section I explore if the same gains are found within industries. That is, forcing $u = v$ for all u in Regression 1. This can be informative on the nature of the learning process behind gains from trade. The results are shown in Table **??**.

Non-linearities As the relationship between the ratio of productivities and the productivity growth of the importer might be non-linear in nature, I also perform a more flexible non-parametric regression, which which aims to capture any non-linear relationship with a polynomial of degree 5. The result when running the regression in Column (2) within industries is shown in Figure 1. A very similar result when pooling across industries is shown in Appendix .

As shown in the figure, there is a non-linear relationship between the importer's productivity growth rate the ratio of productivities of the trading partners, however produc-

[5]Switzerland is Tanzania's main trading partner.

tivity growth is consistently decreasing as the ratio increases. Through the lens of the literature on diffusion of knowledge through trade, this is informative of the bounds to learning that intervene in this diffusion process, especially when studying trade between countries or industries where the gap in productivities might be sizable (like it typically occurs in North-South trade). Note how this non-linearity is different in nature from what one would expect when learning from coworkers (Jarosh et al. (2018), Herkenhoff et al. (2018)), as the nature of learning is different when trying to acquire knowledge solely from trade: this transfer of technology and knowledge is occurring through imported technology and exposure to foreigners, not through face-to-face everyday interactions or direct teaching.

The China Shock and an Instrumental Variables Approach Although suggestive of significant learning from trading partners, these findings could in principle be contaminated by other events occurring in the world market, or be spurious in nature. To address this issue, I use the China shock and revisit all my regressions, constructing instruments for the trade shares in a fashion similar to Autor et al. (2013). Namely, I consider only imports coming from China, and calculate the change in variables from 1995 to 2007 for a subsample of countries for which it is possible to construct a strong Autor, Dorn Hanson instrument, so that:

$$\Delta\% z^{v}_{i2007} - \Delta\% z^{v}_{i1995} = \beta_2 (z^{u}_{j2007} - z^{u}_{j1995}) + \beta_3 (\pi^{vu}_{ij2007} - \pi^{vu}_{ij1995}) + \Psi + \epsilon. \tag{2}$$

Details of this instrument are available in Appendix **??**. The results for the within-industry are shown in in columns (1) and (2) of Table **??**, while the results while pooling across industries are shown in columns (1) and (2) Table **??** . As the tables, I find little no qualitative change in my results, and although the magnitudes are not directly comparable (as these new results are in log-changes), the effects of the ratios are still economically significant in size; even larger than previously found.

Robustness Even after checking the robustness of the results non-parametrically and through the China shock , the results embody an assumption about the relevant time period for the analysis, and about how long it takes for the learning mechanism to kick-in and for trade to have a causal effect on productivity.

Beginning with the latter, one concern about the specification may be that a year is not enough for effect to materialize, or at least not fully. Given how learning works in the model as derived in Section 2.1, the main specification only includes the first lag. To address this concern, Appendix E, includes results in which I inlcude further lags for the independent variables. This exercise allows me to assess if there is a “time-to-build" the stock of knowledge after a trade interaction, and whether or not it is empirically correct to include only the first lag. This exercise, discussed in more detail in Appendix E, reveals no evidence of lags of order higher than 1 having more importance than the first one, as they are not larger in magnitude nor significant for any ratio. Moreover, all the qualitative findings remain unchanged with little impact on the magnitudes of the coefficients.

Finally, I examine the sensitivity of my estimates to the time period studied. Namely, I rerun the regressions over the 1995-2007 period, excluding the great recession from the sample, also with little impact on the results as described in Appendix E.

3 Theoretical Framework

The model presented in this section combines the framework of global diffusion presented in Buera and Oberfield (2016) with the idea of absorption capacity (Cohen and Levinthal, 1989) and sectoral learning. This leads to a model that nests the original one, but has different key results (both qualitative and quantitative).

3.1 Learning Environment

There is a continuum of goods $s \in [0,1]$, and for each good there are m producers, each of them with a knowledge q. The production function is given by $y(s) = q * l(s)$, where $l(s)$ is the labor input and $y(s)$ represents the output of good s. $M_t(q)$ is the fraction of producers with knowledge no greater than q, and the frontier of knowledge takes the form of $\tilde{F}_t(q) = M_t(q)^m$.

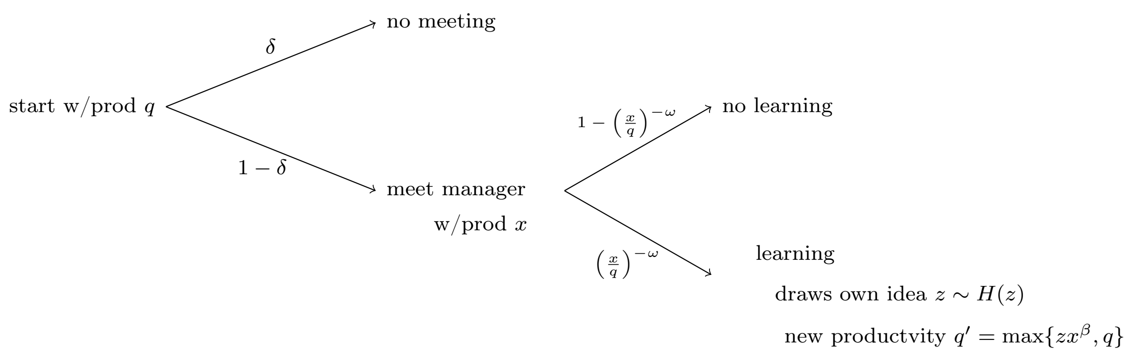
Each period t, a producer begins with a level of knowledge q, and with probability $1-\delta_t$ is not matched with anyone, and thus keeps the same level q next period, i.e. $q_t = q$. With probability δ_t, the producer is matched with another producer with knowledge x drawn from a distribution with CDF $\tilde{G}(x)$.

An important point is that in the model, *not all matches lead to learning*, even if a manager matches with someone with higher productivity than her. In particular, given a match, a producer will be able to assimilate its counterpart's idea with probability $\left(\frac{x}{q}\right)^{\omega}$, which follows the spirit presented in Lucas and Moll (2011). More precisely, the learning probability would be

$$\max\left\{\left(\frac{x}{q_t}\right)^{-\omega},\ 1\right\},\quad \omega > 0$$

If this probability realizes, and the producer learns, then she draws a second idea (can be interpreted as her own idea) z from an exogenous distribution with CDF $H(z)$. Finally, the producer adopts the new hybrid idea if $q < zq'^{\beta}$. The timing of the learning process is summarized in Figure 2.

Figure 2: Timing of the Learning Process



Note that there is a trade-off: on one hand, a producer wants to match with someone with a productivity that is similar to hers, to make the probability of learning high. On the other hand, conditional on learning, she wants to match with someone with a productivity x that is as high as possible. The model is isomorphic to Buera and Oberfield (2016) if $\omega = 0$. Iin this $\omega = 0$ case, the gap in productivity levels plays no role in the probability of learning. Including $\omega > 0$ prevents someone from getting a very good draw of x and becomes a superstar producer overnight. For example, imagine importing books about quantum physics, while struggling with the basics of multiplication; even if there is a match, an insight about quantum physics might not transmitted, *and how hard the idea is to transmit will depend on the gap between both levels of knowledge.*

Dynamics of the Distribution of Knowledge Next, I provide conditions such that the frontier of knowledge converges to a Frechet distribution. Then, the state of knowledge can be summarized using this distribution's level, called the "stock of knowledge". The model is thus compatible with Eaton and Kortum (2002) machinery, and can be used to study trade flows in an environment with asymmetric countries, characterizing the stocks of knowledge only in terms of trade shares and parameters.

Given the distribution of knowledge at time t, $M_t(q)$, the source distribution $\tilde{G}(q)$, and the exogenous distribution of ideas $H(z)$, the distribution of knowledge at time $t + \Delta$

is given by[6]

$$M_{t+\Delta}(q) = M_t(q)[1 - \delta_t \Delta \int_0^\infty \left(\frac{x}{q}\right)^{-\omega} [1 - H(q/x^\beta)]\, d\tilde{G}(x)]$$

Rearranging and taking limits as $\Delta \rightarrow 0$ we obtain

$$\frac{d}{dt} lnM_t(q) = \lim_{\Delta \rightarrow 0} \frac{M_{t+\Delta}(q) - M_t(q)}{\Delta M_t(q)} = -\delta_t \int_0^\infty \left(\frac{x}{q}\right)^{-\omega} [1 - H(q/x^\beta)] d\tilde{G}_t(x) \quad (3)$$

With this limit, we can derive an equation describing the frontier of knowledge. Since $\tilde{F}_t(q) = M_t(q)^m$, the change in the frontier of knowledge evolves as

$$\frac{d}{dt} ln\tilde{F}_t(q) = -m\delta_t \int_0^\infty \left(\frac{x}{q}\right)^{-\omega} [1 - H(q/x^\beta)] d\tilde{G}_t(x)$$

The proof of convergence to a Frechet distribution requires the following assumptions:

i. The distribution of original ideas, $H(z)$, has Pareto tail with exponent θ, such that $\lim_{z \rightarrow \infty} \frac{1 - H(z)}{z^{-\theta}} = 1$.

ii. $\beta \in [0, 1]$, and $\omega < \beta\theta$.

iii. The distribution of ideas from other producers $G_t(q)$ has a thin tail. That is, for all t and $k > 0$, $\lim_{x \rightarrow \infty} x^{\beta\theta}[1 - \tilde{G}(x|x < k)] = 0$.

Define $F(q) = \tilde{F}(m^{\frac{1}{\theta - \theta\beta + \omega}} q)$ and $G(z) = \tilde{G}(m^{\frac{1}{\theta - \theta\beta + \omega}} z)$ to normalize these distributions by the number of producers. Then, as $m \rightarrow \infty$, the frontier of knowledge evolves as[7]

$$\frac{dlnF_t(q)}{dt} = -\delta_t q^{-\theta + \omega} \int_0^\infty x^{\beta\theta - \omega} dG_t(x).$$

Note that a higher q will have 2 different and opposite effects: it will reduce the speed at which the frontier of knowledge evolves ($q^{-\theta}$), but it will also speed up the evolution of this frontier through q^ω; an effect that is absent if $\omega = 0$.

[6]This is because $M_{t+\Delta}(q) = M_t(q)[1 - \alpha_t Prob(q < zq'^\beta)]$. This law of motion is easy to understand thinking $M_{t+\Delta}(q)$ will be equal to $M_t(q)$ unless: there is match (probability α_t), there is adoption given the match according to **??**, and the hybrid idea is better than the original one($q < zx^\beta \rightarrow \frac{q}{x^\beta} < z \rightarrow 1 - H(q/x^\beta)$).

[7]The proof is provided in Appendix A.1.

Now, define $\lambda_t \equiv \int_{-\infty}^{t} \delta_s \int_0^\infty x^{\beta\theta-\omega} dG_s(x) ds$. Then, if $\lambda_t \to \infty$ as $t \to \infty$, the economy's frontier of knowledge can be described as a Frechet distribution, i.e.

$$F_t(\lambda_t^{\frac{1}{\theta-\omega}} q) \to e^{-q^{-\theta+\omega}};$$

the proof of this result is presented in Appendix A.2[8]. Further, the dynamics of the scale parameter λ — called the stock of knowledge — behaves according to

$$\dot{\lambda}_t = \delta_t \int_0^\infty x^{\beta\theta-\omega} dG(x). \tag{4}$$

3.2 International Trade

In this section I will briefly present how to extend the simple model presented before to introduce many asymmetric countries and international trade, following Bernard et al. (2003). There are n countries in the economy. Each country i has a labor supply, Li, stock of knowledge, λ_i,and iceberg trade costs, κ_{ij}. Consumers in i have identical preferences over a continuum of goods;

$$C_i = [\int_0^1 c_i(s)^{\frac{\varepsilon-1}{\varepsilon}} ds]^{\frac{\varepsilon}{\varepsilon-1}},$$

where $c_i(s)$ denotes the consumption of a representative household in i of good s and utility is given by $u(Ci)$.

We assume production is linear in labor,therefore, for a manager in i, the unit cost of selling a good to country j is $\frac{w_i\kappa_{ji}}{q}$. Firms engage in Bertrand competition, therefore, in equilibrium the price index in country i is given by[9]

$$P_i^{-\theta+\omega} \propto \sum_j \lambda_j (w_j\kappa_{ij})^{\theta-\omega}.$$

[8] Namely, it can be proved that the distribution of knowledge will be a Frechet either if the initial distribution F_0 is Frechet, or if we allow $t \to \infty$.

[9] This follows from Bernard et al. (2003), and this proof, along with the one for the trade shares is completely analogous to the one in Appendix B in Buera and Oberfield (2016) using equation 8.

The labor market clearing (with balanced trade) is $w_i L_i = \sum_j \pi_{ji} w_j L_j$, while trade shares are given by

$$\pi_{ij} = \frac{\lambda_j (w_j \kappa_{ij})^{\theta-\omega}}{\sum_k \lambda_k (w_k \kappa_{ik})^{\theta-\omega}}.$$

Diffusion Specification I will find the vector $\lambda = \lambda_1, ..., \lambda_n$ of stocks of knowledge under the assumption that insights are drawn uniformly from the distribution of producers *who sell* goods to the country. Empirical evidence of this being the relevant channel (as opposed to exporters learning) is provided in Appendix **??**. Denoting the set of goods s in country i such that the lowest cost seller is from country j as $S_{ij} \subset [0, 1]$, the source distribution can be written as $G_i^*(q) = \sum_k \int_{s \in S_{ik} | q_j < q} ds$. As proved in Appendix A.3, if $\omega < \beta\theta$ then the general form of Equation 4 becomes

$$G_i^S(q) = \frac{\sum_j \int_{s \in S_{ij} | q_{j1}(s) < q} ds}{\sum_j \int_{s \in S_{ij}} ds}$$

Then, specializing the evolution of the stock of knowledge to this specific source distribution, it evolves according to

$$\dot{\lambda}_{it} = \Gamma\left(1 - \beta + \frac{\omega}{\theta}\right) \delta_{it} \sum_j \pi_{ijt} \left(\frac{\lambda_{jt}}{\pi_{ijt}}\right)^{\beta - \frac{\omega}{\theta}} \tag{5}$$

where $\Gamma(\cdot)$ denotes the Gamma distribution. Thus, opening to trade gives wider access to the most productive sellers in the world (exporters), improving the stock of knowledge.

Moreover, the less a country trades with another country, the better the insights it will receive, and the more the levels of productivity for which the probability of adoption will be very low, given a match. Finally, note that because of the learning externality, the condition that maximizes the stock of knowledge in each country i, $\frac{\pi_{ij}}{\pi_{ik}} = \frac{\lambda_j}{\lambda_k}$, which depends implicitly on ω and δ, is different from

$$\frac{\pi_{ij}}{\pi_{ik}} = \frac{\lambda_j (w_j \kappa_{ij})^{-\theta+\omega}}{\lambda_k (w_k \kappa_{ik})^{-\theta+\omega}},$$

which describes the relationship between the equilibrium expenditure shares, depending explicitly on ω.

4 Quantitative Exploration

I next present an extended quantitative model, in which I impose less restrictive assumptions than in Section 3. I then use model-generated data to validate the model using my reduced-form empirical estimates, and concluding that different learning probabilities are necessary to match the empirical relations I documented. I further compare the TFP dynamics predicted by the model with the data.

4.1 Quantitative Framework

Technology is CRS, such that output of country i with productivity q depends on an aggregate of intermediate inputs d_i and equipped labor l_i, where aggregate equipped labor results from combining aggregate units of capital and efficient units of labor using a Cobb Douglass technology[10]. Each good is denoted by m, and how much of it is used in the production of an intermediate input depends on the function $D_i(m)$. It follows that $\int d_i(m)dm = [\int D_i(m)^{\frac{\rho-1}{\rho}} dm]^{\frac{\rho}{\rho-1}}$, and a good m is produced according to $y_i(m) = q\frac{d_i(m)^{\eta} l_i(m)^{1-\eta}}{\eta^{\eta}(1-\eta)^{1-\eta}}$. Finally, the proportion of non-traded goods (goods with an infinite transportationcost) will be given by ψ. The results for the price index and evolution stock of knowledge for this case are derived in Appendix B.

4.2 Parametrization

In the model derived in Section 3, on a balanced growth path the growth rate of productivity is

[10]The paths of both the aggregate units of capital and the human capital will be taken from the data.

$$\frac{1}{(\theta-\omega)}\frac{\dot{\lambda}}{\lambda}=\frac{\gamma}{(\theta-\omega)(1-\beta+\frac{\omega}{\theta})}, \tag{6}$$

and the growth rate of the stock of knowledge is given by

$$\frac{\gamma}{(1-\beta+\frac{\omega}{\theta})}; \tag{7}$$

where γ represents the growth rate of the arrival of matches. First, I use the mean growth rate of population in the US from 1962 to 2007 to calibrate γ. This is done under the assumption that the US - a developed country - is a benchmark for a country on a balanced growth path. Second, the exponent of distribution of own-ideas H (Pareto) is calibrated to match the value of 4 in Simonovska and Waugh (2014), i.e. $\theta = 4$.

Then, calibrating the mean growth rate of the stock of knowledge to the mean growth rate of the stock of US patents from 1962 to 2007 (2.5%), I can identify $\beta - \frac{\omega}{\theta}$ from (7), It follows that, given θ, and assuming that the growth rate of TFP on the balanced growth path equals the mean groth rate for US (1962 and 2007; 0.8%), it is possible to identify ω from (6) and back-out β. Finally, the trade costs κ_{ij} will be chosen to match bilateral trade flows. The values for these and the rest of parameters are summarized in Table 2.

TABLE 2: PARAMETER VALUES

Parameter	Value	Target
Share of non-traded goods (ψ)	0.34	Fraction of agr, min, man in gross output
Share of intermediate goods (η)	0.48	Intermediate share in gross production
Elasticity of substitution (ε)	1	Waugh (2014) - Robust
Exponent of $H(z)$ (θ)	4	Simonovska and Waugh (2014)
Concavity in learning (β)	0.819	
Absorption (ω)	0.875	

Arrival rates To assign values to the vector of arrival rates $\delta_t = (\delta_{1t}, ..., \delta_{nt})$ there are 2 steps. First, given the evolution of trade flows compute - in each year - the stocks of knowledge λ_{it} needed to match each country's own trade share. Where $\lambda_{it} \propto$

$f(\kappa_{ijt}, \pi_{ijt})(\frac{w_{it}}{P_{it}})^{\eta(\theta-(\omega/\beta))}$. Second, given the evolution of trade flows and stocks of knowledge as well as values for β, ω, and the growth rate of the arrival of ideas γ, back out sequences of δ_t using the law of motion of λ_{it}.[11]. The bilateral data for these calibrations is taken from Feenstra et al. (2005) and on real GDP and the price index from the Penn World tables v.8.

4.3 Model Validation

In this section, I run the regressions presented in Section 2.3 on model-generated data. First, I show the results from the benchmark Regression 1, and compare the results of the model with heterogeneous learning with a model with homogeneous learning like Buera and Oberfield (2016). Then I run the same non-parametric regression on real data and model-generated data. Finally, I construct an Autor et al. (2013) instrument using model-generated data and compare it with the results obtained while using the China shock and the instrument on real data.

Benchmark Regression Looking at the results from Regression 1 shown in Table 3, we can see that the model replicates my empirical findings both qualitatively and quantitatively. Moreover, when running the model with homogeneous learning ($\omega = 0$, coinciding with the original forces in Buera and Oberfield (2016)), the relations between the variables are counterfactual.

TABLE 3: *Dependent variable: percentage change in productivity*

(Log) Variables	Data		Model		$\omega = 0$	
Ratio	-0.56^{***}	-1.46^{***}	-0.51^{***}	-1.01^{***}	0.50^{***}	1.00^{***}
	(0.014)	(0.037)	(0.038)	(0.003)	(0.003)	(0.002)
Trading partner's z^u_{jt}		0.94^{***}		0.99^{***}		1.09^{***}
		(0.038)		(0.003)		(0.004
FE/Robust SE	✓	✓	✓	✓	✓	✓

[11]The explicit forms of these equations are summarized in Appendix A.2

As shown, to answer the two main questions I am interested in: how does learning through trade depend on the trade partner's productivity level, and how does this learning depend on the productivity gap between the parties involved, it is essential to include heterogeneous learning into the model. Moreover, incorporating this heterogeneity through different learning probabilities

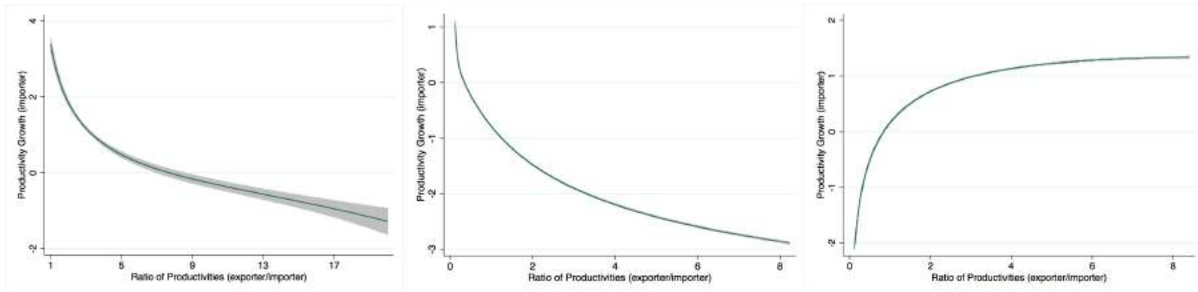
Non-Parametric Regression, China Shock and IV The model-generated data also captures the patterns found in the data through a non-parametric regression. Moreover, the non-parametric regression's results in panels (a)-(c) of Figure 3 also highlight how including differences in learning abilities in model is crucial, as results are counterfactual in the case in which $\omega = 0$ and learning is independent of one's current state of knowledge. Panels (d)-(f) show how using the China shock and constructing instruments in the same fashion as in Section 2.3 delivers qualitative results consistent with the my previous findings: the model can closely replicate the patterns found in the data, and the heterogeneous learning abilities are key in delivering this result; a model without this heterogeneity ($\omega = 0$) does not capture the relations in the data also when using using the instrument and China shock.

4.4 Analysis of TFP Dynamics

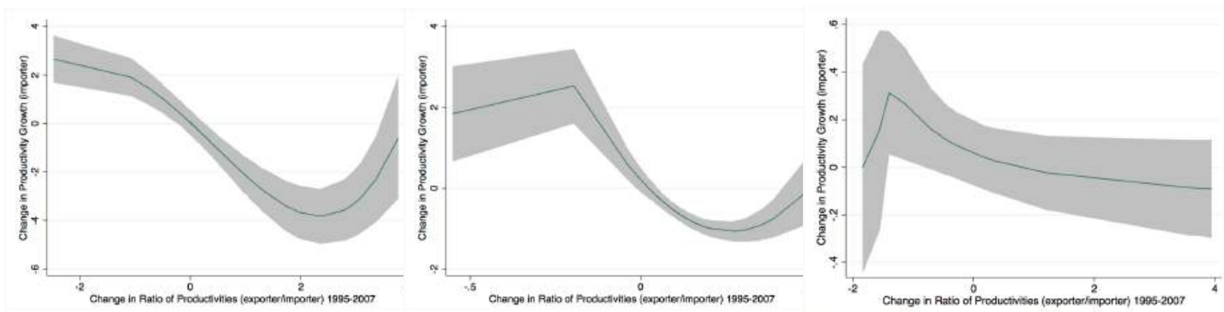
One of the main motivations of this kind of models, is to understand the evolution of TFP across countries, in particular, how can the diffusion of knowledge help explain TFP dynamics and growth miracles. This section will explore the role of heterogeneous learning in explaining these dynamics, in comparison with models without absorptive capacity constraints. In particular, better understand the role of learning, I will quantify and decompose the contribution of learning, isolating the contribution of the (exogenous) arrival of ideas (δ_t).

Figure 4 shows the evolution of TFP both in the model with heterogeneous learning,

Figure 3: Productivity Growth vs Ratio of Productivities



(a)No Instrument (Data) (b)No Instrument(Model) (c)No Instrument($\omega = 0$)



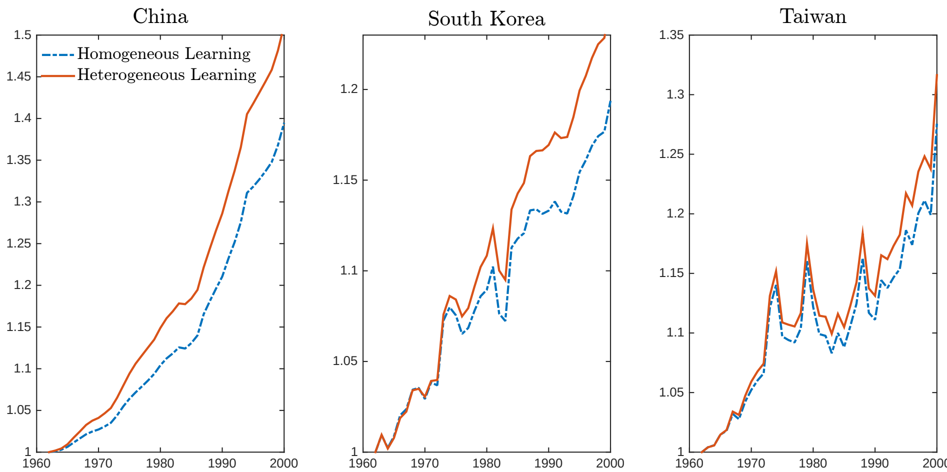
(d)Instrument(Data) (e)Instrument(Model) (f)Instrument($\omega = 0$)

and while turning-off learning heterogeneity ($\omega = 0$). Namely, it plots $\ln \frac{TFP_i(\delta_0,\kappa_t)}{TFP_i(\delta_0,\kappa_0)}$.

This figure assumes that the three countries were on their balanced growth path in 1962. In all cases, the model with heterogeneous learning delivers a higher TFP, especially for the last decades when these countries grew the fastest. For China for instance, the TFP assuming different learning abilities is up to 20 percent larger than ignoring them.

To further understand the effect of learning constraints on TFP growth, I decompose the change in TFP to isolate the effect of changes in the exogenous arrival of ideas, δ_t. First, note that the contribution of an increase in the exogenous arrival of ideas is larger if we assume learning is independent of the current state of knowledge: if anyone can learn from a good producer, more arrivals of good producers will lead to more growth than in a case where only the a subset of productive people are likely to learn. This also means that the difference between how much each model explains from trade will be larger once we net-out the effect of exogenous arrivals of ideas. Figure 5 shows the contribution of

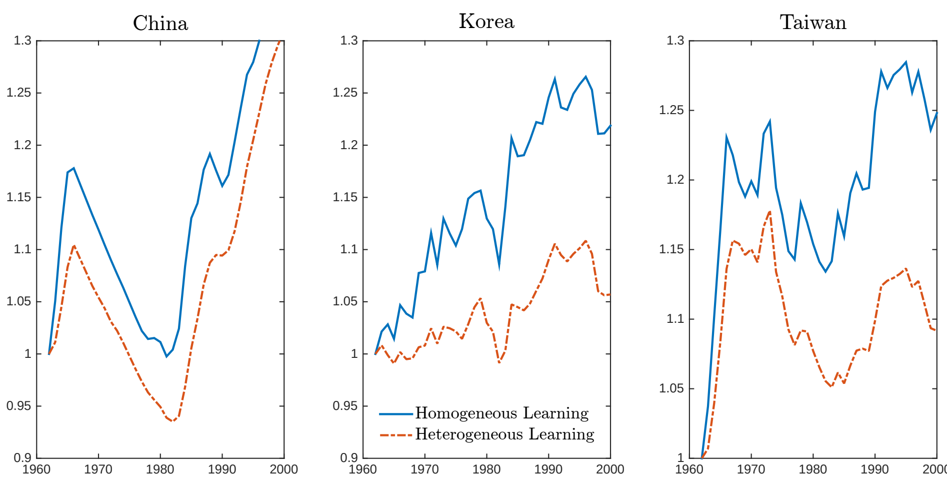
FIGURE 4: EVOLUTION OF TFP WITH DIFFERENT LEARNING ABILITIES



Notes: Figure 4 plots the changes in TFP generated for various growth miracles, detrending TFP by the average growth rate of TFP in the United States. All the values are calibrated as detailed in section 4.2.

changes in the exogenous arrival rate of ideas with and without different learning abilities, exhibiting the expected pattern. Finally, Figure 6 plots the contribution of trade to TFP with both modes of learning, and the evolution of TFP in the data; in all casses netting for the contribution of the exogenous arrival of ideas.[12]

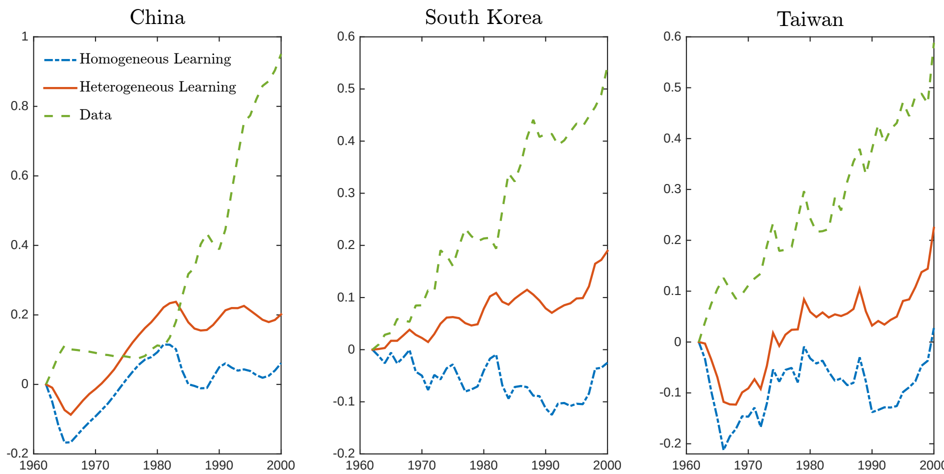
FIGURE 5: CONTRIBUTION OF THE EXOGENOUS ARRIVAL OF IDEAS (δ_t) TO Δ IN TFP



Notes: Figure 6 plots the changes in TFP attributed to increases in the exogenous arrival of ideas δ_t various growth miracles, detrending TFP by the average growth rate of TFP in the United States.

[12]That is, subtracting $\frac{TFP_i(\delta_t,\kappa_0)}{TFP_i(\delta_0,\kappa_0)}$.

Figure 6: Evolution of TFP netted of exogenous changes in arrival rates



Notes: Figure 6 plots the changes in TFP generated for various growth miracles, detrending TFP by the average growth rate of TFP in the United States *minus* changes in TFP due to changes in the exogenous arrival rate of ideas δ_t. All the values are calibrated as detailed in section 4.2.

Figures 5 show how ignoring that learning ability depends on the current state of knowledge can overestimate the role of the exogenous arrival of ideas. Further, Figure 6 shows how it is possible to deliver a relatively high contribution of trade to TFP dynamics for growth miracles by introducing learning heterogeneity. With learning heterogeneity, no country can "buy a lottery ticket" and potentially increase the TFP of its producers overnight by importing from a very productive partner. In th model, this leads to optimal trade shares that "divert" trade towards countries with higher – but relatively close – TFP levels, from which it is more likely to learn given a meeting, resulting in a larger contribution of trade to TFP growth for growth miracles.

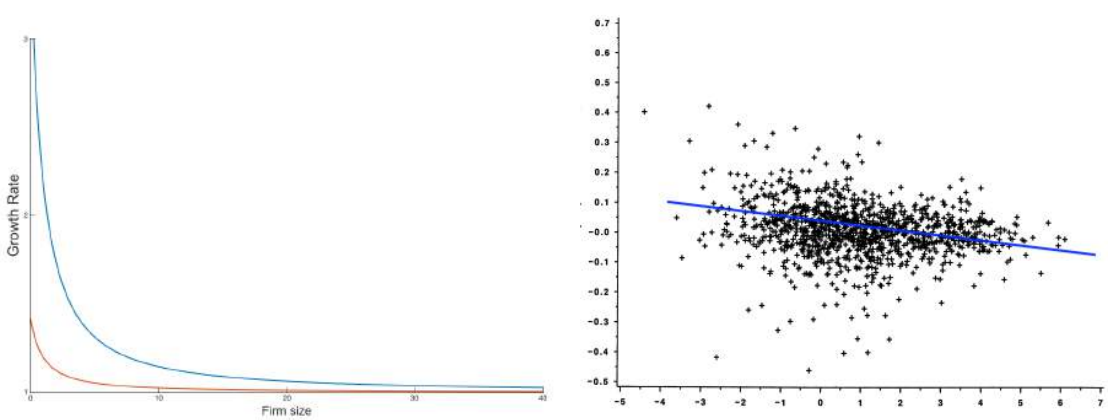
4.5 Further Implications

In Section 4.3, I discussed how learning heterogeneity allows the model to match an empirical dependence of productivity growth on productivity gaps across trading partners. In the previous Section, I showed how, for growth miracles, the contribution of trade to TFP can increase in around 20% due to the introduction of different learning abilities.

In this section, I will discuss other implications of the model. Namely, (1) the model fits the firm size distribution, and follows Gibrat's law; (2) when opening to trade, the most productive sellers drive most of the productivity growth, as it happens empirically (Steinwender, 2015); and (3) policy recommendations drastically change: developing countries should direct more trade to mid-productive countries instead of to the most productive ones.

Consistency with Gibrat's Law for Firms According to Gibrat's Law for Firms, firm growth is independent of firm size. Adding heterogeneous learning allows the model to closely reproduce this fact. The unconstrained model where learning does not depend on the ratio of productivities delivers a relationship between a firm's growth rate and its size that is largely inconsistent with Gibrat's law, as shown in Figure **??**. That is because these firms always have the potential to grow overnight, and in expectation this growth is a lot larger that the one of an already productive firm (who expects most of its matches to be useless in terms of learning). The constrained model provides a much closer match to Gibrat's law. Moreover, in the data, Gibrat's law does not hold perfectly for small firms, something that is true in the model as well.

FIGURE 7: GROWTH RATE VS FIRM SIZE



Homogenous/Heterogenous learning   Data (Hall,1987)

Learning and the Distribution of Productivities Steinwender (2015) finds robust evidence that access to export markets leads to productivity increases within firms, but this only happens for firms that were already highly productive. Since exporters tend to be more productive, introducing a notion of absorptive capacity (namely, preventing firms from learning other firms' ideas when there is a large gap in their productivities), ensures that only ex-ante more productive firms learn from foreigners. This is consistent with Steinwender (2015), who find evidence that there are productivity increases after expanding export markets, but only for firms that were ex-ante the more productive ones.

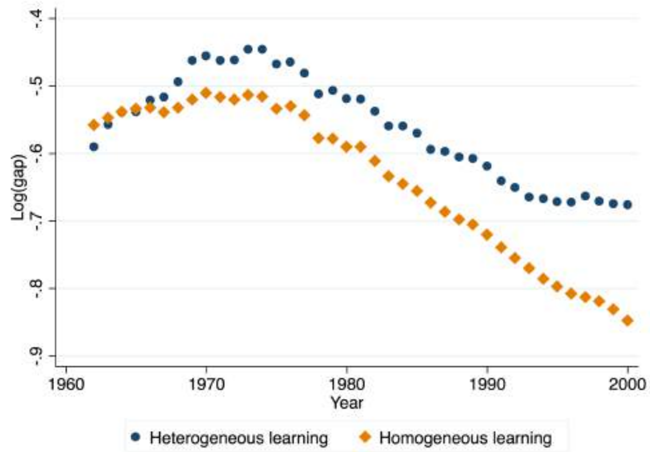
There is also considerable evidence that the time devoted to idea exchanges is greater when the agent has higher ability, as documented by Allen et al. (2010) who found that scientists who worked at more productive firms communicated with outsiders more. Current models imply the opposite: either (1) producers with low productivity benefit the most from diffusion[13](Alvarez et al. (2013), Buera and Oberfield (2016)), or (2) the entire distribution improves due to selection (Perla et al. (2015), and Sampson (2014)). This counterfactual result highlights why it is important to introduce a notion of absorptive capacity in the model. There must be some constraint on how much an agent can learn at once, otherwise the ones who proportionately gain the most will be the ex-ante low-ability agents. In the extended model, once an exporter (highly productive, given the Bertrand structure) enters a country, the most productive sellers are the ones that are more likely to learn, given the probability of learning is decreasing in the ratio of productivities.

Welfare Analysis In this subsection, I analyze the implications of the model on welfare. Figure 8 shows how the "welfare gap", which is computed as the difference in average welfare between the welfare fo the top 25% and the bottom 25% of countries in the sample. I find that convergence, given the data in which the model was estimated, happens both

[13]This is because if a low-ability producer matches with a high-ability one, his gains can be very large, while if a high-ability producer matches his gains are moderate at best, given he is already productive.

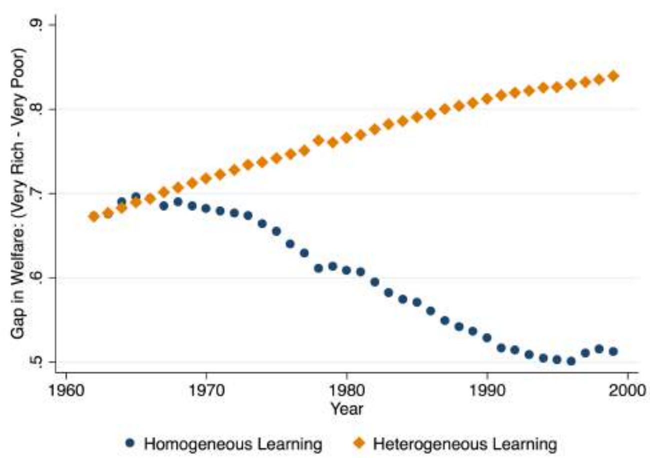
with homogeneous nd heterogeneous learning, but that heterogeneity in learning slows down this convergence by over 20%.

FIGURE 8: WELFARE GAP BETWEEN RICH AND POOR COUNTRIES



Also, I explore the theoretical possibility of divergence in welfare doing a counter-factual exercise, in which I only keep the top and bottom 15% of richest and poorest countries in 1962. This shows how the modle with heterogeneous learning accommodates the possibility of divergence between developed and developing country. Turning down this heterogeneity would make it impossible to allow for this scenario.

FIGURE 9: WELFARE GAP BETWEEN RICH AND POOR COUNTRIES



Other Policy Implications As a result of the optimal trade shares exposed in Section 3, policy implications are different in both models: for developing countries, the model with heterogeneous learning implies inducing larger trade shares with countries with higher –but close– productivity, as opposed with trading more with very productive ones (as in a model where learning is independent of current knowledge). From the point of view of a developing country, this would mean trading more with mid-developed countries instead of very developed ones, in order to maximize the gains from trade coming from knowledge diffusion and technological transfers.

5 Concluding Remarks

In this paper I developed a tractable theory of international diffusion of ideas. Crucially, I incorporated the idea of absorptive capacity, which allows the model to accommodate the fact that the productivity gap between two countries may be a barrier for an economy to have gains from the diffusion of ideas, even if trade is taking place and local firms match with foreign sellers. This idea is introduced in a way that is simple enough to be tractable, but that is able to capture well-documented empirical facts.

The model provides a theory to explain the fact that export markets lead to productivity increases within firms, but only for firms that were already highly productive. It also accounts for heterogeneous diffusion of technologies after conditioning on countries having the same trading partners[14].

The analysis shows how for underdeveloped countries, it is not always better to have very technologically-advanced trading partners in order to improve their state of knowledge. My model includes cases in which there are more gains from diffusion if trading partners have a relatively close state of development. This analysis suggests that previous

[14]Remembering that in an Eaton-Kortum framework, it is the extensive margin the one that is relevant when it comes to trading partners.

models may be too enthusiastic when quantifying how much a low productivity economy may learn from foreign sellers with very high productivities once they start trading, leading to different potential policies to increase productivity.

Qualitatively, previous models were counterfactual: evidence shows high-ability managers and scientists have larger gains from exchanging ideas than low-ability ones. My model can account for this pattern. Aside from exploring this new mechanism theoretically, the quantitative exploration shows that including this feature has important implications. Namely, this framework can reproduce the dynamics of TFP up to 20 per cent more closely than previous models, and is particularly well-suited to examine North-South trade of ideas and the experience of growth miracles, where trading partners have very different levels of development.

As in other models in this literature, I abstract from FDI and purposeful imitation as sources of diffusion of ideas, and omit variation across sectors. A next step could be to focus on a particular industry or a specific country and analyze its learning process.

References

Daron Acemoglu and Fabrizio Zilbotti. Productivity differences. Working Paper 6879, National Bureau of Economic Research, January 1999. URL http://www.nber.org/papers/w6879.

Fernando E. Alvarez, Francisco J. Buera, and Jr. Robert E. Lucas. Idea flows, economic growth, and trade. Working Paper 19667, National Bureau of Economic Research, November 2013. URL http://www.nber.org/papers/w19667.

Kenneth Arrow. *Economic Welfare and the Allocation of Resources for Invention*, pages 609–626. Princeton University Press, 1962. URL http://www.nber.org/chapters/c2144.

Andrew Atkeson and Ariel T. Burstein. Aggregate implications of innovation policy. Working Paper 17493, National Bureau of Economic Research, October 2011. URL http://www.nber.org/papers/w17493.

Anthony B Atkinson and Joseph E Stiglitz. A New View of Technological Change. *Economic Journal*, 79(315):573–78, September 1969. URL https://ideas.repec.org/a/ecj/econjl/v79y1969i315p573-78.html.

David H. Autor, David Dorn, and Gordon H. Hanson. The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–68, October 2013. doi: 10.1257/aer.103.6.2121. URL http://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121.

Andrew B. Bernard, Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. Plants and productivity in international trade. *American Economic Review*, 93(4):1268–1290, September 2003. doi: 10.1257/000282803769206296. URL http://www.aeaweb.org/articles?id=10.1257/000282803769206296.

Francisco J. Buera and Ezra Oberfield. The global diffusion of ideas. Working Paper 21844, National Bureau of Economic Research, January 2016. URL http://www.nber.org/papers/w21844.

Wesley M Cohen and Daniel A Levinthal. Innovation and learning: The two faces of r&d. *Economic Journal*, 99(397):569–96, 1989. URL http://EconPapers.repec.org/RePEc:ecj:econjl:v:99:y:1989:i:397:p:569-96.

Jonathan Eaton and Samuel Kortum. International technology diffusion: Theory and measurement. *International Economic Review*, 40(3):537–570, 1999. ISSN 00206598, 14682354. URL http://www.jstor.org/stable/2648766.

Robert E. Evenson and Larry E. Westphal. Technological change and technology strategy. In Hollis Chenery and T.N. Srinivasan, editors, *Handbook of Development Economics*, volume 3, Part 1, chapter 37, pages 2209–2299. Elsevier, 1 edition, 1995. URL http://EconPapers.repec.org/RePEc:eee:devchp:3-37.

Ronald Findlay. Relative backwardness, direct foreign investment, and the transfer of technology: A simple dynamic model. *The Quarterly Journal of Economics*, 92(1): 1–16, 1978. URL http://EconPapers.repec.org/RePEc:oup:qjecon:v:92:y:1978:i:1:p:1-16.

Daniel Garcia-Macia, Chang-Tai Hsieh, and Peter J Klenow. *How Destructive is Innovation?* January, 2015.

Amy Glass and Kamal Saggi. Foreign direct investment and the nature of r&d. Iser discussion paper, Institute of Social and Economic Research, Osaka University, 1996. URL http://EconPapers.repec.org/RePEc:dpr:wpaper:0420.

Gene M. Grossman and Elhanan Helpman. Trade, knowledge spillovers, and growth.

Working Paper 3485, National Bureau of Economic Research, October 1990. URL http://www.nber.org/papers/w3485.

Kyle Herkenhoff, Lise Jeremy, Guido Menzio, and Gordon Phillips. Worker mobility and the diffusion of knowledge. Technical report, 2018.

Thomas J. Holmes, Ellen R. McGrattan, and Edward C. Prescott. Quid pro quo: Technology capital transfers for market access in china. Working Paper 19249, National Bureau of Economic Research, July 2013. URL http://www.nber.org/papers/w19249.

Gregor Jarosh, Ezra Oberfield, , and Esteban Rossi-Hansberg. Learning from coworkers. Technical report, 2018.

Boyan Jovanovic and Rafael Rob. The growth and diffusion of knowledge. *The Review of Economic Studies*, 56(4):569–582, 1989. ISSN 00346527, 1467937X. URL http://www.jstor.org/stable/2297501.

Samuel Kortum. Exploring Innovation with Firm Level Data. Economics Program Working Papers 08-11, The Conference Board, Economics Program, December 2008. URL https://ideas.repec.org/p/cnf/wpaper/0811.html.

Robert Lucas and Benjamin Moll. Knowledge growth and the allocation of time. Working Paper 17495, National Bureau of Economic Research, October 2011. URL http://www.nber.org/papers/w17495.

Erzo G. J. Luttmer. Technology diffusion and growth. Working Papers 672, Federal Reserve Bank of Minneapolis, 2009. URL https://ideas.repec.org/p/fip/fedmwp/672.html.

Erzo G. J. Luttmer. Eventually, noise and imitation implies balanced growth. Working Papers 699, Federal Reserve Bank of Minneapolis, 2012. URL https://ideas.repec.org/p/fip/fedmwp/699.html.

Jesse Perla and Christopher Tonetti. Equilibrium Imitation and Growth. *Journal of Political Economy*, 122(1):52 – 76, 2014. URL https://ideas.repec.org/a/ucp/jpolec/doi10.1086-674362.html.

Jesse Perla, Christopher Tonetti, and Jess Benhabib. The Growth Dynamics of Innovation, Diffusion, and the Technology Frontier. 2014 Meeting Papers 818, Society for Economic Dynamics, 2014. URL https://ideas.repec.org/p/red/sed014/818.html.

Jesse Perla, Christopher Tonetti, and Michael E. Waugh. Equilibrium technology diffusion, trade, and growth. Working Paper 20881, National Bureau of Economic Research, January 2015. URL http://www.nber.org/papers/w20881.

Natalia Ramondo and Andres Rodriguez-Clare. Trade, Multinational Production, and the Gains from Openness. *Journal of Political Economy*, 121(2):273 – 322, 2013. URL https://ideas.repec.org/a/ucp/jpolec/doi10.1086-670136.html.

Paul M. Romer. Endogenous technological change. *Journal of Political Economy*, 98(5): S71–S102, 1990. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/2937632.

Thomas Sampson. Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth. CEP Discussion Papers dp1288, Centre for Economic Performance, LSE, August 2014. URL https://ideas.repec.org/p/cep/cepdps/dp1288.html.

Ina Simonovska and Michael E. Waugh. The elasticity of trade: Estimates and evidence. *Journal of International Economics*, 92(1):34–50, 2014. doi: 10.1016/j.jinteco.2013.07. URL https://ideas.repec.org/a/eee/inecon/v92y2014i1p34-50.html.

Claudia Steinwender. The roles of import competition and export opportunities for technical change. Cep discussion papers, Centre for Economic Performance, LSE, 2015. URL http://EconPapers.repec.org/RePEc:cep:cepdps:dp1334.

Frances Stewart. Technology and underdevelopment. *Development Policy Review*, A10 (1):92–105, 1977. ISSN 1467-7679.

Timmer, de Vries, and de Vries. Patterns of structural change in developing countries. *Routledge Handbook of Industry and Development.*, pages 65–83, 2015a.

Timmer, Dietzenbacher, Los, Stehrer, and Vries. An illustrated user guide to the world input-output database: the case of global automotive production. *Review of International Economics*, 23(3):575–605, 2015b. doi: 10.1111/roie.12178. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/roie.12178.

Fabrizio Zilibotti, Michael Konig, and Jan Lorenz. Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, 2010. URL https://ideas.repec.org/a/the/publsh/1437.html.

Appendix

A Proofs

A.1 Proof of Convergence of the Frontier of Knowledge

We derived previously how, since $\tilde{F}_t(q) = (M_t(q)^m)$, the change in the frontier of knowledge evolves as

$$\frac{d}{dt} ln\tilde{F}_t(q) = -m\delta_t \int_0^\infty \left(\frac{x}{q}\right)^{-\omega} \left[1 - H\left(\frac{q}{x^\beta}\right)\right] d\tilde{G}_t(x).$$

Define $F_t(q) = \tilde{F}(m^{\frac{1}{\theta-\beta\theta+\omega}} q)$ and evaluate the equation above in $m^{\frac{1}{\theta-\beta\theta+\omega}} q$. We obtain

$$\frac{d}{dt} ln\tilde{F}_t(m^{\frac{1}{\theta-\beta\theta+\omega}} q) = -m\delta_t \int_0^\infty \left(\frac{x}{m^{\frac{1}{\theta-\beta\theta+\omega}} q}\right)^{\omega} [1 - H(m^{\frac{1}{\theta-\beta\theta+\omega}} q/x^\beta)] d\tilde{G}_t(x)$$

Further, using the change of variables $w = m^{\frac{-1}{(\theta-\beta\theta+\omega)}} x$, and defining $G_t(x) = \tilde{G}_t(m^{\frac{1}{(\theta-\beta\theta+\omega}} x)$, we obtain

$$\frac{d}{dt} lnF_t(q) = -m\delta_t \int_0^\infty \left(\frac{x}{q}\right)^{-\omega} [1 - H(q/x)^\beta)] dG_t(x),$$

and we can rewrite the expression as

$$\begin{aligned}
\frac{d}{dt} lnF_t(q) &= -\delta_t q^{-\theta} \int_0^\infty \left(\frac{x}{q}\right)^{-\omega} \frac{[1 - H(m^{\frac{1}{\theta}} q/x^\beta)]}{(m^{\frac{1}{\theta}} q/x^\beta)^{-\theta}} x^{\beta\theta} dG_t(x) \\
&= -\delta_t q^{-\theta+\omega} \int_0^\infty \frac{[1 - H(m^{\frac{1}{\theta}} q/x^\beta)]}{(m^{\frac{1}{\theta}} q/x^\beta)^{-\theta}} x^{\beta\theta-\omega} dG_t(x)
\end{aligned}$$

From assumption (i), we can take the limit as $m \to \infty$ inside these integrals, and by (iii), the integrals $\int_0^\infty x^{\beta\theta} dG_t(x)$ and $\int_0^\infty x^{\beta\theta-\omega} dG_t(x)$ are finite. Therefore, we can take the limit as $m \to \infty$ inside the integral using the dominated convergence theorem to get

$$\frac{dlnF_t(q)}{dt} = \delta_t q^{-\theta+\omega} \int_0^\infty x^{\beta\theta-\omega} dG_t(x).$$

A.2 Proof of Frechet Limit

Solving

$$\frac{dlnF_t(q)}{dt} = \delta_t q^{-\theta+\omega} \int_0^\infty x^{\beta\theta-\omega} dG_t(x).$$

as a differential equation, we obtain $F_t(q) = F_0(q)e^{(-\lambda_t-\lambda_0)q^{-\theta+\omega}}$.

Evaluating this at $\lambda^{1/(\theta+\omega)}q$ gives us that

$$F\left(\lambda^{1/(\theta+\omega)}q\right) = F_0(\lambda^{1/(\theta+\omega)})e^{(-\lambda_t-\lambda_0)\lambda^{-1}q^{-\theta+\omega}}.$$

Asymptotically, this means that

$$\lim_{t\to\infty} F\left(\lambda^{1/(\theta+\omega)}q\right) = e^{-q^{-\theta+\omega}}.$$

A.3 Proof for Law of Motion of the Stock of Knowledge, λ_t

Assuming learning from sellers, and denoting the set of goods s in country i such that the lowest cost seller is from country j as $S_{ij} \subset [0,1]$, the source distribution can be written as $G_i^*(q) = \sum_k \int_{s\in S_{ik}|q_j<q} ds$, and the general form of Equation 4 becomes

$$\begin{aligned}\dot{\lambda_{it}} &= \delta_{it}\int_0^\infty x^{\beta\theta-\omega} dG_{it}^*(x) \\ &= \delta_{it}\Gamma\left(1-\beta+\frac{\omega}{\theta}\right)\sum_j \pi_{ij}\left(\frac{\lambda_{jt}}{\pi_{ij}}\right)^{\beta-\frac{\omega}{\theta}}\end{aligned}$$

where $\Gamma(\,\cdot\,)$ denotes the Gamma distribution.

Proof:

For τ_1 such that $0 \leq \tau_1 < 1$, Buera and Oberfield (2016) show that

$$\int_{s\in S_{ij}} q_{j1}(s)^{\tau_1\theta} p_i^{-\tau_2\theta} = B(\tau_1,\tau_2)\left[\sum_k \lambda_k(w_i\kappa_{ik})^{-\theta}\right]^{\tau_2} \pi_{ij}\left(\frac{\lambda_j}{\pi_{ij}}\right)^{\tau_1} \tag{8}$$

where $B(\tau_1,\tau_2) = \left[1-\frac{\tau_2}{1-\tau_1}+\frac{\tau_2}{1-\tau_1}\left(\frac{\epsilon}{1-\epsilon}\right)^{-\theta(1-\tau_1)}\right]\Gamma(1-\tau_1-\tau_2)$. Using this result, as $\omega < \beta\theta$, we obtain that

$$\int_0^\infty x^{\beta\theta-\omega} dG_i(x) = B\left(\beta - \frac{\omega}{\theta}, 0\right) \sum_j \pi_{ij} \left(\frac{\lambda_j}{\pi_{ij}}\right)^{\beta - \frac{\omega}{\theta}}$$
$$= \Gamma\left(1 - \beta + \frac{\omega}{\theta}\right) \sum_j \pi_{ij} \left(\frac{\lambda_j}{\pi_{ij}}\right)^{\beta - \frac{\omega}{\theta}}.$$

B Extended Model for the Quantitative Analysis

This section derives the price index, expenditure shares, and the law of motion of λ_i for a version of the model that includes intermediate inputs, nontradables and human capital. First, with an elasticity of substitution given by ε, the expression for the price index is

$$p_i^{1-\varepsilon} = \left[\psi(\chi_i\lambda_i)^{\frac{\varepsilon-1}{\theta-\omega}} + (1-\psi)\left(\sum_{j=1}^n \frac{\chi_j\lambda_j}{\kappa_{ij}^{\theta-\omega}}\right)^{\frac{\varepsilon-1}{\theta-\omega}}\right] C,$$

where $\chi_i = \frac{1}{(p_i^{1-\psi} w_i^{\psi})^{\theta-\omega}}$ and C is a constant term; while country i's expendure in non-tradable and tradable goods, respectively, are

$$\pi_i^{NT} = \frac{\psi(\chi_i\lambda_i)^{\frac{\varepsilon-1}{\theta-\omega}}}{\psi(\chi_i\lambda_i)^{\frac{\varepsilon-1}{\theta-\omega}} + (1-\psi)\left(\sum_{j=1}^n \frac{\chi_j\lambda_j}{\kappa_{ij}^{\theta-\omega}}\right)^{\frac{\varepsilon-1}{\theta-\omega}}}, \text{ and}$$

$$\pi_i^{T} = \frac{(1-\psi)\left(\sum_{j=1}^n \frac{\chi_j\lambda_j}{\kappa_{ij}^{\theta-\omega}}\right)^{\frac{\varepsilon-1}{\theta-\omega}}}{\psi(\chi_i\lambda_i)^{\frac{\varepsilon-1}{\theta-\omega}} + (1-\psi)\left(\sum_{j=1}^n \frac{\chi_j\lambda_j}{\kappa_{ij}^{\theta-\omega}}\right)^{\frac{\varepsilon-1}{\theta-\omega}}}.$$

As producers learn from sellers, the dynamics of country i's stock of knowledge depend on two sources improvements: learning coming from locals who produce non-tradables, and learning coming from sellers of tradables (who may or may not be foreigners). These two forces are captured as the two components of the following sum:

$$\dot{\lambda}_i = \left[\psi\lambda_i^{\beta - \frac{\omega}{\theta-\omega}} + (1-\psi)\sum_j \pi_{ij}^T \left(\frac{\lambda_j}{\pi_{ij}^T}\right)^{\beta - \frac{\omega}{\theta-\omega}}\right] C',$$

where C' is a constant term, and $\pi_{ij}^T = \frac{\chi_j \lambda_j / \kappa_{ij}^{\theta-\omega}}{\sum_s \chi_s \lambda_s / \kappa_{is}^{\theta-\omega}}$ is i's expenditure share of *tradables* coming from j.

C Model's Counterpart of Regression 1

Given the probability of learning in 3, an expectation taken using the source distribution $G_t(x)$, we can express average growth as

$$\frac{q_{t+1} - q_t}{q_t} = E_x \left[\delta \left(\frac{x}{q_t} \right)^{-\omega} z \left(\frac{x}{q_t} \right) x^{\beta-1} \mathbb{1}(z x^{\beta} > q_t) + \left(1 - \delta \left(\frac{x}{q_t} \right)^{-\omega} \right) \right].$$

This can be rewritten as

$$\frac{q_{t+1} - q_t}{q_t} = E_x \left[\delta \left(\frac{x}{q_t} \right)^{-\omega} \left(H \left(1 - \frac{q_t}{x} x^{1-\beta} \right) \left(\frac{x}{q_t} \right) x^{\beta-1} - 1 \right) + 1 \right].$$

$$\frac{q_{t+1} - q_t}{q_t} = \delta \rho_{\left(\frac{x}{q_t}, t\right)} \rho_{\left(\frac{x}{q_t}, x, t\right)}$$

where $\rho_{\left(\frac{x}{q_t}, t\right)}$ and $\rho_{\left(\frac{x}{q_t}, x, t\right)}$ capture, respectively, the effects of the ratio's importance and the interaction between the ratio and the partner's productivity. A certain equivalent of this equation would lead to a regression as in Equation (1), without including the trade shares(as explained, trade shares were included to explore their role in the learning process, but do not play a role in the model), which turned out to be insignificant empirically.

D Mapping of Industries from the WIOD into 10 Aggregate Sectors

As explained in Section 2.2, two of the sources I use have different levels of aggregation. Namely, the WIOD traces the flow of goods and services across 35 industries, while the data used to obtain the productivities per sector and per country comes from the GGDC database and world KLEMS data, which divide economic activity into 10 sectors. These sectors are: agriculture, hunting, forestry and fishing; mining and quarrying; manufacturing; electricity, gas and water supply; construction; wholesale and retail trade, hotels and restaurants; transport, storage, and communication; finance, insurance, real estate and business services; government services; community, social and personal services. The mapping was designed according to the ISIC codes presented by Timmer et al. (2015a), to divide the categories in the WIOD into 10 sectors that corresponded with those in the GGCD as follows:

- **Sector 1:** Agriculture, Hunting, Forestry and Fishing
- **Sector 2:** Mining and Quarrying
- **Sector 3:** Food, Beverages and Tobacco, Textiles and Textile Products, Leather, Leather and Footwear, Wood and Products of Wood and Cork, Pulp, Paper, Paper , Printing and Publishing, Coke, Refined Petroleum and Nuclear Fuel, Chemicals and Chemical Products, Rubber and Plastics, Other Non-Metallic Mineral, Basic Metals and Fabricated Metal, Machinery, Nec, Electrical and Optical Equipment, Transport Equipment, Manufacturing, Nec, Recycling
- **Sector 4:** Electricity, Gas and Water Supply
- **Sector 5:** Construction
- **Sector 6:** Sale, Maintenance and Repair of Motor Vehicles Retail Sale of Fuel, Wholesale Trade and Commission Trade, Except of Motor Vehicles, Retail Trade, Except of Motor Vehicles ; Repair of Household Goods, Hotels and Restaurants
- **Sector 7:** Inland Transport, Water Transport, Air Transport, Other Supporting

and Auxiliary Transport Activities; Activities of Travel Agencies, Post and Telecommunications

- **Sector 8:** Financial Intermediation, Real Estate Activities, Renting of M&Eq and Other Business Activities
- **Sector 9:** Public Admin and Defence; Compulsory Social Security, Education, Health and Social Work
- **Sector 10:** Other Community, Social and Personal Services, Private Households with Employed Persons

I compare the aggregate GDP per country-sector that results from aggreagating sectors in this fashion across the WIOD, with production per section and per country reported in the GGDC to verify that this aggregation is correct.

E Additional Tables and Figures

E.1 Robustness: Regressions Using Longer Lags

As explained in Section 2.3, one concern about the specification may be that a year is not enough for the effect to materialize, or at least not fully. This is because the mechanism requires for one party (the importer) to learn from the foreign seller. Given how learning works in the model as derived in Section 2.1, the main specification only includes the first lag. However to address this concern, Table 4 includes results in which I inlcude further lags for the independent variables, namely, 3 lags for each of them. This exercise allows me to assess if there is a "time-to-build" the stock of knowledge after a trade interaction, and whether or not it is empirically correct to include only the first lag.

As shown, the exercise reveals no evidence of lags of order higher than 1 having more importance than the first one. First, none of the lags is larger in magnitude (iwhich seems to decrease on the lag's order). Second, these higher order lags in general are not significant. Column (1) shows how only the ratio for the first lag is significant (and at 1%); the second lag, while not significant, keeps the same sign as the first one with a lower magnitude; the third lag is the smallest in absolute value, and the most insignificant as well. Other independent variables show a similar pattern: the trading partner's productivity is significant and with the expected sign only for the first lag, while the interaction term's significance is also decreasing on the lag. Therefore, all the qualitative findings remain unchanged and consistent with the results presented in section 2.3, with little impact on the magnitudes of the coeficcients.

Table 4: Trade and Changes in Productivity Using Longer Lags

Dependent variable: change in productivity of each country's industry $\left(\Delta z_{it+1}^{k}\right)$.

Log Variables	(1)	(2)	(3)
Ratio $(t-1)$,	-0.232	-0.239^{**}	$-0.362*$
	(0.209)	(0.085)	(0.198)
Ratio $(t-2)$,	0.068	0.076	0.331
	(0.041)	(0.085)	(0.187)
Trading partner's		0.717^{***}	0.824^{***}
productivity (t)		(0.153)	(0.166)
Trading partner's		-0.05	0.066
productivity $(t-1)$		(0.167)	(0.209)
Trading partner's		-0.04	-0.259
productivity $(t-2)$		(0.149)	(0.156)
$\left(\frac{\hat{z}_{it}^{k}}{z_{it}^{k}}\right) \times \hat{z}_{it}^{k}\ (t)$			0.032
			(0.048)
$\left(\frac{\hat{z}_{it}^{k}}{z_{it}^{k}}\right) \times \hat{z}_{it}^{k}(t-1)$			0.023
			(0.030)
$\left(\frac{\hat{z}_{it}^{k}}{z_{it}^{k}}\right) \times \hat{z}_{it}^{k}(t-2)$			-0.047
			(0.034)
FE	✓	✓	✓

Notes: Table 4 reports the regression results when the dependent variable is the change in productivity of each country's industry $\left(\Delta z_{it+1}^{v}\right)$ with longer lags than Table 1. All productivities are demeaned using the mean of a country's industry. Independent variables are in logs. In all cases, the number of observations equals 1,635. Constants not reported. Robust SE are in parentheses. Significance levels: $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.2 Regression Over a Sub-sample

Table 5 presents the results of Table 1, but when dropping observations after year 2007. Namely, I run regression 1 over the period 1995-2007, to exclude the great recession and study the sensitivity of the results to the time period considered. As shown, qualitatively the results are preserved, while quantitatively the use of a subsample has little impact on the coeficcients.

Table 5: Trade and Changes in Productivity in a Subsample

Dependent variable: change in productivity of each country's industry $\left(\Delta z^k_{it+1}\right)$.

	(1)	(2)	(3)
Ratio $\left(\frac{\hat{z}^k_{it}}{z^k_{it}}\right)$	-0.842^{***}	-0.523^{***}	-0.648^{***}
	(0.048)	(0.033)	(0.075)
Trading partner's productivity $\left(\hat{z}^k_{jt}\right)$		0.625^{***}	0.641^{***}
		(0.012)	(0.014)
$\left(\frac{\hat{z}^k_{it}}{z^k_{it}}\right) \times \hat{z}^k_{it}$			0.028^{**}
			(0.006)
FE	✓	✓	✓
Adj R^2	0.67	0.97	0.97

Notes: Table 5 reports the regression results when the dependent variable is log change in productivity of each country's industry $\left(\Delta z^k_{it+1}\right)$. All productivities are demeaned using the mean of a country's industry during the period studied (1995-2010). Independent variables are also in logs. Average productivity of the trading partner of each industry-country is constructed based on industry-country pairs, according to **??**. In all cases, the number of observations equals 1,524. Constants not reported. Robust standard errors are in parentheses. Significance levels: $^{*}p < 0.10$, ** $p < 0.05$, *** $p < 0.01$