Economic Growth among a Network of Industries*

Will Johnson†

April 30, 2015

Abstract

Which industries contribute the most to aggregate economic growth, when taking into account not only each industry's own value added, but also technological spillovers across industries? To shed light on this question, I develop a model of economic growth in which technological spillovers induce a network structure among industries. As an endogenous result of the model, the more similar that two industries' production functions are to each other, the greater the technological spillovers between them, i.e., the stronger the link between the two industries in the network. An industry contributes more to aggregate growth when the industry has a more central position in the network. The closed-economy version of the model is tested using data on growth in each industry in the US from 1960 to 2005, while the open-economy version of the model is tested using data on bilateral, industry-specific trade flows among 72 countries, from 1962 to 2000; I find that the model explains relative growth in each industry and changing patterns of trade over this period better than the null model with no network effects.

JEL Codes: O41, F11, D85

Keywords: economic growth, inter-industry networks, learning-by-doing spillovers

*First and foremost I would like to thank Dilip Mookherjee for his intellectual guidance and support, as well as Samuel Bazzi and Stefania Garetto for many valuable contributions. This research has also been shaped by helpful comments from Kehinde Ajayi, Amrit Amirapu, Dany Bahar, Sebastian Bustos, Ricardo Hausmann, Frank Neffke, Andrew Newman, Michael Peters, John van Reenen, Esteban Rossi-Hansberg, and seminar and conference participants at Azim Premji University, Boston University, Columbia University, Concordia University, and Harvard University.

†Boston University's Department of Economics, and Harvard University's Center for International Development. Email: willj@bu.edu

1 Introduction

There are striking asymmetries in the goods that rich countries and poor countries produce and trade with each other. For example, the leading exports of South Korea are semiconductors, wireless telecommunications, and motor vehicles, while the leading exports of Mozambique are aluminum, prawns, and cashews.[1] This paper explores the hypothesis that moving into different industries is an important pathway to economic growth, and poor countries are poor partly because government or market failures block that pathway.

This raises the question, which industries are the most conducive to aggregate growth? An answer to this question must account not only for growth in each industry's own value added, but also technological spillovers across industries. The hypothesis underlying this paper is that the size of technological spillovers from one industry to another depends on how *similar* the two industries are. I develop a model of economic growth in which, as an endogenous result, this notion of similarity ends up playing a front-and-center role. The intuition is simple. In the model, the source of economic growth is learning-by-doing among workers; the learning-by-doing spills over among all workers within a given occupation, regardless of the industry for which they are working. This generates inter-industry learning spillovers, where the size of the spillovers between any two industries depends on how similar the two industries are to each other, in terms of the intensity with which they use each occupation.

Given these inter-industry learning spillovers, we can then think of industries as forming a network, where each industry is a node in the network, and the strength of the link between any two industries corresponds to the size of the learning spillovers between them. The amount that an industry contributes to long-run aggregate economic growth corresponds to how central the industry is in this network.

The key parameters of the model are simply the intensity with which each industry uses each occupation, which can be directly taken from data. That makes this, to the best of my knowledge, the first model of endogenous long-run growth in which "not all industries are created equal" and which can be taken directly to data. I test the closed-economy version of the model against data on growth in each industry in the US from 1960 to 2005, and I test the open-economy version of the model against data on bilateral, industry-specific trade flows among 72 countries, from 1962 to 2000; in each case I find that the model performs significantly better than the null model with no network effects.

This paper combines two broad strands of literature. At the heart of my model is the notion of economic growth through learning-by-doing, which dates back to Arrow (1962). The first to analyze this within a multi-industry framework were Clemhout and Wan (1970) and Bardhan (1971), who showed that if certain industries exhibit more learning-by-doing than others, and if this learning is external to individual firms, then this gives theoretical (although not necessarily practical) justification for subsidizing the industries with more learning. These considerations are further amplified by international trade[2]: if rich coun-

[1]Source: CIA World Factbook. For more systematic documentation of such differences, see Rodrik (2006).

[2]Besides the seminal studies mentioned above, see, for example, Succar (1987), Young (1991), and

tries have a comparative advantage (in the static sense) in high-growth industries (i.e., industries with large learning-by-doing externalities) while poor countries have a (static) comparative advantage in low-growth industries, then international trade between rich and poor countries will boost rich countries' growth but stifle poor countries' growth.[3]

The inter-industry learning spillovers that endogenously result from my model connect this paper with a recently flourishing literature on the macroeconomic implications of network structures among industries. Particularly large is the literature on the business cycle implications of the input-output structure of the economy, dating back to Long and Plosser (1983), and more recently Carvalho (2010) and Acemoglu et al (2012).[4]

Long-run growth implications of the input-output structure of the economy were first emphasized by Hirschman (1958), who argued for directing investment toward industries that use many inputs ("backward linkages") and that in turn are used by many industries as inputs ("forward linkages"); this view was first formalized by Rodriguez-Clare (1996a) and Rodriguez-Clare (1996b). Jones (2011) develops a static model in which forward and backward linkages amplify the effects of exogenous sector-specific distortions on aggregate total factor productivity, using US input-output data for illustration.

Note, however, that the network of industries in this paper is *not* based on input-output linkages, but instead on industries learning from one another, where the size of the learning spillovers between two industries depends on their *similarity*. In this respect this paper complements the work of Hidalgo, Klinger, Barabasi, and Hausmann (2007), in which, using a very different methodology, they envision industries as forming a network (which they call the *product space*), in which a link between two industries represents overlap in the capabilities required to produce in those industries.

Although the two ideas (input-output linkages and learning spillovers) sound superficially similar (they both involve "links" in a "network"), they are in fact very different. Consider, for example, the oil industry. In terms of the input-output structure of the economy, since many other industries use oil, then oil has a central place in the network, and a model of long-run development that stresses input-output linkages would imply that if a country produces oil, this will spur development in those other industries. In contrast, the framework in this paper calls to our attention the fact that the oil industry has very little in common with most other industries – the skills that oil workers develop have little use in other industries. Therefore, in the framework of this paper, the oil industry is at the periphery of the network[5], and according to this model, a country that produces oil is likely to get stuck producing little else.

The rest of this paper is organized as follows. In section 2, I introduce the model in

Dehejia (1993), among many others; a brief survey of this literature is given by Acemoglu (2008). Note that these papers are purely theoretical.

[3]There are also certainly mechanisms by which international trade can stimulate growth in poor countries, such as technology transfer and access to cheaper physical and human capital. In this paper, however, I abstract away from these important mechanisms.

[4]In future drafts, I plan to discuss in further depth how my analysis ties in with this business cycle literature.

[5]See Figure 1 in section 3.3.

the context of a single closed economy and derive theoretical results. In section 3, I test the closed-economy model through a quantitative exercise using US data. In section 4, I extend the theoretical and quantitative analysis into a multi-country framework. In section 5, I conclude and discuss further research.

2 Closed-economy model

The basic intuition underlying the model is as follows. There are multiple industries and multiple occupations – think of "industries" as the car industry, the finance industry, and so on, while "occupations" are engineers, economists, and so on.[6] There is learning-by-doing within each occupation, which spills over to everyone in the occupation regardless of the industry for which they are working.

Consider, then, what happens if production increases in the car industry. Since the car industry employs a large number of engineers but only a small number of economists, this will cause a significant increase in learning-by-doing among engineers, not so much among economists. The extent to which this benefits another industry corresponds to how much that other industry is engineer-intensive rather than economist-intensive – it will benefit the airplane manufacturing industry more than the finance industry, since the former is engineer-intensive while the latter is economist-intensive.

We can then think of industries as forming a network, where, for any two industries, the strength of the link between them corresponds to how similar they are in their intensity of usage of different occupations. An industry that is more central in this network generates more learning spillovers and thereby contributes more to aggregate economic growth.

With this intuition in the back of our minds, let us now turn to the formal model.

2.1 The model

Consider a closed economy with J industries, indexed by j; K occupations, indexed by k; and an arbitrary finite or countably infinite number of time periods, indexed by t.

Let X_{jkt} be the effective units of labor in occupation k used by industry j at date t, let Y_{jt} be the amount of industry j's output at date t, and let Y_t be total utility at date t.

At each date t, a representative agent[7] chooses how much of each occupation to use in each industry – that is, she chooses $\{X_{jkt}\}_{j,k}$ – in order to maximize date-t utility, which is a Cobb-Douglas function of how much she produces in each industry:

[6] We will revisit the issue of what the specific industries and occupations are when discussing different sources of data in sections 3.1, 3.4, 4.3, and 4.4.

[7] Rather than a single representative agent, we can alternatively think of a representative consumer, a representative firm for each industry, and a representative worker in each occupation, with these agents buying and selling with each other in perfectly competitive markets. It is straightforward to show that such an alternative formulation would not affect any of the model's predictions – the key is that I assume the representative agent efficiently allocates resources in a static sense at each date t but does not internalize learning-by-doing externalities, exactly as would be the case in the alternative version of the model with multiple types of agents.

$$Y_t = \prod_j Y_{jt}^{\beta_j}$$

I assume $0 \leq \beta_j \leq 1 \; \forall j$ and $\sum_j \beta_j = 1$.

Each industry j is produced according to a Leontief production function[8]; industries can vary in which occupations they use intensively:

$$Y_{jt} = \min\{\frac{1}{\alpha_{jk}} X_{jkt}\}_k \quad \forall j$$

I assume $0 < \alpha_{jk} \; \forall j, k$ and $\sum_k \alpha_{jk} = 1 \; \forall j$.

The agent faces the following labor endowment constraint at each date t:

$$\sum_j \sum_k \psi_{kt} X_{jkt} = L$$

where L is the economy's total labor force, while ψ_{kt} represents the cost at time t of producing an occupation-k worker of a fixed level of productivity, or equivalently, $\psi_{kt} = \frac{1}{\pi_{kt}}$, where π_{kt} is the productivity of an occupation-k worker at time t.

Occupational productivity evolves over time from learning-by-doing:

$$\psi_{k,t+1} = \psi_{kt}(\tilde{X}_{kt})^{-\rho} \quad \forall k, t \tag{1}$$

where $\rho > 0$ is the elasticity of learning-by-doing with respect to relative occupational usage, $\psi_{k,0}$ is given $\forall k$, and $\tilde{X}_{kt}$ denotes one plus the fraction of the effective labor force employed in occupation k (across all industries) at time t – that is,

$$\tilde{X}_{kt} \equiv 1 + \frac{\sum_j X_{jkt}}{\sum_j \sum_k X_{jkt}}$$

However, the representative agent does not internalize the dynamic returns from learning-by-doing, and so she does not take into account equation 1 when deciding on $\{X_{jkt}\}_{j,k}$.

Equation (1) is saying, for example, that the higher the fraction of the work force working as engineers, the more learning-by-doing there will be among engineers. This learning-by-doing spills over across all engineers, regardless of which industry they are working in. Note that ρ, the elasticity of learning-by-doing with respect to relative occupational usage, does not depend on the occupation. One might suspect, however, that some occupations have more learning-by-doing than others – e.g., only a very exuberant economist would believe that his field has as much learning-by-doing as engineering. But this is precisely the kind of asymmetry I do *not* want to take a stand on, for the reasons expressed by Hausmann, Hwang, and Rodrik (2007) above, so I let ρ be the same for all occupations.

[8] In future work I aim to determine how much the results generalize beyond Leontief production functions.

2.2 Equilibrium

Given the parameters $\{\psi_{k0}\}_k$, $\{\alpha_{jk}\}_{j,k}$, $\{\beta_j\}_j$, and L, an *equilibrium* of the economy is defined as a path $\{X_{jkt},\, Y_{jt},\, Y_t,\, \psi_{kt}\}_{jkt}$ such that at each date t,

1. Occupational employment choices $\{X_{jkt}\}_{jk}$ (along with the corresponding industrial outputs $\{Y_{jt}\}_i$ and utility Y_t) solve the representative agent's date-t maximization problem given date-t occupational costs $\{\psi_{kt}\}_k$.

2. Occupational costs $\{\psi_{kt}\}_k$ evolve over time according to equation 1.

In equilibrium, because the production functions are Leontief, occupations are always in fixed proportions with industrial outputs:

$$X_{jkt} = \alpha_{jk} Y_{jt} \quad \forall j, k, t \tag{2}$$

Thus the cost of producing one unit of Y_{jt} can be expressed as

$$P_{jt} \equiv \sum_k \alpha_{jk} \psi_{kt} \tag{3}$$

The agent's date-t problem then collapses to choosing output in each industry $\{Y_{jt}\}_j$ to maximize utility Y_t given the date-t costs of producing in each industry $\{P_{jt}\}_j$. The solution to this collapsed problem is the standard solution to a Cobb-Douglas maximization problem:

$$Y_{jt} = \frac{\beta_j L}{P_{jt}} = \frac{\beta_j L}{\sum_k \alpha_{jk} \psi_{kt}} \quad \forall j, t \tag{4}$$

For a given specification of parameters, there is a unique equilibrium characterized by equations 1, 2, and 4.

2.3 Inter-industry spillovers

To see how the learning-by-doing induces a network structure among industries, consider an increase in production in some arbitrary industry j and what effect this has on some other arbitrary industry h.

If Y_{jt} increases by one unit, then by equation 2, for each occupation k, X_{jkt} increases by α_{jk} units. By taking the derivative of equation 1, we see that for a given occupation k, a one unit increase in X_{jkt} causes the next-period cost of occupation k, $\psi_{k,t+1}$, to decrease by Q_{kt} units, where

$$Q_{kt} \equiv \rho \psi_{kt} (\tilde{X_{kt}})^{-\rho-1} [(\sum_k \sum_j X_{jkt})^{-1} - (\sum_j X_{jkt})(\sum_k \sum_j X_{jkt})^{-2}] \tag{5}$$

Meanwhile, for arbitrary industry h, by (3), a one unit decrease in the cost of occupation k, $\psi_{k,t+1}$, causes the cost of a unit of production in industry h, $P_{h,t+1}$, to decrease by α_{hk} units.

Combining the above results, we have, for arbitrary industries j and h, a one unit increase in industry-j production causes the next-period cost of a unit of industry-h production to decrease by $\sum_k Q_{kt}\alpha_{jk}\alpha_{hk}$ units. Industries can therefore be seen as forming a network in which the strength of the link between industries j and h at date t is $\sum_k Q_{kt}\alpha_{jk}\alpha_{hk}$.

To gain a more intuitive grasp of this, ignore the Q_{kt} terms for a moment and note that $\sum_k \alpha_{jk}\alpha_{hk}$ is the *correlation coefficient* between the occupational intensities of industries j and h, which is a natural way of measuring how similar industries j and h are in terms of their production functions. All else being equal, the more similar two industries are to each other, the more technological spillovers there will be between them. This is what the above result is saying, except we weight each occupation k by Q_{kt}, since that tells us where we are on occupation k's learning curve.

It is important not to confuse the result in this section with the older idea, a la Hirschman (1958), that the economy can be thought of as a network of industries in which the links in the network represent the industries' direct buying and selling relationships with one another. Instead, as discussed in the introduction, my paper captures a very different idea, in which links in the network of industries represent technological spillovers based on how *similar* industries are to each other.[9]

Note that, by (5), the Q_{kt} terms involve endogenous variables, and so the network structure evolves endogenously over time. Exploring this endogenous evolution of the network is potentially an interesting direction for future research, but it is difficult to gain any analytical traction without making further assumptions. For this reason, in the next section I analyze a tractable log-linear approximation of the model.

2.4 Log-linear approximation

Recall that J is the number of industries. Consider the *symmetric* case in which, at some date t, all occupational productivities are the same and production is equal across industries, that is, $\psi_{kt} = \psi_{mt}\ \forall k, m$ and $Y_{jt} = Y_{ht}\ \forall j, h$.[10] I will make use of the following fact, the proof of which is trivial:

Fact For any arbitrary set of variables $\{z_n\}_n$ and constants $\{a_n\}_n$, if $\sum_n a_n = 1$ and $z_n = z_m\ \forall n, m$, then $log(\sum_n a_n z_n) = \sum_n a_n log(z_n)$.

Combining this fact with the three equilibrium equations 1, 2, and 4, we have

[9] In future drafts of this paper I plan to do a more thorough comparison between my analysis and what we would get from a Hirschman-type model.

[10] However, we do *not* need to assume any symmetry with regard to how much each industry uses each occupation, which is important − such differences in occupational usage are what we are most interested in; as we saw in the previous section, these are what give us the network structure of industries.

$$
\begin{aligned}
log(Y_{h,t+1}) = log(\beta_h L) - \sum_k \alpha_{hk} log(\psi_{kt}) \\
+ \rho \sum_k \alpha_{hk} log(\sum_j \alpha_{jk}) + \rho \sum_k \alpha_{hk} \sum_j (\frac{\alpha_{jk}}{\sum_m \alpha_{mk}}) log(Y_{jt}) \\
- \rho \sum_k \alpha_{hk} log(J) - \rho \sum_k \alpha_{hk} \sum_j (\frac{1}{J}) log(Y_{jt}) \quad (6)
\end{aligned}
$$

By taking the derivative of (6) with respect to $log(Y_{jt})$, we see that a 1% increase in Y_{jt} causes an A_{jh}% increase in $Y_{h,t+1}$, where

$$
A_{jh} \equiv \rho \sum_k \alpha_{hk} (\frac{\alpha_{jk}}{\sum_m \alpha_{mk}} - \frac{1}{J}) \quad (7)
$$

The intuition for this result is similar to the intuition from the previous section. An increase in Y_{jt} corresponds to an increase in the usage of each occupation k. The extent to which this increases learning-by-doing within occupation k is proprotional to ρ (since ρ is the elasticity of learning-by-doing with respect to occupational usage) as well as $(\frac{\alpha_{jk}}{\sum_m \alpha_{mk}} - \frac{1}{J})$, which captures how much usage of occupation k is concentrated in industry j relative to how much the overall economy is concentrated in industry j (the latter being $\frac{1}{J}$, since there are J industries and we are assuming symmetry) $-$ if the latter is greater than the former, then increasing Y_{jt} causes *less* learning-by-doing in occupation k, since learning-by-doing is a function of the *fraction* of the economy's effective labor force employed in occupation k. Meanwhile, the extent to which an increase in learning-by-doing for occupation k benefits industry h is proportional to α_{hk}.

The beauty of this result, compared to the result from the previous section, is that, as we see from (7), the effect that growth in one industry has on growth in another industry is now only a function of exogenous parameters. It does not vary over time, which is important for the theoretical results that follow in the rest of this section.

This result holds exactly only in the symmetric case, but let us take it as an approximation of the behavior of the model outside of it.[11] Consider, then, an arbitrary equilibrium path $\{Y^\star_{jt}\}_{j,t}$ and corresponding $\{Y^\star_t\}_t$. Suppose, starting from this equilibrium path, we increase Y_{jt} by 1% for some industry j at some date t. What effect will this have on total discounted social welfare from date t onward? How does our answer depend on which industry we are giving the shock to?

To answer these questions, let lowercase letters denote log-deviations from the previous equilibrium path. Specifically, let $y_{jt} \equiv log(Y_{jt}) - log(Y^\star_{jt})$ and $y_t \equiv log(Y_t) - log(Y^\star_t)$. Let $\vec{y_t}$ denote the J-dimensional vector specifying y_{jt} for each industry j. Let y denote the discounted sum of log-deviations in utility over time, i.e., let $y \equiv \sum_t \delta^t y_t$, where $0 < \delta < 1$ is the representative agent's discount factor. Let β denote the J-dimensional

[11]In future work I plan to investigate, analytically and/or numerically, how good the approximation is.

vector specifying the Cobb-Douglas exponent β_j for each industry j, and let I denote the JxJ identity matrix.

Recall, we are maintaining, as an approximation, that starting from any arbitrary point, a 1% increase in Y_{jt} causes an A_{jh}% increase in $Y_{h,t+1}$, where A_{jh} is defined by (7). It follows, then, that if we let A denote the JxJ matrix whose (j, h) element is A_{jh}, then $\vec{y}_{t+1} \approx A\vec{y}_t$. Let us now calculate $\frac{dy}{d\vec{y}_t}$, which is the J-dimensional vector whose jth element gives the total discounted sum of percent increases in utility from a 1% increase in Y_{jt}. We have the following result:

$$\frac{dy}{d\vec{y}_t} \approx \beta + \delta A\beta + \delta^2 A^2\beta + \delta^3 A^3\beta + \dots$$
$$= \beta + \delta(\mathrm{I} + \delta A + \delta^2 A^2 + \dots)A\beta$$
$$= \beta + \delta(\mathrm{I} - \delta A)^{-1}A\beta \quad (8)$$

The right-hand side of (8) is the J-dimensional vector whose jth element is the *Bonacich centrality* of industry j in the network of industries. Bonacich centrality is a measure of how important a node is in a network. For example, in a network of friends, the Bonacich centrality of an individual is her number of friends plus (for some discount factor δ) δ times the number of friends her friends have, plus δ^2 times the number of friends her friends of friends have, and so on. Equation (8) involves the same thing, except instead of weighting all nodes equally (as one usually would when analyzing a network of friends), each industry j is weighted by its Cobb-Douglas exponent β_j, since that is how much it contributes directly to utility.

3 Closed-economy quantitative exercise

In this section I test to see how well the model explains the U.S.'s economic growth, broken down by industry, over the past half century.

3.1 Data

There are two kinds of data that shed light on the key parameters of the model, namely the α terms, where α_{jk} is the intensity with which industry j uses occupation k. The most direct source of data is an industry-occupation table, which tells us the number of people in each occupation employed by each industry. A more indirect source of data is an input-output table, which tells us how much each industry uses the output of each other industry as an input.

We can use an input-output table as a proxy for an industry-occupation table by relabeling the columns of the input-output table as occupations (e.g., relabeling "coal mining" to "coal miners") – if the electric utilities industry uses a large amount of coal mining output as an input, then the electric utilities industry uses (indirectly) a large number of coal miners. When using an input-output table in this way (unlike the typical way of using an input-output table), my measure of the strength of the link between two industries (which I get from the model) is how *similar* the two industries are in terms of how intensely they use different inputs, *not* how much the two industries purchase from each other, just as when using an industry-occupation table, the measure is how similar the two industries are in how intensely they use each occupation. These two methods should give similar results.

In this closed-economy section of the paper I use input-output tables, while in the international section of the paper I use industry-occupation tables. This choice was arbitrary; in the future, for the sake of robustness, I plan to use both for both.

Here I use KLEMS data from Jorgenson (2007) that divide the U.S. economy into 35 industries (mostly at the 2-digit SIC level), spanning from 1960 to 2005.[12] These data include, for each year, each industry's output as well as each industry's usage of 37 inputs (capital, labor, and 35 intermediate inputs, one for each of the 35 industries). All output and input variables are adjusted for industry-specific price changes, with 1996 used as the base year for all prices. The 35 industries, and their corresponding identifying numbers (which will be used when reporting results) are given in Table 1.

Table 1: The 35 industries and their ID #'s

ID #	Industry	ID #	Industry
1	Agriculture, forestry, fisheries	19	Stone, clay and glass products
2	Metal mining	20	Primary metals
3	Coal mining	21	Fabricated metal products
4	Crude oil and gas extraction	22	Non-electrical machinery
5	Non-metallic mineral mining	23	Electrical machinery
6	Construction	24	Motor vehicles
7	Food and kindred products	25	Other transportation equipment
8	Tobacco manufactures	26	Instruments
9	Textile mill products	27	Miscellaneous manufacturing
10	Apparel and other textile products	28	Transportation and warehousing
11	Lumber and wood products	29	Communications
12	Furniture and fixtures	30	Electric utilities (services)
13	Paper and allied products	31	Gas utilities (services)
14	Printing and publishing	32	Wholesale and retail trade
15	Chemicals and allied products	33	Finance, insurance and real estate
16	Petroleum refining	34	Personal and business services
17	Rubber and plastic products	35	Government enterprises
18	Leather and leather products		

[12]The ultimate source of these data are apparently the Bureau of Economic Analysis (BEA) and Bureau of Labor Statistics (BLS), but I have yet to find any documentation that pinpoints the specific BEA and BLS data series used or the specific procedures used to transform the BEA and BLS data into the dataset with which I am working.

3.2 The exercise

I perform the following quantitative exercise:

1. Calibrate $\{\beta_j\}_j$ (the Cobb-Douglas exponents in the utility function) and L (the size of the economy, which in the model is the size of the labor force but in these data I will take to be GDP, which can be thought of as the same thing in a model where labor is the only factor of production) using the 1960 data on output in each industry. Summing output across all industries in 1960 gives us L. For each industry j, dividing industry j's 1960 output by total economy-wide 1960 output gives us β_j.

2. Calibrate $\{\alpha_{jk}\}_{j,k}$ (the occupational intensities in the Leontief production functions) using the 1960 data on each industry's usage of the 35 different intermediate goods. Specifically, α_{jk} equals industry j's 1960 usage of intermediate good k divided by industry j's 1960 total usage of all intermediate goods.[13] (By construction, $\sum_k \alpha_{jk} = 1\ \forall j$.)

3. Set $\{\psi_{k0}\}_k$ (the starting cost of each occupation) equal to one for every occupation. This may seem strange, but it actually fits with the way I calibrate the above parameters, in the sense that, with these values, the model will, by construction, perfectly fit the 1960 data, whereas any other values for $\{\psi_{k0}\}_k$ would throw the model off from the 1960 data.

4. Use grid search to select ρ (the speed of learning-by-doing). Allow it to take on any of the following values: 0.00, 0.01, 0.02, ... , 0.98, 0.99, 1.00. The remaining steps below are for a specific value of ρ; the criterion for selecting ρ is described below.

5. Given the above parameters, the model makes predictions of Y_{jt} (output of industry j at date t) for all j and t starting from 1960. Run the model for 46 periods, representing 1960 to 2005.

6. Compare the model's predictions for 2005 with the 2005 data. More specifically, calculate $Corr(Y_{j,2005}, \hat{Y}_{j,2005})$ where $\hat{Y}_{j,2005}$ is what the model predicts for output of industry j in 2005, and $Y_{j,2005}$ is from the actual data.

7. Choose the value of ρ that maximizes $Corr(Y_{j,2005}, \hat{Y}_{j,2005})$.

8. Report the optimal value of ρ and the maximized value of $Corr(Y_{j,2005}, \hat{Y}_{j,2005})$ (along with the scatterplot for $Y_{j,2005}$ vs. $\hat{Y}_{j,2005}$).

[13] An alternative approach would be to use all 37 inputs, including labor and capital. However, for most industries, labor and capital usage dwarf the usage of any given intermediate good, and so if we were to take such an approach, our measure of how related two industries are would be disproportionately based on the similarity of their capital and labor intensities, which is not a good measure of relatedness, since it contains very little information.

The maximized value of $Corr(Y_{j,2005}, \hat{Y}_{j,2005})$, as a measure of how successful the model is in explaining the data, has little meaning in and of itself; we have to ask, the model is successful relative to what? A natural benchmark is to see how well the model matches the data when setting ρ to zero, which corresponds to no learning-by-doing and hence no network effects. This is equivalent to letting the 1960 equilibrium repeat itself over and over again, i.e., it gives us $Corr(Y_{j,2005}, Y_{j,1960})$. So we compare $Corr(Y_{j,2005}, \hat{Y}_{j,2005})$ vs. $Corr(Y_{j,2005}, Y_{j,1960})$ – that is, we see how successful the model is at predicting future data relative to the null model that says that future data will just be the same as past data.

One important remark about this quantitative exercise is that these kind of input-output data provide us with more information about the structure of the economy than what we are using here. In particular, if industry j uses a lot of industry h's output as an input, this, in and of itself, means that industries j and h have a direct purchasing relationship with one another, but we are ignoring this aspect of the data in this exercise. Instead we are using these input-output data to measure the technological relatedness of industries to each other. This goes back to the difference between this paper and the traditional way of thinking about the economy as a network, as discussed in the introduction and in section 2.3.

3.3 Results

Before examining the results of the quantitative exercise, let us first examine what the network of industries actually looks like. To calculate the network structure, we take the values of ρ (the elasticity of learning-by-doing with respect to occupational usage) and $\{\alpha_{jk}\}_{j,k}$ (the intensities with which each industry j uses each occupation k) that we obtained from the quantitative exercise and then plug these values into equation (7) for each pair of industries j and h, which gives us the strength of the learning-spillover link between any two industries. Figure 1 gives a visual depiction of the network. Note that the strength of the link between any two industries is a continuous variable that can be positive or negative; in Figure 1, a line is drawn between two industries if they have a positive link, and no line is drawn if their link is zero or negative.

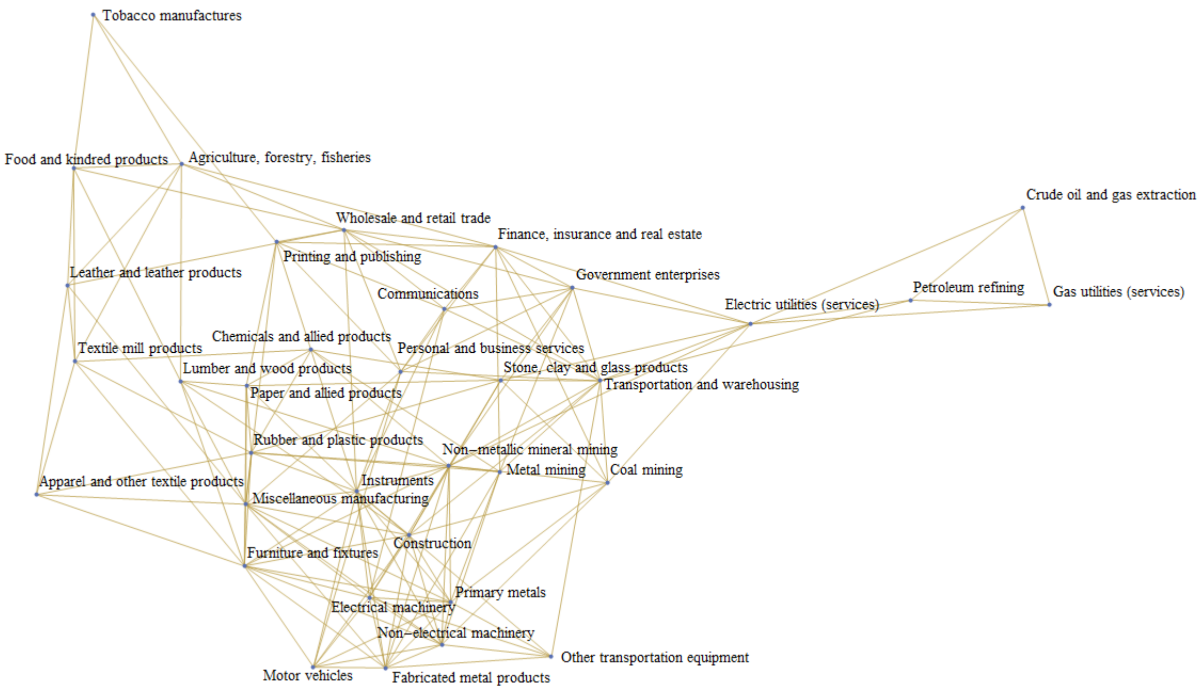
The value of ρ (the elasticity of learning-by-doing with respect to occupational usage) that minimizes the objective function[14] is 0.46, meaning a 1% increase in relative usage of an occupation one year increases the productivity of that occupation the next year by 0.46%.[15]

Figure 2 plots each industry's share of 2005 aggregate output along the X axis against the model's predictions along the Y axis. Industries are marked by their ID# from Table 1. If the model perfectly fit the data, each industry would lie along the 45-degree line,

[14] Recall from section 3.2 that the objective function here is the sum of squared errors between each industry's share of aggregate output in 2005 and the model's predictions for these shares.

[15] In the future I plan to examine how this result compares with other studies' estimates of the rate of learning-by-doing. Note that making such comparisons is a nontrivial exercise, since different studies assume a variety of functional forms for learning-by-doing.

Figure 1: Visual depiction of the network of industries



which is represented by the dashed line. As we can see, the model fits the data well; the correlation coefficient between the data and the model's predictions is .96.

Now consider the null model, i.e., the model when plugging in $\rho = 0$, or in other words, the model with no learning-by-doing and hence no network effects. Since the null model simply predicts the same static equilibrium repeated over and over again, this is equivalent to examining how well we can predict each industry's share of 2005 aggregate output by simply using each industry's share in 1960. This is plotted in Figure 3.

As we can see, the model matches the data better than the null model; .88 is the correlation coefficient between the data and the null model's predictions[16], compared to .96 for the model. Another way of comparing the two is that the sum of squared errors is 0.017 for the null model and 0.0057 for the actual model[17]; the sum of squared errors of the actual model is 66% lower than the sum of squared errors of the null model.

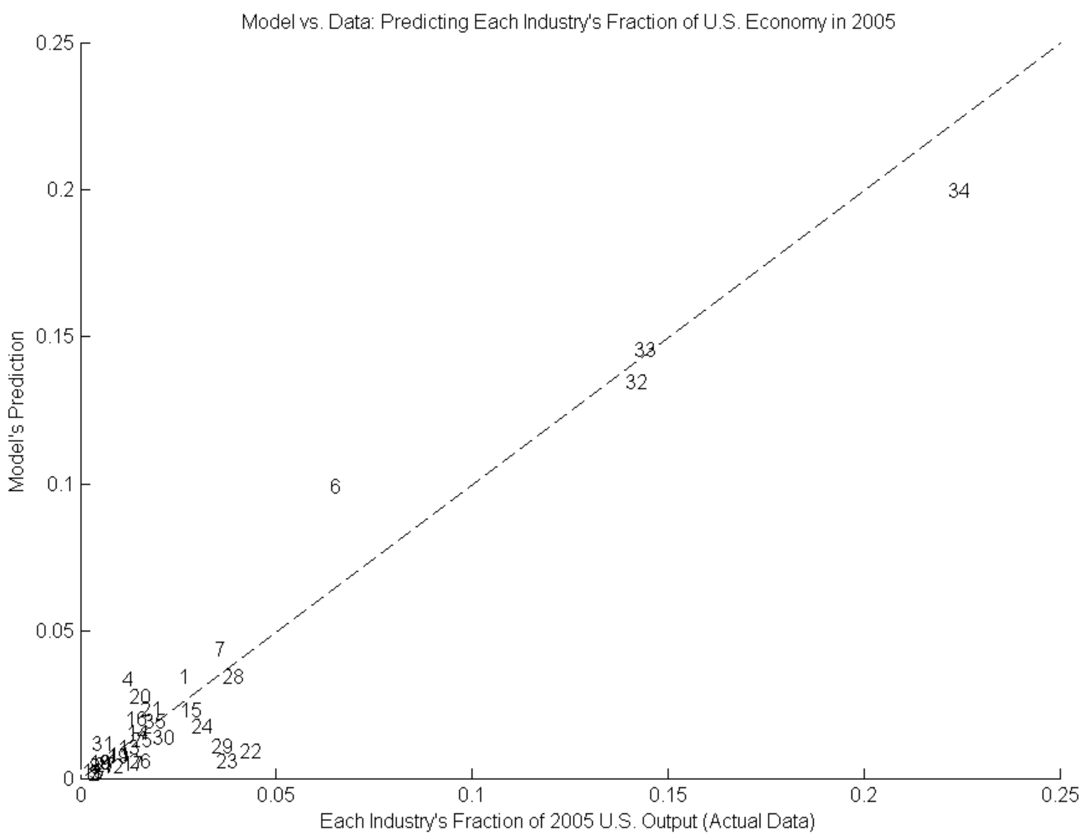
Now let us look more directly at how well the model (the actual model, not the null model) predicts changes in industrial composition. Figure 4 plots the *change* between 1960 and 2005 in each industry's share of aggregate output along the X axis against the model's predictions along the Y axis. (Note: The reason the 45-degree line [the dashed line] looks steeper is because of the different axes.)

By this metric, the model still performs reasonably well; the correlation coefficient

[16] Equivalently, .88 is the correlation between the data in 1960 and the data in 2005.

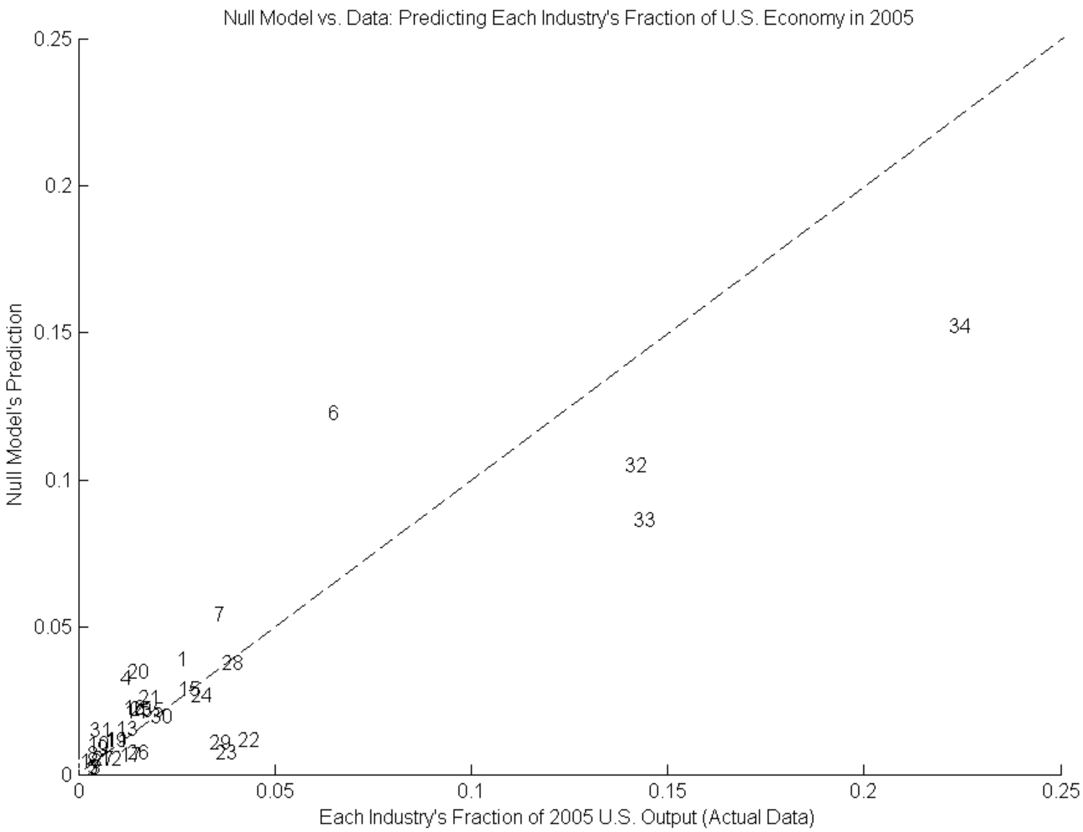
[17] These numbers do not have units, since the outcome being measured here is an industry's *fraction* of aggregate output.

Figure 2:



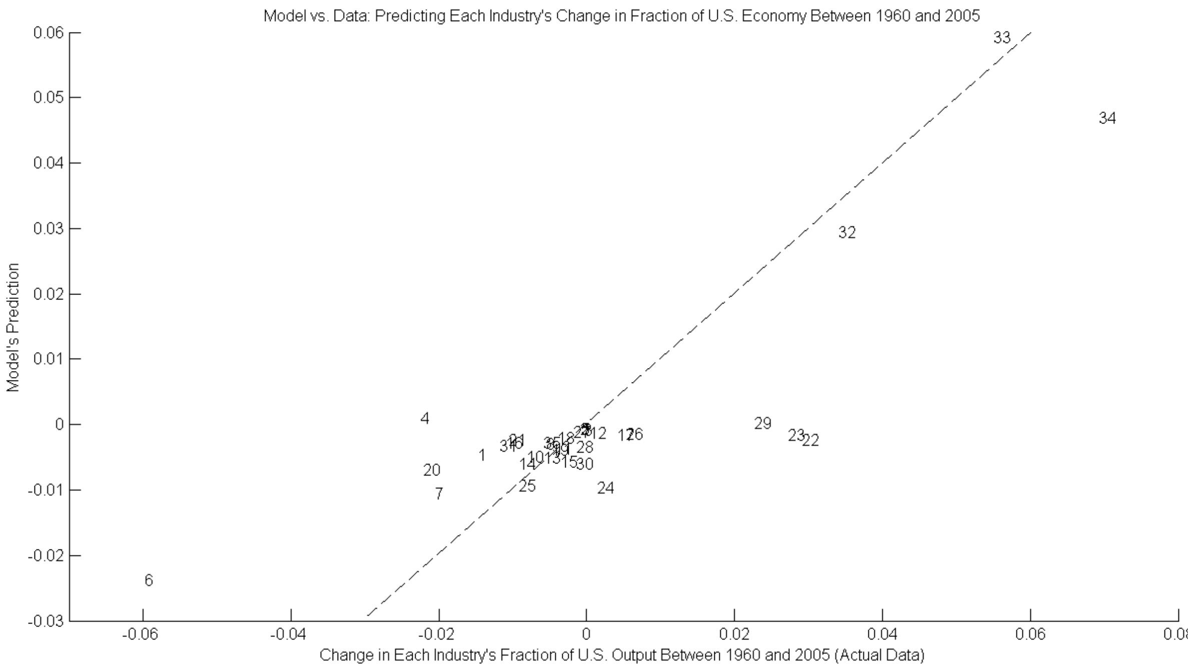
Model vs. Data: Predicting Each Industry's Fraction of U.S. Economy in 2005
Model's Prediction
Each Industry's Fraction of 2005 U.S. Output (Actual Data)
0.25
0.2
0.15
0.1
0.05
0
0
0.05
0.1
0.15
0.2
0.25
34
33
32
6
7
4
1
28
20
15
24
31
29
22
23

Figure 3:



Null Model vs. Data: Predicting Each Industry's Fraction of U.S. Economy in 2005
Null Model's Prediction
Each Industry's Fraction of 2005 U.S. Output (Actual Data)
0.25
0.2
0.15
0.1
0.05
0
0.05
0.1
0.15
0.2
0.25
34
6
32
33
7
1
28
20
4
15
24
21
30
31
13
19
16
29
22
23
26

Figure 4:



between the data and the model's predictions here is .84. This is "infinitely" better than the null model here, since the null model does not predict any changes between 1960 and 2005.

These results are suggestive evidence that the model is indeed getting at something. Further possible tests of the model are offered in sections 3.4 and 4.5.

3.4 Other sources of data

One can carry out the exercise in section 3.2 with other sources of data. One interesting direction to go in is to use data at a more disaggregated level. For example, in other work of mine that is available upon request, I use data from the BEA and BLS at roughly the three-digit NAICS industry level, in which the U.S. economy is broken into 69 industries. The results are similar to the results of section 3.3, except that even the null model performs extremely well, because the data only span from 2007 to 2012, during which time there was very little change in the U.S.'s industrial composition.[18]

[18] These BEA data go back to 1997, but many of the BLS price indices only go back to 2007, which is why my analysis of these data starts at 2007. If I use the data going back to 1997 and simply do not adjust for industry-specific price changes, the calibrated value of ρ ends up being nearly zero (specifically, 0.08), which is what we would expect, because when we do not adjust for industry-specific price changes, what we are effectively examining is changes in industries' *expenditure shares*, and expenditure shares do not change if the Cobb-Douglas assumption holds.

Given that Jorgenson's KLEMS data are ultimately derived from BEA and BLS data, I imagine it is possible to get more disaggregated price-adjusted data spanning a similar time frame as the KLEMS data, What these data sources have in common is that I have to proxy for different industries' usage of different occupations using industries' usage of intermediate goods. An alternative source of data is data directly on how many people each industry employs in different occupations, i.e., an industry-occupation table. This is what I use in the international section of the paper; in the future, for the sake of robustness, I plan to re-do the closed-economy quantitative exercise in section 3.2 using industry-occupation tables to see if the results change.

Lastly, the quantitative exercise in section 3.2 can be performed for any country that has the relevant data, not just the U.S. It would be particularly interesting to look at developing countries with data of sufficiently high quality.

4 Open-economy analysis

The closed-economy analysis of sections 2 and 3 extends easily into an international context, thanks to recent advances in trade theory.

The theory from section 2 involves technological progress through learning-by-doing; we thus want to extend this into a model in which countries trade with each other based on technological differences, i.e., a Ricardian model. It was not until Eaton and Kortum (2002) that this kind of model could be taken to data in a multi-country framework. Costinot, Donaldson, and Komunjer (2012) were the first to perform this kind of quantitative analysis in a way that allows for asymmetries across industries, which is what we are interested in.

In this section I combine the model from section 2 with the model from Costinot, Donaldson, and Komunjer (2012) (henceforth "CDK"), which is a static model of trade. At each date t, the model in this section is essentially the CDK model (the only difference is that there are multiple occupations, but this only matters for dynamics). The dynamics of the model, just as in section 2, are governed by a learning-by-doing equation − as before, this learning-by-doing is within each occupation, which spills over to everyone in the occupation regardless of the industry for which they are working, generating network effects among industries. The important thing to note here, which was a moot point in the single closed economy case, is that these spillovers are *within* countries, not across countries.

4.1 The open-economy model

As before, time is discrete and indexed by t.

There are now I countries, indexed by i. Country i is endowed exogenously with L_i workers; each worker in country i at date t is paid wage w_{it}, which will be determined in equilibrium.

As before, labor is the only factor of production; workers can work in K different occupations, indexed by k. Labor is perfectly mobile across occupations but immobile across countries.

and I am in the process of figuring out how.

As before, there are J final goods (i.e., industries $-$ the two terms are used interchangeably here), but now each final good j comes in a countably infinite number of varieties indexed by $\omega \in \Omega \equiv \{1, ..., +\infty\}$.[19]

The production structure of the economy is analogous to the closed-economy version of the model, except with the addition of total factor productivity terms, which will be discussed below. Specifically, the production function for variety ω of final good j in country i is as follows:

$$y_{ijt}(\omega) = z_{ij}(\omega) \min\{\frac{1}{\alpha_{jk}} x_{ijkt}(\omega)\}_k \tag{9}$$

where $y_{ijt}(\omega)$ is the quantity of variety ω of final good j produced in country i at date t; $x_{ijkt}(\omega)$ is the effective units of labor in occupation k used in the production of variety ω of final good j in country i at time t; $\alpha_{jk} > 0 \ \forall j, k$; and $z_{ij}(\omega)$ is the total factor productivity of variety ω of final good j in country i, to be discussed below.

The TFP term $z_{ij}(\omega)$ is a random variable drawn independently for each triplet (i, j, ω) from a Fréchet distribution $F_{ij}(\cdot)$ such that[20]

$$F_{ij}(z) = \exp[-(\frac{z}{z_{ij}})^{-\theta}]$$

for all $z \geq 0$, where $z_{ij} > 0 \ \forall i, j$ and $\theta > 1$. z_{ij} is the total factor productivity of country i in good j when averaged across good j's infinite varieties, while θ governs the dispersion of productivity, which is an important parameter in the Ricardian trade literature, because the more that productivity varies, the more important is the force of comparative advantage. Note that, for my purposes, these parameters are fixed over time; this will be discussed further below.

From (9) it follows that, in equilibrium,

$$x_{ijkt}(\omega) = \frac{\alpha_{jk}}{z_{ij}(\omega)} y_{ijt}(\omega) \tag{10}$$

Let ψ_{ikt} be the inverse of the productivity of occupation k in country i at time t. This will endogenously evolve over time from learning-by-doing (with starting values ψ_{ik0} given for all i and k), as will be discussed below.

Then, from (10) it follows that

[19] This is a modeling trick from the literature on Ricardian trade; the reason for it is as follows. If we did not have the infinite-variety structure, then for each good j and each country i, one hundred percent of country i's consumption of good j would be sourced by whichever country n can produce and deliver good j to country i most cheaply (or, in knife-edge cases in which two or more countries can do so equally cheaply, the solution is indeterminate). With the infinite-variety structure, this is exactly what happens at the variety level, but when we aggregate up to the good level we have interior solutions.

[20] This is another modeling trick from the Ricardian trade literature, first used by Eaton and Kortum (2002), who realized that if we want a tractable framework in which the distributions of labor requirements, costs of production, and prices are all in the same family, the distribution that gives us this is the Fréchet distribution.

$$\frac{\sum_k \alpha_{jk}\psi_{ikt}}{z_{ij}(\omega)}$$

is the number of units of raw labor needed to produce one unit of variety ω of final good j in country i at time t.

Now let us consider trade between countries. I will make the standard assumption of iceberg trade costs, meaning that for each unit of good j shipped from country i to country n, only $\frac{1}{d_{ijn}} < 1$ units arrive, with d_{ijn} such that $d_{iji} = 1 \ \forall i$ and $d_{ijn} \leq d_{ijl}d_{ljn}$ for any third country l.

It follows, then, that

$$c_{ijnt}(\omega) = \frac{d_{ijn}w_{it}\sum_k \alpha_{jk}\psi_{ikt}}{z_{ij}(\omega)}$$

is the cost of producing and delivering one unit of variety ω of good j from country i to country n at date t. Aggregating up to the good level, define c_{ijnt} as follows:

$$c_{ijnt} \equiv \frac{d_{ijn}w_{it}\sum_k \alpha_{jk}\psi_{ikt}}{z_{ij}} \tag{11}$$

Markets are perfectly competitive, and therefore the price $p_{jnt}(\omega)$ paid by buyers in country n for variety ω of good j at date t is

$$p_{jnt}(\omega) = \min_{1\leq i\leq I}[c_{ijnt}(\omega)]$$

and the set of varieties of good j that are exported by country i to country n at date t is given by

$$\Omega_{ijnt} \equiv \{\omega \in \Omega | c_{ijnt}(\omega) = \min_{1\leq m\leq I} c_{mjnt}(\omega)\}$$

Each country has a representative consumer whose utility function is a Cobb-Douglas function of the composite goods, where each composite good is a CES function of its infinite varieties. Let β_{ij} be country i's Cobb-Douglas exponent on good j, and let σ_{ij} be country i's elasticity of substitution among the infinite varieties of good j. (As in CDK (2012), I assume $0 \leq \beta_{ij} \leq 1$ and $\sigma_{ij} < 1 + \theta \ \forall i, j$.) Accordingly, define p_{ijt} as follows:

$$p_{ijt} \equiv [\sum_{\omega\in\Omega} p_{ijt}(\omega)^{1-\sigma_{ij}}]^{\frac{1}{1-\sigma_{ij}}}$$

Then, defining $e_{ijt}(\omega)$ as total expenditure by country i on variety ω of good j at date t, we have

$$e_{ijt}(\omega) = (\frac{p_{ijt}(\omega)}{p_{ijt}})^{1-\sigma_{ij}}\beta_{ij}w_{it}L_{it}$$

Furthermore, define e_{ijnt} as the value (in dollar terms) of total exports of good j from country i to country n at date t; that is,

$$e_{ijnt} \equiv \sum_{\omega \in \Omega_{ijnt}} e_{njt}(\omega)$$

Then we get the result that

$$e_{ijnt} = \frac{(c_{ijnt})^{-\theta}}{\sum_{m=1}^{I}(c_{mjnt})^{-\theta}} \beta_{nj} w_{nt} L_{nt} \tag{12}$$

The date-t equilibrium of the world economy is pinned down by a balanced trade condition. Let π_{ijnt} be country i's share of the world exports (in dollar terms) of good j to country n at date t; that is,

$$\pi_{ijnt} \equiv \frac{e_{ijnt}}{\sum_{m=1}^{I} e_{mjnt}}$$

Then, for a given wage vector $w_t = (w_{it})_i$,

$$Z_{it} = (\sum_{n=1}^{I}\sum_{j=1}^{J} \pi_{ijnt}\beta_{jn} w_{nt} L_{nt}) - w_{it} L_i$$

is the excess demand for country i's labor at date t. The equilibrium at date t is pinned down by specifying that $Z_{it} = 0$ for every country i.

As is typical in the Ricardian trade literature, there is no closed form solution for this date-t equilibrium, but it can be computed using an algorithm from Alvarez and Lucas (2007). The basic idea behind their algorithm is simple: start with an arbitrary guess for the equilibrium wage vector $w_t = (w_{it})_i$, calculate each country's excess demand for labor Z_{it}, and then raise the wage of any country i for which $Z_{it} > 0$ while lowering the wage of any country i for which $Z_{it} < 0$. Keep doing this, and, under regularity conditions discussed by Alvarez and Lucas, the algorithm will converge to a unique equilibrium wage (from which the equilibrium values of all other variables can be straightforwardly computed).

That completes the description of the economy at date t. The evolution of the economy from date t to date $t+1$ is given by the learning-by-doing equation, which will be given below. Let $\overline{x}_{ikt}$ be the total efficiency units of labor that country i uses in occupation k (across all varieties of all goods) at date t, i.e.,

$$\overline{x}_{ikt} = \sum_{j=1}^{J}\sum_{\omega \in \Omega} x_{ijkt}(\omega)$$

and let $\tilde{x}_{ikt}$ be one plus the share of country i's effective labor force that is employed in occupation k (across all industries) at date t, i.e.,

$$\tilde{x}_{ikt} \equiv 1 + \frac{\overline{x}_{ikt}}{\sum_m \overline{x}_{imt}}$$

The learning-by-doing equation is given by

$$\psi_{i,k,t+1} = \psi_{ikt}[\tilde{x}_{ikt}]^{-\rho} \tag{13}$$

where $\rho > 0$, with ψ_{ik0} given for every country i and occupation k.

Equation (13) is saying that the larger the fraction of country i's effective labor force working in occupation k at date t, the more that occupation k's productivity increases in country i from date t to date $t+1$.

This learning-by-doing is external to individual agents, and hence individual agents do not take the learning-by-doing equation into account in their decision-making.

4.2 Theoretical analysis of the open-economy model

In sections 2.3 and 2.4 we asked, in the context of the closed-economy version of the model, when we give an exogenous positive shock to production in a specific industry, what are the effects on every other industry? In this section we ask, using the open-economy version of the model, when we give an exogenous positive shock to production in a specific industry in a specific country, what are the effects on every other industry in every other country? Furthermore, what are the effects on each country's welfare?

As we will see below, the intuition for why there were network effects among industries in the closed-economy model carries over to the open-economy model. Now, however, thanks to international trade, the learning-by-doing induced in a country by extra production in an industry will affect other industries in that country not only through direct learning spillovers, but also through indirect general equilibrium effects; an increase in learning-by-doing in a country pushes up the country's equilibrium wage, which − all else being equal − makes industries in that country less competitive, and furthermore, consumers in all countries benefit from the fall in the costs of production (and hence prices) induced by learning-by-doing, not just the learning country.

In this section I will assume that trade costs are zero. Note that there is nothing important about the number zero; what is important is that the trade costs are symmetric across all countries and industries − as discussed in Alvarez and Lucas (2007), asymmetric trade costs make it impossible to get any analytical traction in a Ricardian trade model like this one. I will further assume that each country's labor force is the same size (normalized to 1), and each country's Cobb-Douglas utility parameter for each industry is the same (namely, $1/J$, where J is the number of industries), in order to keep expressions simple.

Under the above simplifying assumptions, the first-order approximation of country i's wage at date t is[21]

$$w_{it} \approx [\sum_j (\frac{z_{ij}}{\sum_k \alpha_{jk}\psi_{ikt}})^{\theta}]^{\frac{1}{1+\theta}} \tag{14}$$

Note that this is a weighted average of country i's date-t productivity in industry j across all j, as one would intuitively expect. This result trivially holds with exact equality when the productivity terms are symmetric across countries and industries, but it is only

[21]The derivation of this result is available upon request, along with all the other results of this section.

an approximation otherwise. The results in this section, which are only first-order approximations rather than exact equalities, are derived by plugging (14) into the equations of the model and then log-linearizing the resulting system of equations[22] − hence, this section is the open-economy analogue of section 2.4.

For the purposes of this section, let $\hat{y}_{ijt}$ denote the logarithm of production in country i in industry j at date t, and let $\hat{W}_{it}$ denote the logarithm of country i's welfare at date t. Further, let

$$\tilde{\alpha}_{ijk} \equiv \frac{\frac{\alpha_{jk}}{z_{ij}}}{\sum_j \frac{\alpha_{jk}}{z_{ij}}} - \frac{\frac{1}{z_{ij}}}{\sum_j \frac{1}{z_{ij}}} \tag{15}$$

which is the *relative* intensity with which industry j uses occupation k in country i.

4.2.1 Production analysis

Results (16) and (17) below answer the question, given a positive shock at date t to production in industry j in country i, what effect does this have on production at date $t+1$ in industry h in country m?

For any country i and any pair of industries j and h:

$$\frac{d\hat{y}_{i,h,t+1}}{d\hat{y}_{i,j,t}} \approx [1 + (\frac{I-1}{I})\theta]\rho \sum_k \alpha_{hk}\tilde{\alpha}_{ijk} - [(\frac{1}{J})(\frac{I-1}{I})\theta]\rho \sum_{j'}\sum_k \alpha_{j'k}\tilde{\alpha}_{ijk} \tag{16}$$

and for any pair of countries i and $m \neq i$ and any pair of industries j and h:

$$\frac{d\hat{y}_{m,h,t+1}}{d\hat{y}_{i,j,t}} \approx -[(\frac{1}{I})\theta]\rho \sum_k \alpha_{hk}\tilde{\alpha}_{ijk} + [(\frac{1}{J})(\frac{1}{I})\theta]\rho \sum_{j'}\sum_k \alpha_{j'k}\tilde{\alpha}_{ijk} \tag{17}$$

The intuition behind (16) and (17) is as follows. The increase in production in industry j in country i has a *direct effect* and an *indirect effect*.

The direct effect is as follows. For each occupation k, the extent to which an increase in production in industry j in country i corresponds to an increase in usage of occupation k relative to other occupations (and hence an increase in learning-by-doing in occupation k) is given by $\tilde{\alpha}_{ijk}$ (which, examining (15), can be positive or negative, since learning-by-doing is based on *relative* occupational usage). The extent to which this extra learning-by-doing in occupation k benefits industry h is given by α_{hk}. Hence, the size of learning spillovers between industries j and h is $\sum_k \alpha_{hk}\tilde{\alpha}_{ijk}$. If industries j and h are similar (dissimilar) enough to each other in their occupational usage, then $\sum_k \alpha_{hk}\tilde{\alpha}_{ijk}$ is greater (less) than zero, and the direct effect on industry h within country i is positive (negative), while it is negative (positive) in every other country, because in every other country industry h becomes relatively less (more) competitive compared to country i.

The direct effect is scaled by ρ, since ρ is the rate of learning-by-doing. Furthermore, the direct effect on each other country is scaled by $\frac{1}{I}$, where I is the number of countries,

[22] Again, the details are available upon request.

as well as θ, since θ is the trade elasticity. This is balanced by the fact that the direct effect on country i itself is scaled by $[1 + (\frac{I-1}{I})\theta]$; note that $[1 + (\frac{I-1}{I})\theta] - (I-1)(\frac{1}{I})\theta = 1$, i.e., the scale factors on the direct effects across the world sum to one.

The indirect effect on industry h is as follows. Industry h is, of course, not the only industry directly affected by industry j. Summing the term $\sum_k \alpha_{hk}\tilde{\alpha}_{ijk}$ across all industries gives us $\sum_{j'}\sum_k \alpha_{j'k}\tilde{\alpha}_{ijk}$, which is the size of the total learning spillovers from industry j to all other industries – or, using network terminology, it is the first-degree centrality of industry j in the network of industries. If industry j is sufficiently central (sufficiently peripheral), then $\sum_{j'}\sum_k \alpha_{j'k}\tilde{\alpha}_{ijk}$ is greater (less) than zero, and the high (low) amount of learning-by-doing induced by the increase in production in industry j in country i raises (lowers) country i's equilibrium wage, which (all else being equal) makes each industry in country i less (more) competitive and makes each industry in every other country more (less) competitive.

As with the direct effect, the indirect effect is scaled by ρ, since ρ is the rate of learning-by-doing. Moreover, the indirect effect (which, bear in mind, is capturing an individual industry's effect on the entire economy) is scaled by $\frac{1}{J}$, where J is the number of industries. As with the direct effect, the indirect effect on each other country is scaled by $\frac{1}{I}$, where I is the number of countries, as well as θ, since θ is the trade elasticity. In the case of the indirect effect, this is balanced by the fact that the indirect effect on country i itself is scaled by $(\frac{I-1}{I})\theta$; note that $-(\frac{I-1}{I})\theta + (I-1)(\frac{1}{I})\theta = 0$, i.e., the scale factors on the indirect effects across the world sum to zero.

Note that if we set $I = 1$ (i.e., if country i is the only country in the world), then the indirect effect is zero, and the total effect of the industry-j shock on industry h is $\rho\sum_k \alpha_{hk}\tilde{\alpha}_{ijk}$, which is exactly the same as the closed-economy results from section 2.4.[23]

4.2.2 Welfare analysis

Results (18) and (19) below answer the question, given the aforementioned positive shock at date t to production in industry j in country i. what effect does this have on the date $t+1$ welfare of country m?

For any country i and any industry j:

$$\frac{d\hat{W}_{i,t+1}}{d\hat{y}_{ijt}} \approx (\frac{1}{J})[1 - (\frac{I-1}{I})(\frac{1}{1+\theta})]\rho\sum_{j'}\sum_k \alpha_{j'k}\tilde{\alpha}_{ijk} \tag{18}$$

and for any pair of countries i and $m \neq i$ and any industry j:

$$\frac{d\hat{W}_{m,t+1}}{d\hat{y}_{ijt}} \approx (\frac{1}{J})[(\frac{1}{I})(\frac{1}{1+\theta})]\rho\sum_{j'}\sum_k \alpha_{j'k}\tilde{\alpha}_{ijk} \tag{19}$$

[23]There is a trivial difference, namely, the z terms, which by (15) are part of the $\tilde{\alpha}$ terms, made no appearance in the closed-economy results, but that is just because there were no z terms in the closed-economy model. It would be easy to add them in, in which case we would get exactly the same result as here.

The intuition behind (18) and (19) partly carries over from the intuition above for the indirect effects in Results (16) and (17) $-$ the effects of the date t shock to industry j on countries' welfare at date $t+1$ is a function of industry j's first-degree centrality $\sum_{j'} \sum_k \alpha_{j'k} \tilde{\alpha}_{ijk}$, and again this is scaled by ρ and $\frac{1}{J}$ (and by $\frac{1}{I}$ for countries other than i) for the same reasons as above.

Note, though, that the right-hand sides of (18) and (19) have the same sign rather than opposite signs; $\sum_{j'} \sum_k \alpha_{j'k} \tilde{\alpha}_{ijk}$ is greater (less) than zero when industry j is sufficiently central (peripheral) in the network that an increase in production in industry j in country i induces more (less) learning in the aggregate economy of country i, in which case other countries benefit (are hurt) as well, due to the opportunity to buy products from country i at a lower (higher) cost.

Furthermore, note that the effect on other countries is scaled by $\frac{1}{1+\theta}$ rather than θ; a higher θ dampens the effect on other countries rather than exacerbating it $-$ a higher θ means less heterogeneity in intra-industry productivity, meaning (all else being equal) international trade is less important for a country's welfare, meaning extra economy-wide learning in country i benefits other countries less. (This is in contrast with Results (16) and (17), which were looking at the effects on a specific industry h, which are exacerbated when intra-industry productivity varies less.)

Given that the effect on other countries is scaled by $(\frac{1}{I})(\frac{1}{1+\theta})$, this is balanced by the effect on country i itself being scaled by $[1-(\frac{I-1}{I})(\frac{1}{1+\theta})]$; note that $[1-(\frac{I-1}{I})(\frac{1}{1+\theta})]+(I-1)(\frac{1}{I})(\frac{1}{1+\theta})=1$, i.e., the scale effects on welfare across the world sum to one.

4.2.3 Analyzing more than one period ahead

Results (16) through (19) are only telling us the effects at date $t+1$; now let us consider the effects arbitrarily far into the future. First we will consider the effects over time on production in each industry in each country. Let A_{im} denote the J X J matrix whose (j,h) element is $\frac{d\hat{y}_{m,h,t+1}}{d\hat{y}_{ijt}}$ (which we found an approximation for above, which does not depend on t). Let A be the (IJ) X (IJ) matrix formed by appending the A_{im} matrices to each other, so that the (i,m) block of A is A_{im}.

Start from an arbitrary equilibrium path $\{\hat{y}^{\star}_{ijt}\}_{i,j,t}$ and consider an arbitrary vector of shocks to production in each industry in each country at date t: let y_t be the (IJ)-dimensional vector whose first J elements are $\hat{y}_{1jt}-\hat{y}^{\star}_{1jt}$ for each industry j in country 1; the next J elements of y_t are $\hat{y}_{2jt}-\hat{y}^{\star}_{2jt}$ for each industry j in country 2; and so on.

If we take the first-order approximations that we found above and suppose that they approximately hold at any arbitrary point, then we have the result that for any date t and any length of time τ beyond t:

$$y_{t+\tau} \approx (A')^{\tau} y_t \tag{20}$$

While the closed-economy model involved a network of industries, (20) is saying that we can think of the open-economy model as involving a network of countries and industries, where each node in the network is a country-industry pair, and the network matrix A

(whose entries we found above) gives us the effect of an increase in production in industry j in country i on every other industry in every other country, with these effects being the aforementioned sum of direct learning spillovers and general equilibrium effects via international trade.[24]

Now let us consider the effects on each country's discounted sum of welfare, summing from date t to ∞. Let w_m be the (IJ)-dimensional vector whose first J elements are $\frac{d\hat{W}_{m,t+1}}{d\hat{y}_{1jt}}$ for each industry j in country 1 (which we found an approximation for above, which does not depend on t), whose second J elements are $\frac{d\hat{W}_{m,t+1}}{d\hat{y}_{2jt}}$ for each industry j in country 2, and so on. Let $\bar{W}_m$ denote the discounted sum of country m's logarithm of welfare over time, discounted at the rate δ – that is, $\bar{W}_m \equiv \sum_{t=0}^{\infty} \delta^t \hat{W}_{mt}$. Note, then, that $\frac{d\overline{W}_m}{dy_t}$ is the (IJ)-dimensional vector whose first J entries are $\frac{d\overline{W}_m}{d\hat{y}_{1jt}}$ for each industry j in country 1, whose second J entries are $\frac{d\overline{W}_m}{d\hat{y}_{2jt}}$ for each industry j in country 2, and so on. Then, for any arbitrary country m, we have the following result:

$$\begin{aligned} \frac{d\overline{W}_m}{dy_t} &\approx \delta w_m + \delta^2 A w_m + \delta^3 A^2 w_m + ... \\ &= \delta(I + \delta A + \delta^2 A^2 + ...)w_m \\ &= \delta(I - \delta A)^{-1} w_m \end{aligned}$$

So we have, for any arbitrary country m:

$$\frac{d\overline{W}_m}{dy_t} \approx \delta(I - \delta A)^{-1} w_m \tag{21}$$

The intuition behind (21) is as follows. The left-hand side is the vector telling us how much a shock at date t to each country-industry pair affects country m's discounted infinite sum of welfare from date t onward. The right-hand side is the vector of each country-industry pair's *Bonacich centrality* (from country m's perspective) in the network of country-industry pairs, defined as follows. Recall from section 2.4 that for any arbitrary network, a node's Bonacich centrality is equal to the sum of its first-degree links with every other node discounted by δ, plus the sum of its second-degree links with other nodes discounted by δ^2, and so on. In this case the links are weighted by the vector w_m, which tells

[24]This idea of a "network of countries and industries" is, on the one hand, fairly intuitive, but on the other hand, I think it would be nice if I could somehow re-interpret these results as simply a network of industries, where countries have different "positions" in the network based on which industries they specialize in. I think there would be quite a bit of value added in pinning that down, especially for purposes of visualization and intuition.

Furthermore, if I can precisely pin down the notion of a country's "position" in the network of industries, then an interesting direction to go in would be to try to characterize certain positions in the network as more "advantageous" than other positions, perhaps using the model to somehow pin down a sufficient statistic for a country's future economic growth, as a function only of its initial position in the network of industries.

Table 2: The 37 countries

ID #	Country	ID #	Country
1	Morocco	19	Malaysia
2	Tunisia	20	Pakistan
3	Nigeria	21	Philippines
4	Canada	22	Singapore
5	USA	23	Thailand
6	Argentina	24	Taiwan
7	Brazil	25	Belgium and Luxembourg
8	Chile	26	Denmark
9	Colombia	27	France and Monaco
10	Ecuador	28	Greece
11	Mexico	29	Ireland
12	Peru	30	Italy
13	Venezuela	31	Netherlands
14	Israel	32	Portugal
15	Japan	33	Spain
16	Turkey	34	UK
17	China, Hong Kong, and S.A.R.'s	35	Norway
18	South Korea	36	Sweden
		37	Switzerland and Lichtenstein

us how much a shock to a given country-industry pair in a given time period affects country m's next-period welfare – which, as we found above, relates to each country-industry pair's first-degree centrality in the network.

4.3 International trade data

I take the model from Section 4.1 to international trade data from Feenstra et al (2005). The data report bilateral exports among 72 countries, at the 4-digit SITC Revision 2 product level, annually over the years 1962-2000. This means that a typical observation is that, in 1978, Japan exported to Italy $2,447,000 (in 1978 nominal US dollars) worth of silk worm cocoons and silk waste (SITC Rev. 2 code 2614). 696 different products appear in the 1962 data, and 1288 in the 2000 data. These data include values reported by importing countries as well as values reported by exporting countries; I use the values reported by importing countries, which is standard practice, since importing countries are seen as having more of an incentive to carefully keep track of goods crossing borders.

In order to merge these data with the industry-occupation table described below, I reclassify exports from 4-digit SITC Rev. 2 product codes into 3-digit 1997 NAICS industry codes, using a concordance table from Feenstra and Lipsey (n.d.). I restrict the sample to industries that appear in the industry-occupation table, and I further restrict the sample to the importer-exporter-industry triplets that appear in both 1962 and 2000. We are left, then, with 37 countries and 28 industries, which are listed in Tables 2 and 3, respectively.

Table 3: The 28 industries

NAICS	Industry	NAICS	Industry
113	Forestry and Logging	325	Chemical Manufacturing
211	Oil and Gas Extraction	326	Plastics and Rubber Products
212	Mining (except Oil and Gas)	327	Nonmetallic Mineral Products
221	Utilities	331	Primary Metal Manufacturing
311	Food Manufacturing	332	Fabricated Metal Products
312	Beverages and Tobacco Products	333	Machinery
313	Textile Mills	334	Computers and Electronic Products
314	Textile Product Mills	335	Electrical Equipment and Appliances
315	Apparel Manufacturing	336	Transportation Equipment
316	Leather and Allied Products	339	Miscellaneous Manufacturing
321	Wood Products	483	Water Transportation
322	Paper Manufacturing	512	Motion Pictures and Sound Recording
323	Printing and Related Activities	541	Professional and Scientific Services
324	Petroleum and Coal Products	562	Waste Management

4.4 Calibrating the open-economy model

Recall from the model that α_{jk} is the intensity with which industry j uses occupation k. I calibrate these parameters (which, by assumption, are fixed over time and across countries) using the May 2013 industry-occupation table from the Occupational Employment Statistics (OES) program at the Bureau of Labor Statistics (BLS) in the US. For each industry j and occupation k, I set α_{jk} equal to the number of people in occupation k that industry j employs, divided by the total number of people that industry j employs. The industries are at the 3-digit NAICS level and are reported above in Table 3. The occupations are at the 2-digit SOC level and are reported in Table 4.

Further recall from the model that β_j is industry j's exponent in the Cobb-Douglas utility function. I calibrate these parameters (which, as with the alpha parameters, I assume in the model to be fixed over time and across countries) using the 2012 Use Table from the Bureau of Economic Analysis (BEA) in the US. For each industry j, I set β_j equal to personal consumption expenditures on industry j divided by total personal consumption expenditures. The industries are at the 3-digit NAICS level and are reported in Table 3.

To calibrate the size of each country's labor force (which is fixed over time by assumption), I use the 2012 World Development Indicators (WDI) from the World Bank.[25] For each country i, I set L_i equal to the number of people in country i's labor force in 2012.

There is a sizeable literature on how to calibrate trade costs in this kind of model, but for now, for simplicity, I only consider the case of free (i.e., costless) trade − that is, for each exporter i, industry j, and importer n, I set d_{ijn} equal to 1, meaning that trade costs are zero.

Recall that the parameter θ governs the dispersion of total factor productivity and hence the strength of the force of comparative advantage. There is a substantial literature

[25]Taiwan is not included in the World Development Indicators, so for Taiwan I use the 2012 Monthly Bulletin of Statistics of the Republic of China.

Table 4: The 22 occupations

SOC code	Occupation
11	Management Occupations
13	Business and Financial Operations Occupations
15	Computer and Mathematical Occupations
17	Architecture and Engineering Occupations
19	Life, Physical, and Social Science Occupations
21	Community and Social Services Occupations
23	Legal Occupations
25	Education, Training, and Library Occupations
27	Arts, Design, Entertainment, Sports, and Media Occupations
29	Healthcare Practitioners and Technical Occupations
31	Healthcare Support Occupations
33	Protective Service Occupations
35	Food Preparation and Serving Related Occupations
37	Building and Grounds Cleaning and Maintenance Occupations
39	Personal Care and Service Occupations
41	Sales and Related Occupations
43	Office and Administrative Support Occupations
45	Farming, Fishing, and Forestry Occupations
47	Construction and Extraction Occupations
49	Installation, Maintenance, and Repair Occupations
51	Production Occupations
53	Transportation and Material Moving Occupations

on calibrating θ; I borrow the value 6.53 from Costinot et al (2012).

Now consider the industrial productivity parameters. Recall that z_{ij} is the average productivity of country i in industry j. I calibrate these parameters using the *revealed productivity* method of Costinot et al (2012). Using the 1962 trade data, I run the following regression:

$$\ln e_{ijn} = \delta_{in} + \delta_{nj} + \delta_{ij} + \varepsilon_{ijn} \tag{22}$$

where e_{ijn} is the value of exports in industry j from country i to country n, and δ_{in}, δ_{nj}, δ_{ij} are exporter-importer, importer-industry, and exporter-industry fixed effects, respectively.[26]

Note that according to the model[27],

$$\ln e_{ijn} = \delta_{in} + \delta_{nj} + \theta \ln z_{ij} + \varepsilon_{ijn}$$

Combining the above two equations, we have

[26]This is a computationally demanding regression, and so I do this at the 2-digit level. Thus, for any country i and any industries j and m, industries j and m are given the same calibrated values of z_{ij} and z_{im} if those industries are within the same 2-digit NAICS group of industries.

[27]This can be seen by plugging equation (11) into equation (12) and noting that at the initial date 1962, ψ_{ik0} is normalized to 1 for every country i and every occupation k, and $\sum_k \alpha_{jk} = 1$ for every industry j.

$$z_{ij} = e^{\frac{\delta_{ij}}{\theta}} \tag{23}$$

Thus, running regression (22) and plugging the exporter-industry fixed effects into equation (23) gives us revealed measures of the productivity terms z_{ij}.

The last parameters to consider are the learning-by-doing parameters. For each occupation k and country i, I normalize the inverse of the initial productivity of occupation k in country i (ψ_{ik0}) to 1. Finally, we are left with ρ, the rate of learning-by-doing in equation (13), which is the only parameter of the model for which there is no immediately obvious way of calibrating. I grid-search over the values -1.00, -0.99, ..., 0.00, 0.01, ... , 0.99, 1.00 to find the value of ρ that best fits the data, as described in the next section.

4.5 Open-economy quantitative exercise

To test the international model against the international trade data described in section 4.3, I first calibrate the parameters of the model as described in section 4.4 and then perform the following quantitative exercise.

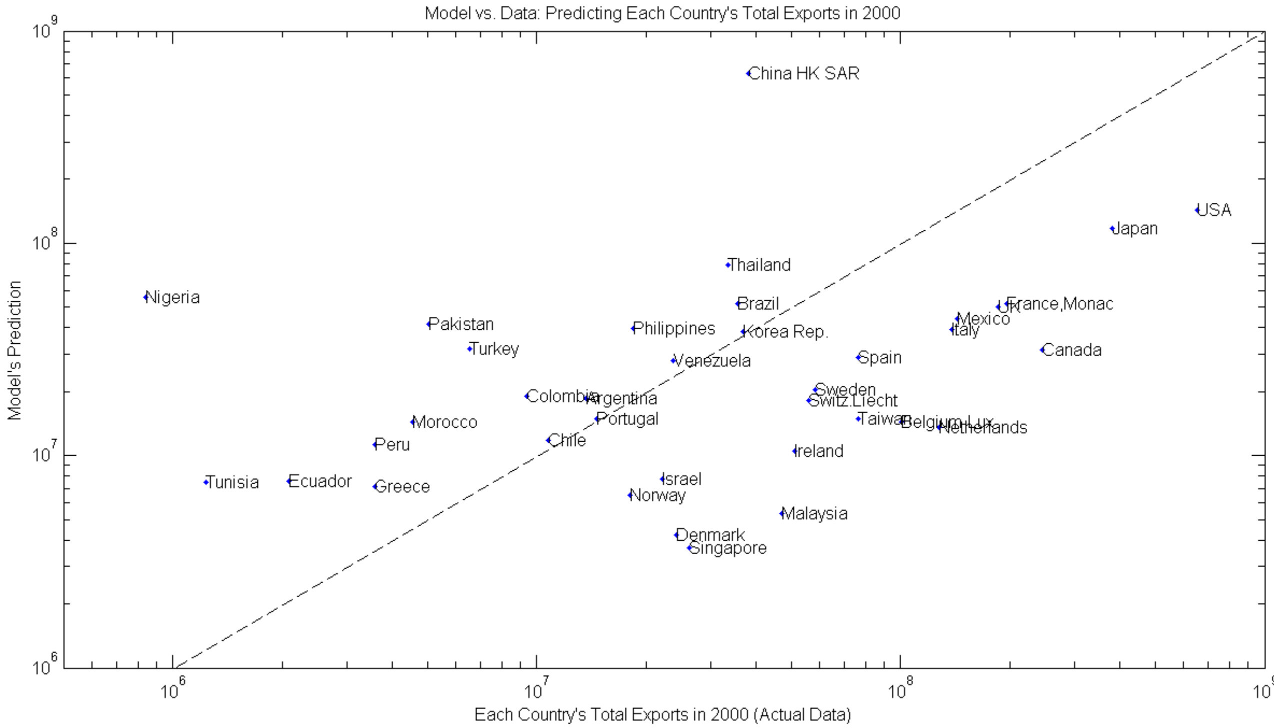
For each value of $\rho \in \{0.0, 0.1, ..., 9.9, 10.0\}$, I run the model for 39 periods, representing 1962 to 2000,[28] collect the model's predictions for each country's total exports in 2000, and calculate the correlation coefficient (the "goodness of fit") between these predictions and the actual 2000 data. I then choose the value of ρ that best fits the data, and I compare this goodness of fit with the goodness of fit of the null model (in which $\rho = 0$ and thus there there is no learning-by-doing and hence no network effects).

The resulting best-fit value of ρ is 2.8, meaning that an increase of 1% in the share of a country's effective labor force in a particular occupation at a particular year corresponds to a 2.8% increase in that country's productivity in that occupation the next year. Figure 5 plots the predictions of the model (with ρ set to its best-fit value of 2.8) for the total exports of each country in 2000 against the actual 2000 data. The goodness of fit of the model to the data corresponds to how close the points in Figure 5 are to the 45-degree line. The correlation coefficient between the model's predictions and the actual data is .181, whereas for the null model it is .087, so the model outperforms the null model by more than two-fold.

The model thus does a good job of predicting relative growth in different countries' total exports. Since the model actually makes predictions for exports at date t in industry j from country i to country n, for every i, n, j, and t, there are many additional ways to evaluate the model, which I intend to explore in future drafts of this paper.

[28]Note that, by normalizing the initial occupational productivity parameters $\{\psi_{ik0}\}$ all to 1, and by calibrating the industrial productivity parameters $\{z_{ij}\}$ using 1962 data, I am essentially calibrating the model to best match 1962 trade patterns, and then running the model for 39 periods to see how well the dynamics of the model (coming from the learning-by-doing equation) predict the evolution of world trade patterns from 1962 to 2000.

Figure 5: Model vs. Data: Predicting Each Country's Total Exports in 2000



5 Conclusions

This paper gives us, to my knowledge, the first model that offers an explanation for how different industries contribute differently to long-run economic growth and that can be taken directly to data. I show that the model helps explain the evolution of the industrial composition of the U.S., as well as growth in different countries' total exports, over the last half-century. When tested against such data, the model performs significantly better than the null model in which the strength of learning spillovers across industries is set to zero.

In addition to a variety of possible robustness checks, future steps include incorporating input-output linkages, seeking more reduced-form evidence that supports the broad ideas of the theory without hinging on the theory's structural assumptions, and microfounding the learning-by-doing equation that lies at the heart of the model.

References

[1] Acemoglu, D. (2008). "Learning-by-Doing, Trade, and Growth." Section 19.7 of *Introduction to Modern Economic Growth*, Princeton University Press.

[2] Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). "The network origins of aggregate fluctuations." *Econometrica* 80 (5), 19772016.

[3] Alvarez and Lucas (2007), "General equilibrium analysis of the Eaton-Kortum model of international trade." *Journal of Monetary Economics*, Vol. 54, pp. 17261768.

[4] Arrow, Kenneth (1962). "The Economic Implications of Learning by Doing." *The Review of Economic Studies*, Vol. 29, No. 3, pp. 155-173.

[5] Autor, Dorn, and Hanson (2013), "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6).

[6] Bardhan (1971). "On Optimum Subsidy to a Learning Industry: An Aspect of the Theory of Infant-Industry Protection." *International Economic Review*, Vol. 12, No. 1, pp. 54-70.

[7] Carvalho, V. M. (2010). "Aggregate fluctuations and the network structure of intersectoral trade." Mimeo, CREI, Barcelona.

[8] Cavalcanti and Giannitsarou (2013). "Growth and Human Capital: A Network Approach." Working paper.

[9] Clemhout, S. and H. Y. Wan, Jr. (1970). "Learning-by-Doing and Infant Industry Protection." *The Review of Economic Studies*, Vol. 37, No. 1, pp. 33-56.

[10] Costinot, Donaldson, and Komunjer (2012), "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies*, Vol. 79, pp. 581-608.

[11] Dehejia, Vivek (1993). "Optimal Endogenous Growth in a Two-Sector Model with Learning-by-Doing and Spillovers." *Journal of Economic Integration*.

[12] Eaton and Kortum (2002), "Technology, Geography, and Trade." *Econometrica*, Vol. 70, No. 5.

[13] Feenstra and Lipsey (n.d.). "Concordance between SITC Rev 2 at the 4-digit level and NAICS (1997)." Downloaded from < http://www.nber.org/lipsey/sitc22naics97/ >.

[14] Feenstra, Lipsey, Deng, Ma, and Mo (2005). "World Trade Flows: 1962-2000." NBER Working Paper 11040.

[15] Fogli, A. and L. Veldkamp (2012). "Germs, social networks and growth." NBER Working Paper 18470.

[16] Hamilton, Alexander (1791). "Report on Manufactures." Communicated to the United States House of Representatives on December 5, 1791.

[17] Hausmann, Hwang, and Rodrik (2007), "What you export matters," *Journal of Economic Growth* 12:125.

[18] Hidalgo, C. A.; Klinger, B.; Barabasi, A. L.; and Hausmann, R (2007). "The Product Space Conditions the Development of Nations." *Science*. Vol. 317, Iss. 482.

[19] Hirschman, Albert O. (1958). *The Strategy of Economic Development*. Yale University Press: New Haven, CT.

[20] Jones, Charles (2011). "Misallocation, Economic Growth, and Input-Output Economics." NBER Working Paper 16742.

[21] Jorgenson, Dale (2007), "35 Sector KLEM." Downloaded from ¡ http://scholar.harvard.edu/jorgenson/data ¿

[22] Long, John; and Plosser, Charles (1983), "Real Business Cycles." *Journal of Political Economy*, vol. 91, no. 1.

[23] Matsuyama, Kiminori (1992). "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory*. Vol. 58, Iss. 2.

[24] Matsuyama, Kiminori (2008). "Structural Change." From *The New Palgrave Dictionary of Economics*, Second Edition, edited by Steven Durlauf and Lawrence Blume.

[25] Ramey, Valerie (2011). "Identifying Government Spending Shocks: It's all in the Timing." *Quarterly Journal of Economics*, Vol. 126 (1), pp. 1-50.

[26] Ramey, Valerie; and Matthew Shapiro (1998). "Costly capital reallocation and the effects of government spending." *Carnegie-Rochester Conference Series on Public Policy*, Vol. 48, pp. 145-194.

[27] Rodriguez-Clare, Andres (1996a). "Multinationals, Linkages and Economic Development." *American Economic Review*. Vol. 86, No. 4.

[28] Rodriguez-Clare, Andres (1996b). "The Division of Labor and Economic Development." *Journal of Development Economics*. Vol. 49, p. 3-32.

[29] Succar, Patricia (1987). "The Need For Industrial Policy In LDC's - A Re-statement Of The Infant Industry Argument." *International Economic Review*.

[30] Young (1991). "Learning by Doing and the Dynamic Effects of International Trade." *The Quarterly Journal of Economics*, Vol. 106, No. 2, pp. 369-405.