No. 8495

A MANY-COUNTRY MODEL OF INDUSTRIALIZATION

Holger Breinlich and Alejandro Cuñat

INTERNATIONAL TRADE AND
REGIONAL ECONOMICS

ISSN 0265-8003

A MANY-COUNTRY MODEL OF INDUSTRIALIZATION

Holger Breinlich, Universität Mannheim, CEP and CEPR
Alejandro Cuñat, Universitat Wien

Discussion Paper No. 8495
July 2011

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 8495

July 2011

ABSTRACT

A Many-Country Model of Industrialization*

We draw attention to the role of economic geography in explaining important cross-sectional facts which are difficult to account for in existing models of industrialization. By construction, closed-economy models that stress the role of local demand in generating sufficient expenditure on manufacturing goods are not suited to explain the strong and negative correlation between distance to the world's main markets and levels of manufacturing activity in the developing world. Secondly, open-economy models that emphasize the importance of comparative advantage are at odds with a positive correlation between the ratio of agricultural to manufacturing productivity and shares of manufacturing in GDP. This paper provides a potential explanation for these puzzles by nesting the above theories in a multi-location model with trade costs. Using a number of simple analytical examples and a full-scale multi-country calibration, we show that the model can replicate the above stylized facts.

JEL Classification: F11, F12, F14 and O14
Keywords: economic geography, industrialization and international trade

Holger Breinlich
Universität Mannheim
Mannheim
GERMANY

Email: h.breinlich@uni-mannheim.de

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=160700

Alejandro Cuñat
Department of Economics
University of Vienna
Hohenstaufengasse 9
A-1010 Vienna
AUSTRIA

Email: alejandro.cunat@univie.ac.at

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=147859

* This paper is partly based on the unpublished 2005 paper “Economic Geography and Industrialization” which was chapter 2 of Breinlich’s PhD dissertation. We are grateful to Harald Fadinger, Gabriel Felbermayr and seminar participants in Mannheim, Munich and Vienna for helpful suggestions. Stephen Redding, Anthony Venables and Silvana Tenreyro provided very useful comments on the earlier PhD chapter. All remaining errors are ours.

Submitted 11 July 2011

1 Introduction

One of the most striking aspects of economic development is the decline of agriculture's share in GDP and the corresponding rise of manufacturing and services. Economists have proposed a number of theories to explain this transformation. Among the most influential approaches are explanations that focus on differences in the income elasticity of demand across sectors (*e.g.*, Rosenstein-Rodan (1943); Murphy *et al.* (1989b); Kongsamut *et al.* (2001)). As *per-capita* income increases, non-homothetic preferences lead to a shift of demand from agriculture to manufacturing goods, thus increasing manufacturing's share in GDP. Traditionally, these approaches have analyzed closed-economy models and stressed the role of local demand. More recently, authors such as Matsuyama (1992, 2009) have provided extensions to open-economy settings and have shown that some key results of the closed-economy literature, such as the positive impact of increased agricultural productivity on industrialization, can be reversed in such models.

The present paper draws attention to two cross-sectional facts which, taken together, are not easily explained by either closed-economy or open-economy models of demand-driven industrialization. We argue that to understand these facts we need to move beyond the closed-*versus*-open-economy dichotomy prevalent in the literature, and to consider multi-country settings in which countries interact with each other through international trade, but in which bilateral interactions are partly hampered (to a different extent across country pairs) by the fact that trade is not costless.

Our first observation is that proximity to foreign sources of demand seems to matter for industrialization. For example, it has long been noted that Hong Kong, Singapore, and Taiwan not only benefitted from an outward-oriented trade policy but also close proximity to the large Japanese market (*e.g.*, Puga and Venables (1996)). A cursory look at the data suggests that distance to foreign markets has a more general relevance: Figure 1 plots the manufacturing share in GDP against the minimum distance to the European Union, Japan and the U.S. for a cross-section of developing countries in 2000.[1] The figure shows that developing economies close to one of these main markets of the world show proportionally higher levels of industrialization.

Whereas this first fact suggests that interactions between economies are important, and thus points to the relevance of open-economy models, our second fact seems to suggest the opposite: Figure 2 plots manufacturing shares against a standard proxy for comparative advantage in agriculture, labor productivity in agriculture relative to manufacturing, for a cross-section of developing countries for the year 2000.[2] The fitted line has a positive, albeit statistically insignificant slope. As we show in our more detailed econometric analysis in Section 2, extending the sample to include more countries and years leaves this positive correlation intact and actually makes it statistically significant as well. This is of course puzzling for open-economy theories of

---

[1] We use the Netherlands as the approximate geographic centre of the European Union in Figure 1. Developing countries are defined as countries belonging to the income categories "low", "lower middle" and "upper middle" published by the World Bank (corresponding to less than 9,265 USD in 1999). The simple OLS regression underlying the fitted line in Figure 1 yields a negative slope coefficient which is statistically significant at the 1% level.

[2] Developing countries are defined as in footnote 1. Relative productivy in agriculture *vs.* manufacturing is the proxy of choice in many studies of Ricardian comparative advantage, *e.g.* Golub and Hsieh (2000).

industrialization such as Matsuyama (1992). If countries are indeed integrated through trade, should we not expect them to specialize according to their comparative advantages?

We argue that both facts can be understood in a standard model of industrialization in which there are differences in the income elasticity of demand across sectors and in which comparative advantage forces are present and active. The key difference of our approach in comparison with existing closed-economy approaches is that we allow for a setting with many countries which are integrated through trade. But crucially, and in contrast to open-economy models such as Matsuyama (1992), trade is not costless and geographic position is therefore important.

In our model, developing countries closer to foreign sources of demand will experience higher demand for both the agricultural and manufacturing goods they produce than more distant countries, ceteris paribus. We outline conditions under which this translates into higher manufacturing shares in GDP. Most importantly, higher overall demand will lead to higher wages which, in the presence of non-homotheticity in demand combined with positive trade costs, will shift local production towards the manufacturing sector. Trade costs for agricultural products also hamper the comparative-advantage mechanism put forward by free-trade models. High agricultural productivity leads to higher wages which, again because of the combination of agricultural trade costs and non-homothetic demand, leads countries to specialize in manufacturing (we call this the "relative-demand effect" of agricultural productivity). The standard comparative-advantage effect, which would drive specialization patterns in the opposite direction, is also present but can be overcompensated by the relative-demand effect for intermediate levels of trade costs.

Given that our model nests free trade and autarky as special cases and that its predictions vary depending on the level of trade costs and other parameters (such as the degree of non-homotheticity of preferences), we complement our theoretical analysis with a full-scale multi-country calibration. That is, we ask to what extent our model matches the above stylized facts for empirically plausible parameter values. We choose parameters to match international trade and expenditure data and demonstrate that this calibrated model generates the same positive correlation observed in the data between access to markets and comparative advantage in agriculture, on the one hand, and manufacturing shares on the other hand. Crucially, this is not true when we constrain our trade cost estimates to zero (free trade) or infinity (autarky). Interestingly, allowing for positive but finite levels of trade costs also improves the predictive power (in terms of matching observed and predicted shares) as opposed to autarky and free trade.

Our paper relates to at least three sets of contributions in the literature. In terms of the questions addressed, we contribute most directly to the literature on industrialization that relies on differences in the income elasticity of demand across sectors for explaining structural change ("demand-driven industrialization"). We add to this literature by drawing attention to the role of economic geography in shaping cross-sectional patterns of industrialization. Similar to papers such as Murphy *et al.* (1989a, b), Matsuyama (1992) or Laitner (2000) we focus on the initial shift from agriculture to manufacturing, which is the key transition for the group of countries we are interested in in this paper, *i.e.* low- to middle-income countries. That is, for most of the paper we do not model the services sector, which rises with income *per capita* at all levels of economic development, and which is not subject to open-economy analytical treatments due to its non-tradability. However, as we show in our robustness checks, explicitly modelling a non-tradable services sector leaves our results unchanged.

Consistent with our focus on explaining cross-sectional facts, we also disregard the dynamic aspects of the industrialization process and rely on an entirely static model. In this respect, we are similar to the contributions by Murphy *et al.* (1989a, b) and Matsuyama (2009) but different from most other papers in the industrialization literature. Our approach, however, avoids the criticism by Ventura (1997) and Matsuyama (2009), among others, of closed-economy dynamic approaches to issues such as industrialization; namely that explaining cross-country patterns taking place in a globalized world on the basis of dynamic closed-economy arguments can be quite misleading. In fact, we extend this methodological criticism to the standard two-country, free-trade way of thinking about trade and development: once one recognizes that bilateral distances and geographic position matter, one must extend the model to many countries and allow for differences in bilateral trade costs.

While our criticisms apply most directly to theories that approach the phenomenon of industrialization from the demand side, the stylized facts we have presented are not easily explained by models that focus on supply-side determinants either, be it contributions from the barriers-to-modern-growth literature (*e.g.,* Parente and Prescott (1994) and (2000), Goodfriend and McDermott (1995), or Galor and Weil (2000)), attempts to reconcile balanced (neoclassical) growth or convergence with structural transformations (*e.g.*, Caselli and Coleman (2001) or Ngai and Pissarides (2007)) or approaches from traditional international trade theory (*e.g.*, Leamer (1987) and Schott (2003)). The ageographical nature of these approaches means that they are not well suited to explain phenomena that have an inherent geographic component, such as the ones described above. Thus, while not denying the importance of these models and theories for many aspects of industrialization, this paper draws attention to geographical proximity as a new and potentially important factor in explaining the dramatic differences in levels of industrialization across the world.[3]

Methodologically, our paper is most closely related to work in international trade and economic geography which is interested in the effects of comparative advantage and relative location on trade flows, wages, and production structures (*e.g.*, Krugman (1980), Puga and Venables (1999), Golub and Hsieh (2000), Davis and Weinstein (2003), or Redding and Venables (2004), to name but a few). To the best of our knowledge, however, the insights from this literature have never been applied to the aforementioned stylized facts, nor to the modelling of cross-sectional patterns in levels of industrialization more generally. Some of our results are also relevant for the international trade literature beyond our immediate focus on industrialization. For example, the role of trade costs in modifying the impact of comparative advantage on production structures

---

[3] In a recent working paper, Yi and Zhang (2010) share our concern that many aspects of industrialization cannot be analyzed neither within a closed-economy setting nor under free trade. They analyze the effects of changes in productivity and declining trade barriers on production structures within a three-sector, two-country model, but focus on dynamic rather than cross-sectional aspects of industrialization. Restricting their analysis to a two-country setting also prevents them from adequately modelling economic geography. (For this, at least three countries are needed as will become clear in Section 3).

has been mostly ignored in the literature, although our results suggest that models based on a free-trade assumption may have poor explanatory power.[4] Theoretically, we contribute to the home-market effect literature by outlining conditions under which more central locations specialize in manufacturing once we leave the standard setting of monopolistic competition and factor price equalization.[5]

The remainder of this paper is organized as follows. Section 2 shows that the two correlations displayed in Figures 1 and 2 also survive in a broader cross-section of countries, and are robust to the inclusion of proxies for local demand and other domestic factors. Section 3 develops a multi-country model with trade costs. This model is used in Section 4 to shed light on the puzzles raised in the introduction. In Section 5 we show that a fully calibrated version of the model is able to replicate our stylized facts. Finally, section 6 concludes.

2 Empirical Evidence

In this section, we examine the robustness of the correlations from the introduction through variations in sample composition and by including a number of control variables.[6] Our full econometric specification will be

$$ltShareM_{lt} = \alpha + d_t + \beta_1 RP_{lt} + \beta_2 CEN_{lt} + \beta_3 AP_{lt} + \beta_4 POP_{lt} + \varepsilon_{lt}, \tag{1}$$

where RP_{lt} is relative productivity (of agriculture to manufacturing) and CEN_{lt} the 'centrality' of country l, *i.e.*, its access to foreign markets (to be defined below). AP_{lt} denotes agricultural productivity, POP_{lt} the population size of country l, and d_t is a full set of year fixed effects. The dependent variable is the logistic transformation of a country's share of manufacturing value added in GDP. We use a logistic transformation to account for the fact the manufacturing share is limited to a range between 0 and 1.[7] Concerning the regressors, we discuss the choice of suitable empirical proxies in turn. Additional details on the data and their sources, as well as a list of countries used in the regressions below are contained in Appendix A.

Keeping in line with existing studies on Ricardian comparative advantage (*e.g.,* Golub and Hsieh, 2000), we use labor productivity as a proxy for productivity. In contrast to total factor productivity, this has the advantage of considerably increasing the number of available observations.

We measure country l's centrality (CEN_{lt}) as the sum of all other countries' GNP, weighted

---

[4] An exception is Deardorff (2004).

[5] Also see Davis (1998) and Hanson and Xiang (2004).

[6] These are the correlations we will aim at reproducing in our calibration exercise.

[7] Using untransformed manufacturing shares instead does not change any of the qualitative results reported below. We have also experimented with including the share of services in GDP as an additional control variable, again without finding any significant changes in the other coefficient estimates (both sets of results are available from the authors upon request).

by the inverse of bilateral distances, which are taken to proxy for trade costs between locations:

$$CEN_l = \sum_{j \neq l} GNP_j \times dist_{jl}^{-1}. \tag{2}$$

This specification reflects the basic intuition of our discussion. What matters is centrality in an economic geography sense, that is proximity to markets for domestic products. Of course, the above centrality index is closely related to the concept of market potential first proposed by Harris (1954), which has been frequently used in both geography and – more recently – in economics. A number of studies have demonstrated that this simple proxy has strong explanatory power and yields results very similar to more complex approaches that estimate trade costs from trade flow gravity equations (see, for example, Head and Mayer (2006), or Breinlich (2006)).[8]

As additional control variables, we also include agricultural productiviy (AP) to account for the pro-industrializing relative-demand effect discussed above, and population size (POP) as an additional proxy for the extent of the domestic market. We have data for all the required variables for 112 countries in 2000. Keeping in line with the focus of this paper on the industrialization of developing countries, however, we exclude high-income countries from our regression sample (although of course all available countries are used to calculate the centrality measure).[9] In our robustness checks, we will also briefly present results for the full sample.[10]

In Table 1, we present a number of univariate correlations between the logistic transformation of manufacturing shares and our proxies for comparative advantage (relative productivity, RP) and centrality. Columns 1-2 replicate the correlations from Figures 1 and 2 and show that using a logistic transformation of manufacturing shares as the dependent variable leads to similar results. In column 3, we use our more sophisticated measure of centrality (2). Note that we would now expect to find a positive and significant sign, which is indeed what we do. We also note that both measures of centrality seem to be important determinants of levels of industrialization. They explain around 10% of the cross-sectional variation of manufacturing shares in our sample. This is comparable in magnitude to *per-capita* income whose positive correlation with manufacturing shares in the initial phase of development is a key variable in much of the existing empirical literature on cross-country patterns of industrialization (e.g., Syrquin and Chenery (1989)).

In columns 4-8, we undertake a first series of robustness checks. Column 4 uses data on sector-specific purchasing power parities to strip out the variation in prices from our relative productivity measure, so that the remaining variation more closely reflects physical productivity differences (Appendix C provides additional details). As seen, using this refined measure leaves the correlation with manufacturing shares basically unchanged. In columns 5-6, we include a dummy for China and the South-East Asian economies of Korea, Thailand, Malaysia, Indonesia and the Philippines. These countries are arguably special cases because of their very successful

---

[8] Using a nonstructural measure also seems to be better in line with the more explorative character of this section.

[9] We use the World Bank's income classification and exclude all countries with gross national income per capita in excess of 9,265 USD in 1999 ("high income countries").

[10] See footnote 34 in Section 5 and Appendix Table A.2.

export-oriented industrialization strategies and are also potentially influential outliers in both Figures 1 and 2. The corresponding dummy variable (not reported) is indeed positive and highly significant but the coefficient on our centrality measure remains almost unchanged. The positive correlation between manufacturing shares and relative productivity is increased and becomes statistically significant. In columns 7 and 8, we present results for additional years for which comparable cross-sectional data on relative productivities is available (1980 and 2000, yielding a unbalanced panel of 256 observations in total). Again, using these additional data makes the results from columns 1 and 3 stronger.

In Table 2, we gradually build up our results to the full specification (1). In column 1 we include population size, column 2 uses agricultural productivity as an additional regressor, and column 3 includes both population and agricultural productivity. In column 4, we drop agricultural productivity and replace it with *per-capita* GDP. *Per-capita* GDP helps controlling for the purchasing power of the local population, skill levels, and other potentially confounding factors. Note, however, that it is very highly correlated with agricultural productivity so that in practice both variables are likely to pick up the influence of similar omitted variables. The high correlation also makes the inclusion of both variables in the same regression impossible.[11] In columns 5-8, we again use our larger sample for the years 1980, 1990 and 2000.

Three main insights arise from these regressions. First, proxies for the size of the domestic market are strongly positively correlated with levels of industrialization, as was to be expected from prior results in the literature. Second, centrality retains its positive and significant influence throughout. Third, comparative advantage in agriculture has a positive and significant effect on industrialization whenever we do not control for absolute agricultural productivity, and an insignificant effect whenever we do. This suggests that relative productivity might be picking up the influence of absolute productivity levels in agriculture.

Limited data availability for relative and absolute agricultural productivity prevents us from estimating specification (1) for a yet larger sample. In columns 9-11, we exclude these variables which increases the sample size more than tenfold since we can now use observations for every year from 1980 to 2005. This allows us to provide some further results on the importance of centrality for industrialization by running variations of the following specification:

$$ltShareM_{lt} = \alpha + d_t + d_l + \delta_1 CEN_{lt} + \delta_2 PCGDP_{lt} + \delta_3 POP_{lt} + \varepsilon_{lt}, \tag{3}$$

where $PCGDP_{lt}$ denotes *per-capita* GDP and d_t and d_l are a full set of time and country fixed effects. Column 9 of Table 2 reports results for an OLS regression pooled over the period 1980-2005 with year dummies only. Column 10 estimates the full specification (3) by including country fixed effects, thus eliminating any time-invariant heterogeneity across countries from our correlations. Column 11 uses long first differences between 1980 and 2005. All regressions give a similar picture as the results for the smaller sample: both the size of the domestic market and access to foreign markets are positively correlated with levels of industrialization. If anything, controlling

---

[11] The correlations of the variables in logs is 84% in our sample.

for country-specific effects in columns 10 and 11 implies an even stronger role for centrality.[12]

3 The Model

Consider a world with countries $j = 1, ..., R$, each with L_j consumers, each of which supplies one unit of labor inelastically. There are two sectors, agriculture and manufacturing; we assume perfect labor mobility between sectors, and no international labor mobility. As we discuss in Section 5.2 and Appendix D below, adding a third, non-tradable sector (*i.e.* services) complicates the analysis but yields similar results, both qualitatively and (in our calibration) quantitatively. Thus, for the sake of simplicity we abstract from the services sector for most of our analysis.

3.1 Demand Side

Preferences are identical across countries. Country-j individuals maximize a Stone-Geary utility function over consumption of an agricultural and a manufacturing composite good:

$$U_j = \alpha \ln(M_j) + (1-\alpha)\ln(A_j - \underline{A}), \tag{4}$$

$\alpha \in (0,1)$, where

$$M_j = \left[\sum_{l=1}^{R} m_{lj}^{(\sigma_M-1)/\sigma_M}\right]^{\sigma_M/(\sigma_M-1)}, \tag{5}$$

$$A_j = \left[\sum_{l=1}^{R} a_{lj}^{(\sigma_A-1)/\sigma_A}\right]^{\sigma_A/(\sigma_A-1)}. \tag{6}$$

Both M_j and A_j are Armington aggregators of country-specific varieties: every country is assumed to produce one differentiated variety.[13] M_j is consumption of the manufacturing composite and m_{lj} is the amount of the variety produced in l that is consumed by an individual consumer in j. Similarly, A_j is consumption of the agricultural composite and a_{lj} is the amount of the variety produced in l that is consumed by an individual consumer in j. The elasticities of substitution between varieties are constant at $\sigma_M, \sigma_A > 1$.

$\underline{A} > 0$ denotes minimal consumption of agricultural goods, *i.e.* the subsistence level. These preferences guarantee that above the level $\underline{A}$ the expenditure share of agricultural goods declines with rising *per capita* income. This is the so-called Engel's law, which has strong empirical foundations (see Crafts (1980)). We assume $\underline{A} < \theta_{Al}$ for all l, where θ_{Al} is agricultural productivity in country l (to be defined more precisely below). This assumption guarantees that *per-capita*

---

[12] The downside of omitting relative and absolute agricultural productivity is of course that their exclusion is likely to lead to omitted variable bias. To verify the likely magnitude of this bias, we estimated both (1) and (3) on the same samples used in columns 1-4 and 5-8. Comparing the coefficient on CEN in these regressions does indeed suggest that omitting AP and RP leads to an upward bias, albeit a small one.

[13] The Armington assumption ensures that all countries consume all varieties provided trade costs and elasticities of substitution are finite. This implies that all countries have diversified production structures in equilibrium, which is the empirically relevant case and renders the model relatively tractable below.

income in each country is sufficient to reach the subsistence level. Thus, at least some expenditure will be devoted to manufacturing products.

Below we impose enough structure so that labor is the only source of income. The individual's budget constraint in country j is therefore given by $P_{Mj}M_j + P_{Aj}A_j = w_j$, with w_j denoting the wage in j, equal across sectors. P_{Mj} and P_{Aj} are price indices for the manufacturing and agricultural composite goods. Prices paid for the different products in the importing location j, p_{Mlj} and p_{Alj}, consist of the mill price charged in country l plus industry-specific bilateral trade costs $T^M_{lj}, T^A_{lj} \geq 1$. ($T^M_{jj} = T^A_{jj} = 1$ for all j.) These trade costs are of the iceberg-type form: for every unit of a good that is shipped from l to j, only $1/T_{lj}$ arrive while the rest "melts" *en route*.

Utility maximization yields country-j individual's demand for manufactured and agricultural goods produced in l. Aggregating across individuals and countries, total demands (inclusive of trade costs) for country-l goods are

$$m^D_l = p^{-\sigma_M}_{Ml} \sum_{j=1}^{R} \left(T^M_{lj}\right)^{1-\sigma_M} P^{\sigma_M - 1}_{Mj} E_{Mj}, \tag{7}$$

$$a^D_l = p^{-\sigma_A}_{Al} \sum_{j=1}^{R} \left(T^A_{lj}\right)^{1-\sigma_A} P^{\sigma_A - 1}_{Aj} E_{Aj}, \tag{8}$$

where

$$P_{Mj} = \left(\sum_{l=1}^{R} p^{1-\sigma_M}_{Mlj}\right)^{\frac{1}{1-\sigma_M}} = \left[\sum_{l=1}^{R} \left(p_{Ml} T^M_{lj}\right)^{1-\sigma_M}\right]^{\frac{1}{1-\sigma_M}}, \tag{9}$$

$$P_{Aj} = \left(\sum_{l=1}^{R} p^{1-\sigma_A}_{Alj}\right)^{\frac{1}{1-\sigma_A}} = \left[\sum_{l=1}^{R} \left(p_{Al} T^A_{lj}\right)^{1-\sigma_A}\right]^{\frac{1}{1-\sigma_A}}. \tag{10}$$

$E_{Mj} = \alpha \left(w_j - P_{Aj}\underline{A}\right) L_j$ and $E_{Aj} = \left[(1-\alpha)w_j + \alpha P_{Aj}\underline{A}\right] L_j$ denote total expenditures on manufacturing and agricultural goods in country j, respectively.

3.2 Production

Each country produces a differentiated variety of the manufacturing and the agricultural goods. Sectors are perfectly competitive, operate under constant returns to scale, and use labor as the only input. The amount of labor employed in manufacturing in country l is denoted by L_{Ml}, and supply of the local variety is $m_l = \theta_{Ml} L_{Ml}$, where θ_{Ml} denotes productivity in manufacturing in country l. The amount of labor employed in agriculture in country l is denoted by L_{Al}, and supply of the local variety is $a_l = \theta_{Al} L_{Al}$. Productivity levels are allowed to vary across countries and sectors. Positive production implies f.o.b. prices equal the cost of producing one unit of output: $p_{Ml} = w_l/\theta_{Ml}$ and $p_{Al} = w_l/\theta_{Al}$.

3.3 Equilibrium

Equilibrium in the manufacturing and agricultural goods markets requires $a_l = a_l^D$ and $m_l = m_l^D$, respectively. This yields

$$\begin{aligned} m_l &= \theta_{Ml}^{\sigma_M} w_l^{-\sigma_M} \left[\sum_{j=1}^{R} \left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M - 1} E_{Mj} \right], && (11) \\ a_l &= \theta_{Al}^{\sigma_A} w_l^{-\sigma_A} \left[\sum_{j=1}^{R} \left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A - 1} E_{Aj} \right]. && (12) \end{aligned}$$

Labor demands in manufacturing and agriculture are, respectively,

$$\begin{aligned} L_{Ml} &= \theta_{Ml}^{\sigma_M - 1} w_l^{-\sigma_M} \left[\sum_{j=1}^{R} \left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M - 1} E_{Mj} \right], && (13) \\ L_{Al} &= \theta_{Al}^{\sigma_A - 1} w_l^{-\sigma_A} \left[\sum_{j=1}^{R} \left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A - 1} E_{Aj} \right]. && (14) \end{aligned}$$

Notice these are functions of the vector of wages of all countries. Full employment requires $L_{Ml} + L_{Al} = L_l$, which can be rewritten as

$$\theta_{Ml}^{\sigma_M - 1} w_l^{-\sigma_M} \left[\sum_{j=1}^{R} \left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M - 1} E_{Mj} \right] + \theta_{Al}^{\sigma_A - 1} w_l^{-\sigma_A} \left[\sum_{j=1}^{R} \left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A - 1} E_{Aj} \right] = L_l. \quad (15)$$

These are R non-linear equations in the R wage rates, determining the vector of wages and subsequently all other equilibrium variables of the model.[14]

To summarize, the crucial features of this model that will drive the results in the remaining sections are as follows. First, there are varying levels of agricultural and manufacturing productivity across locations, which together with non-homothetic preferences will drive industrialization and de-industrialization through comparative advantage and Engel's law. Second, positive trade costs render relative geographical positions important, both by conferring a market size advantage to more central regions and by softening the impact of comparative advantage across space. As the focus of this paper is on industrialization, we also introduce the share of manufacturing in GDP as a further variable which in the following is also referred to as the level of industrialization of a location:

$$ShareM_l = \frac{p_{Ml} m_l}{p_{Ml} m_l + p_{Al} a_l} = \frac{w L_{Ml}}{w L_{Ml} + w L_{Al}} = \frac{L_{Ml}}{L_l}. \quad \text{(MS)}$$

---

[14] By Walras's Law, one of these R equations is redundant.

4 Analysis

This section analyzes the properties of the model just developed and uses it to shed light on the puzzles raised in the introduction. As will become clear, the present model nests some of the existing approaches in the literature on demand-driven industrialization as the two special cases of infinite and zero trade costs ("closed economy" and "free trade"). We will demonstrate that in the case with positive trade costs new results arise that help resolve our two puzzles.

4.1 Closed Economy

With infinitely high levels of trade costs, *i.e.* under autarky, it is easy to show that the expression for the share of manufacturing in GDP simplifies to

$$ShareM_l = \alpha \left(1 - \frac{\underline{A}}{\theta_{Al}}\right). \qquad (MS_{AUT})$$

As is apparent from equation (MS_{AUT}), the manufacturing share in GDP increases with agricultural productivity. Non-homothetic preferences (due to the positive subsistence consumption level in agriculture, $\underline{A}> 0$) are crucial for this result. Intuitively, the increases in *per capita* income resulting from higher values of θ_{Al} lead to a decline in the share of subsistence consumption in total expenditure. As every unit of income above the subsistence level is spent in fixed proportions on agricultural and manufacturing varieties, the expenditure share of the latter rises. In a closed economy, this leads in turn to a shift of labor into manufacturing and an increase in $ShareM_l$. As discussed, we refer to this positive impact of agricultural productivity on industrialization as the "relative-demand effect" of agricultural productivity shocks. Very similar effects are obtained in the existing literature (*e.g.* Matsuyama (1992) or Murphy *et al.* (1989b)). Needless to say, the autarky assumption renders any cross-country differences in centrality or comparative advantage completely irrelevant.

4.2 Free Trade

Under free trade, Ricardian comparative advantage emerges as the key factor for the determination of the level of industrialization. With costless trade, and assuming $\sigma_M = \sigma_A = \sigma$, it is easy to show that

$$\frac{L_{Ml}/L_{Al}}{L_{Ml'}/L_{Al'}} = \left(\frac{\theta_{Ml}/\theta_{Al}}{\theta_{Ml'}/\theta_{Al'}}\right)^{\sigma-1}. \qquad (16)$$

Since

$$ShareM_l = \frac{L_{Ml}}{L_l} = \frac{1}{1 + \left(L_{Ml}/L_{Al}\right)^{-1}}, \qquad (17)$$

lower ratios θ_{Ml}/θ_{Al} imply a stronger bias towards agriculture in the production structures of countries. As in standard Ricardian models of international trade, being relatively productive in agriculture biases a country's production structure towards agriculture, thus reducing the share of manufacturing in GDP as the location specializes accordingly. We will refer to this as the "comparative advantage effect". Notice that free trade eliminates any independent influence of the productivity level θ_{Al} on industrialization via the non-homothetic preferences channel we discussed above.

4.3 Costly Trade

In the presence of positive trade costs that are different across country pairs, the model becomes much less tractable. We therefore use simplified versions of the full model in a number of examples that illustrate the new types of results our model can yield in this new environment. In Section 5 we relax these simplifying assumptions and solve the model numerically, using calibrated parameter values.

It is a long-standing theoretical result in international trade theory that the size of the home market matters for industrial structure (Krugman (1980), Krugman and Helpman (1985)). More recently, Davis and Weinstein (1998, 2003) found empirical support for home market effects in a study on OECD countries. However, their finding depended crucially on taking into account demand linkages across locations, indicating the importance of foreign demand.[15] In models of industrialization, however, the role of access to foreign markets has been ignored so far, even though its inclusion seems to be a logical extension of the existing literature. In a world with positive trade costs, central locations have effectively a larger market size as they are closer to sources of demand, *ceteris paribus*. Note that this holds in addition to any size advantage the domestic economy may have and depends on its position relative to other locations.

More central countries can benefit from their position to industrialize even in the absence of any technological or size advantage, simply because being more central raises relative demand for the central country's manufacturing good. There are several theoretical reasons why one would expect central locations to experience a larger relative demand for their manufacturing goods than peripheral ones. First, being more central raises demand for both agricultural and manufacturing goods and raises wages.[16] With non-homothetic preferences, this leads to an expansion of domestic manufacturing expenditure which, with positive trade costs, will translate into a domestic manufacturing share higher than in other countries, as the resulting increase in manufacturing expenditure will have a stronger effect on the domestic manufacturing good than on those produced by other countries. The following particular case of our model illustrates this mechanism.

Example 1 Consider a three-country world, $R = 3$, and a geographic structure such that country 1 takes a "central" position while countries 2 and 3, which are fully symmetric, are in the "periphery": we model this by assuming that country 1 can trade with both 2 and 3 at positive but finite trade costs ($T_{12} = T_{21} = T_{13} = T_{31} = T > 1$) and that countries 2 and 3 cannot trade with one another ($T_{23} = T_{32} = \infty$).[17] Trade costs are assumed equal across sectors. We simplify

---

[15] Indeed, in an earlier version of the same paper, Davis and Weinstein (1996) interpreted local demand as purely domestic and ignored linkages across borders, and were unable to detect home market effects.

[16] See Redding and Venables (2004) for empirical evidence on the positive effect of centrality on income levels.

[17] For the sake of the argument, we rule out the possibility that countries 2 and 3 can trade via country 1.

further by assuming $\sigma_M = \sigma_A = \sigma$. Finally, we choose all parameters to be identical across countries (except for the bilateral trade costs) and, in particular, we set $\theta_{Aj} = \theta_{Mj} = L_j = 1$. Profiting from the symmetry we have imposed, let us normalize $w_2 = w_3 = 1$.

It is easy to show that we cannot have an equilibrium in which $w_1 = 1$, as the model's market clearing conditions would be violated. We can prove this by contradiction. If it were the case that $w_1 = w_2 = w_3 = 1$, then aggregate labor demand would be different across countries:

$$L_{M1} + L_{A1} = \frac{1}{2T^{1-\sigma} + 1} + \frac{2}{T^{\sigma-1} + 1} > L_{M2} + L_{A2} = \frac{1}{2 + T^{\sigma-1}} + \frac{1}{1 + T^{1-\sigma}}. \tag{18}$$

Thus, it must be the case that $w_1 > w_2 = w_3$. Due to the non-homotheticity of preferences, this implies that country 1's expenditure is biased towards manufacturing: $E_{M1} > E_{M2}$. As discussed above, positive trade costs lead this bias in demand for manufacturing goods to favor country 1's manufacturing industry primarily:

$$\begin{aligned} L_{M1} &= \frac{1}{w_1^{1-\sigma} + 2T^{1-\sigma}} E_{M1} + \frac{2}{w_1^{1-\sigma} + T^{\sigma-1}} E_{M2}, & (19) \\ L_{M2} &= \frac{1}{w_1^{1-\sigma} T^{\sigma-1} + 2} E_{M1} + \frac{1}{(w_1 T)^{1-\sigma} + 1} E_{M2}. & (20) \end{aligned}$$

Establishing analytical results here is difficult, but the condition $2 > T^{\sigma-1}$, for example, is sufficient for $L_{M1} > L_{M2}$, which implies a larger manufacturing share in the central country.

A second reason why centrality favors industrialization is based on the different elasticities of substitution of manufacturing and agricultural products. Higher wages due to a more central position lead to higher prices of both types of goods. If agricultural goods are more homogeneous than manufacturing goods (this would correspond to $\sigma_A > \sigma_M$ in our model), as is usually the case, central locations will specialize in manufacturing, *ceteris paribus*. This is since demand for locally produced manufacturing varieties will be less sensitive to higher prices than demand for the country's agricultural variety. The following example illustrates this mechanism.

Example 2 Again assume $R = 3$ and that all parameters are identical across countries (except for the bilateral trade costs) and, in particular, that $\theta_{Aj} = \theta_{Mj} = L_j = 1$, $\sigma_A = \infty$, and $\sigma_M > 1$ but finite. Again, we consider a geographic structure such that country 1 takes a "central" position while countries 2 and 3 are in the "periphery": here we model this by assuming that country 1 can trade freely with both 2 and 3 ($T_{12} = T_{21} = T_{13} = T_{31} = 1$) and that countries 2 and 3 cannot trade with one another ($T_{23} = T_{32} = \infty$).[18] Trade costs are again equal across sectors. We take the agricultural good as the numéraire. Under incomplete specialization for all countries, the labor market equilibrium conditions comprise equations

$$\begin{aligned} L_{M1} &= \frac{1}{3}\alpha\left(1 - \underline{A}\right) + \alpha\left(1 - \underline{A}\right) = \frac{4}{3}\alpha\left(1 - \underline{A}\right), & (21) \\ L_{M2} &= L_{M3} = \frac{1}{3}\alpha\left(1 - \underline{A}\right) + \frac{\alpha}{2}\left(1 - \underline{A}\right) = \frac{5}{6}\alpha\left(1 - \underline{A}\right). & (22) \end{aligned}$$

---

[18] We again rule out the possibility that countries 2 and 3 can trade via country 1.

It is easy to show that in this case country 1's manufacturing share is larger than that of countries 2 and 3, since $L_{M1} > L_{M2} = L_{M3}$. If parameter values in this incomplete specialization scenario yielded $L_{M1} > 1$, then country 1 would specialize completely in manufacturing.[19] In this case, the labor market equilibrium conditions comprise equations

$$
\begin{aligned}
L_{M1} &= w_1^{-\sigma_M}\left[\frac{\alpha\left(w_1-\underline{A}\right)}{\left(2+w_1^{1-\sigma_M}\right)}+\frac{2\alpha\left(1-\underline{A}\right)}{\left(1+w_1^{1-\sigma_M}\right)}\right]=1, && (23)\\
L_{M2} &= L_{M3}=\frac{\alpha\left(w_1-\underline{A}\right)}{\left(2+w_1^{1-\sigma_M}\right)}+\frac{\alpha\left(1-\underline{A}\right)}{\left(1+w_1^{1-\sigma_M}\right)}<1, && (24)
\end{aligned}
$$

which imply $w_1 > w_2 = w_3 = 1$. Notice that the mechanism discussed in this example does not depend on the non-homotheticity of preferences: assuming $\underline{A}= 0$ would not change the result here.[20]

Finally, trade costs can also affect the response of specialization patterns to changes in productivity. An increase in agricultural productivity, for example, will generate a "relative-demand effect" in favor of the manufacturing industry through the non-homothetic preferences, and a "comparative-advantage effect" in favor of agriculture. Which effect dominates depends on the link between domestic expenditure and production and thus the level of trade costs. Under autarky, where consumption and production are perfectly linked, the relative-demand effect dominates, as we already saw above. Under free trade, where consumption and production are separate choices, the comparative advantage effect dominates. Outside these two extreme cases, with intermediate values for trade costs, which effect dominates depends on parameter values. The following example sketches some intuition for this case.

Example 3 Consider many countries (R large). For simplicity, we assume again $\sigma_M = \sigma_A = \sigma$. All country-pairs face the same bilateral trade costs: $T_{jl}^M = T_{jl}^A = T > 1$ for all $j \neq l$. All countries have the same population size and productivities, $\theta_{Aj} = \theta_{Mj} = L_j = 1$ for all j, except for $\theta_{A1} > 1$. By symmetry, we can normalize $w_j = 1$ for all $j \neq l$. From the model's equilibrium conditions,

$$
\begin{aligned}
\frac{L_{M1}}{L_{A1}} &= \frac{P_{M1}^{\sigma-1}E_{M1}+\sum_{l\neq 1}T^{1-\sigma}P_{Ml}^{\sigma-1}E_{Ml}}{\theta_{A1}^{\sigma-1}\left[P_{A1}^{\sigma-1}E_{A1}+\sum_{l\neq 1}T^{1-\sigma}P_{Al}^{\sigma-1}E_{Al}\right]}, && (25)\\
\frac{L_{Mj}}{L_{Aj}} &= \frac{P_{Mj}^{\sigma-1}E_{Mj}+\sum_{l\neq j}T^{1-\sigma}P_{Ml}^{\sigma-1}E_{Ml}}{P_{Aj}^{\sigma-1}E_{Aj}+\sum_{l\neq j}T^{1-\sigma}P_{Al}^{\sigma-1}E_{Al}}, && (26)
\end{aligned}
$$

---

[19] Under the assumption $\sigma_A = \infty$, there is no need for every country to produce its own "variety" of the agricultural good.

[20] A third mechanism which could generate higher levels of industrialization in the center is based on the manufacturing industry having access to both a constant returns to scale and an increasing returns to scale (IRS) production technique (see Murphy et al., 1989a/b). In this case, central locations would be the first, *ceteris paribus*, to reach the critical level of demand that makes IRS production profitable. This mechanism is absent from our model, as we assume constant returns to scale across sectors.

for countries 1 and j. Assuming that trade costs are such that countries consume sizable amounts of foreign goods, one can neglect the effect of θ_{A1} on the price levels P_{Ml} and P_{Al}. A high θ_{A1} therefore has a direct effect in the denominator of equation (25) and an indirect effect via a high w_1 in the terms E_{M1} and E_{A1} of both equations. Notice first that the direct effect of θ_{A1} raises country 1's agricultural share in GDP (the comparative-advantage effect). Second, a higher w_1 (due to a higher θ_{A1}) tilts relative expenditure towards manufacturing in both country 1 and country j because of the non-homotheticity in demand, but more so in country 1 due to the presence of trade costs. As discussed above, this relative-demand effect operates in the direction opposite to the comparative-advantage effect.

5 A Calibrated Multi-Country Model of Industrialization

The discussion in Section 4 has shown that our model is, in principle, able to replicate the stylized facts from the introduction. However, our results relied on a number of simplifying assumptions and may not generalize to the full model from Section 3 which, as discussed, does not have an analytical solution. Also, the exact magnitude of parameter values mattered a great deal for the direction of results, especially in example 3.

This is why we complement the analytical results with a full-scale multi-country calibration of our model. That is, we ask to what extent the model matches our stylized facts for empirically plausible parameter values. In the following, we choose parameters to match international trade and expenditure data and use this calibrated model to generate data on manufacturing shares and the independent variables used in the regressions in Tables 1 and 2 (more details on how exactly this is done are provided below). Intuitively, if the true data generating process for our variables of interest is similar to the one postulated by our model, we should expect to find comparable multivariate correlations in both the real and the generated data.[21]

5.1 Parameter Values

For a calibrated version of our model, we need data on the size of countries' workforces (L_l) and productivity levels (θ_{Al}, θ_{Ml}), and values for the parameters governing substitution elasticities (σ_A, σ_M), trade costs (T^A_{lj}, T^M_{lj}), the manufacturing expenditure share (α), and subsistence consumption ($\underline{A}$). Table 3 provides parameter estimates and a brief description of the calibration procedure and data sources used. In the following, we describe the calibration in more detail. Data requirements limit the sample to 107 countries for the year 2000, 79 of which are classified

---

[21] Note that we are not primarily interested in matching the cross-section of manufacturing shares and the independent variables in Tables 1 and 2 as closely as possible, but only ask our model to reproduce (univariate and multivariate) correlations found in the real data. This is of course a strictly weaker test than trying to match the above variables exactly. (If we succeeded in doing so, we would naturally be able to reproduce the correlations as well). To be sure, matching the entire distribution of the variables from Tables 1 and 2 is also interesting but would require a yet more complicated model and is beyond the research objective of this paper (which, to reiterate, is to explain correlations in the data which are at odds with existing theories). Having said this, we provide some evidence below that the version of our model with positive but finite trade costs is actually more successful in matching the cross-sectional variation in manufacturing shares than versions based on autarky or free trade.

as developing and will be used in our regression analysis of the simulated data.[22]

We follow Feenstra (1994) in using variation in import quantities and prices to identify elasticities of substitution among manufacturing and agricultural varieties (σ_A, σ_M). This approach, as extended by Broda and Weinstein (2006) and Broda *et al.* (2006), has become the dominant method for estimating substitution elasticities in the international trade literature in recent years. In our setting, it has the additional advantage of building on a very similar demand structure as our paper (CES and Armington varieties), while allowing for more general supply side features. We adapt this approach to our setting by using data which correspond to our calibration exercise in terms of country coverage, time period and the definition of sectors for which we estimate elasticities. We focus on a discussion of our estimates in the following and refer the reader to Appendix B for a more detailed description of the Feenstra-Broda-Weinstein methodology and how we adapt it to our setting.

For our baseline elasticity estimates, we use cross-country trade data for the year 2000 but restrict the estimation sample to the 102 countries which are in our calibration sample and for which we have the necessary information on import prices and quantities.[23] We obtain $\sigma_M = 2.3$ and $\sigma_A = 2.3$. For comparison, Broda *et al.* (2006) estimate elasticities of substitution between varieties of goods produced in each of approximately 200 sectors, separately for 73 countries (rather than imposing a common elasticity as we do in accordance with our model). The median across these estimates for the 60 countries also present in our data is 3.4. Given the much higher degree of aggregation in our data (two instead of 200 sectors), our lower estimates seem plausible. This is because both economic theory and the empirical results of Broda and Weinstein (2006) and Broda *et al.* (2006) suggest that estimated elasticities should fall as the level of aggregation increases and varieties become less similar.[24]

As a robustness check, we also obtain estimates using data on imports by the U.S. from the countries in our calibration sample.[25] These data are likely to be of higher quality than the cross-country data used before (see Feenstra, Romalis and Schott, 2002), although of course we only have one importer now instead of 102. Using these data yields comparable coefficient magnitudes as before although agricultural varities are now estimated to be slightly more substitutable across countries ($\sigma_M = 2.0$ and $\sigma_A = 2.6$).

Since labor is the only factor of production in our model, we proxy θ_{Ml} and θ_{Al} by labor productivity in manufacturing and agriculture, respectively. However, as already briefly discussed in Section 2, the cross-sectional variation in labor productivity across countries which we observe

---

[22] See Appendix A for a list of countries included in the calibration sample. All 107 countries will be used to generate our synthetic data set as developed countries do of course play a major role in determining manufacturing shares and centrality of developing countries.

[23] Three groups of countries only report one common set of trade data, explaining the five missing observations: Botswana, Lesotho, Namibia and South Africa; Belgium and Luxembourg; and St. Lucia and St. Vincent and the Grenadines.

[24] Closer to our level of aggregation but obtained via a different methodology is the estimate by Eaton *et al.* (2008) who use French firm-level data to estimate an elasticity of substitution between individual manufacturing varieties of $\sigma_M = 1.7$.

[25] Again, we lose five countries due to aggregation in the trade data (see footnote 23), leaving us with 101 exporters (the U.S. is of course excluded as an exporter).

in the data is driven by both differences in technological efficiency and differences in prices. That is, $lp_{.l} = VA_{.l}/L_{.l} = p_{.l}x_{.l}/L_{.l}$ in terms of our model because we abstract from intermediate inputs. Since we are only interested in $\theta_{.l} = x_{.l}/L_{.l}$, we use data on purchasing power parities for agriculture and manufacturing goods consumption from the International Comparison Program (ICP) to construct proxies for $p_{.l}$ and strip out price variation from the data (see Appendix C for details).[26]

Estimates of the trade cost matrices can be obtained via gravity equation regressions using cross-country manufacturing and agricultural trade data. To see this, note that exports in the model are:

$$\begin{aligned} X_{lj}^{M} &= p_{Ml}^{1-\sigma_M}\left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M-1}E_{Mj}, \\ X_{lj}^{A} &= p_{Al}^{1-\sigma_A}\left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A-1}E_{Aj}. \end{aligned}$$

The only bilateral variable on the right-hand side in the above expressions is trade cost T_{lj}. We proxy for these costs by $T_{lj} = dist_{lj}^{\delta_1} e^{\delta_2 d_{int,lj}}$, where $dist_{lj}$ denotes the bilateral distance between countries l and j, and δ_1 denotes the elasticity of trade cost with respect to distance. The dummy variable $d_{int,lj}$ indicates if a trade flow crosses national borders (*i.e.*, $d_{int,lj} = 1$ if $l \neq j$ and 0 if $l = j$). This is a parameterisation of trade cost which is common in the international trade literature (*e.g.*, Wei (1996)). Proxying all other variables by importer and exporter fixed effects and adding an error term, we can rewrite bilateral exports as

$$\begin{aligned} X_{lj}^{M} &= d_{\exp,M} \times d_{imp,M} \times dist_{lj}^{(1-\sigma_M)\delta_{M1}} e^{(1-\sigma_M)\delta_{M2} d_{int,M}} \times \varepsilon_{lj,M}, \\ X_{lj}^{A} &= d_{\exp,A} \times d_{imp,A} \times dist_{lj}^{(1-\sigma_A)\delta_{A1}} e^{(1-\sigma_A)\delta_{A2} d_{int,A}} \times \varepsilon_{lj,A}. \end{aligned} \tag{27}$$

We estimate (27) in its original multiplicative form via Poisson QMLE, using data from the sources listed in Table 3 and following Wei (1996) in proxying internal trade flows as domestic production (gross output) minus exports.[27] As has been pointed out by Santos-Silva and Tenreyro (2006), Poisson QMLE can accomodate zero trade flows, which are common in our data, and leads to consistent parameter estimates even in the presence of heteroskedasticity in ε_{lj}. Appendix Table

---

[26] Echevarria (1997) uses a similar approach based on U.S. data. One way of interpreting our price correction of measured productivity is that we want to allow for more general price-setting mechanisms in the data than the restrictive mechanism assumed in our theory. There, the simplifying assumption of perfect competition leads to differences in productivity being exactly compensated by differences in prices. In principle, one could also address this issue as in Bernard *et al.* (2003), who assume a specific form of imperfect competition to generate variation in mark-ups across producers (*i.e.*, Bertrand competition and limit pricing in which the lowest-cost supplier is constrained not to charge more than the second-lowest cost supplier). The drawback of this approach is that results are potentially sensitive to the particular choice of mechanism generating variation in mark-ups. In our context, data availability is an additional serious issue, since firm-level data or at least information about within-sector, across-firm productivity differences are required to implement the Bernard *et al.* (2003) methodology. Such data are not available in comparable form for the countries in our sample.

[27] Again, we restrict our sample to countries which are in our calibration sample and for which we have the necessary data. For manufacturing trade, we lose the same five countries as before due to aggregation (see footnote 23), plus Uzbekistan due to missing production data. Agricultural production data are unfortunately much less complete, restricting the estimation sample to 66 countries. Using all 78 countries for which production data is available only leads to minor changes in the estimates for the trade cost elasticities which have no impact on the following results (details available from the authors).

A.1 contains details of the estimation results, which are broadly in line with those from comparable specifications in the literature.[28] In robustness checks below, we will also use estimates of T_{lj} obtained by estimating a log-linearized version of (27) via OLS (see Appendix Table A.1 for results). The distance coefficients in both sets of estimations provide estimates for $(1-\sigma)\,\delta$ which, together with our assumed values of σ, yield estimates for δ and thus for $T_{lj} = dist_{lj}^{\delta_1} e^{\delta_2 d_{int,lj}}$.

We again use data on nominal and real expenditure on food and manufacturing goods from the ICP to estimate α and $\underline{A}$. In the model, the nominal expenditure share of manufacturing in GDP, and real consumption of agricultural goods per head, respectively, are given by:

$$S_{EMj} = \frac{E_{Mj}}{w_j L_j} = \alpha - \alpha \underline{A} \frac{P_{Aj}}{w_j}, \tag{28}$$

$$\frac{E_{Aj}}{P_{Aj} L_j} = (1-\alpha) w_j P_{Aj}^{-1} + \alpha \underline{A}. \tag{29}$$

Note that as $w_j \rightarrow \infty$, $S_{EMj} \rightarrow \alpha$; and as $w_j \rightarrow \underline{A} P_{Aj}$, $E_{Aj} / \left(P_{Aj} L_j\right) \rightarrow \underline{A}$. For our simulations below, we thus use the nominal expenditure share of manufacturing in total expenditure on food and manufacturing $\left(S_{EMj} / \left(S_{EMj} + S_{EAj}\right)\right)$ of the richest country (Luxembourg) in our data as a proxy for α. Likewise, we use the real food expenditure per worker of the poorest country (Zambia) as a proxy for $\underline{A}$.[29]

5.2 Results

We now present results for the same regressions as in Tables 1 and 2, but this time we use simulated rather than actual data for the year 2000.[30] That is, we use the calibrated model to generate artificial data on manufacturing shares for the developing countries in our simulation sample. Note that our model also generates data for *per-capita* GDP (equal to wages in the model), GDP (wages times population size) and centrality (calculated according to (2), using the same distance data but replacing GNP with model generated GDP). Thus, we use generated data for both dependent and independent variables in the regressions below, consistent with the notion that we would like to evaluate whether our model is comparable to the data generating process in the real world. Population size and productivity data are of course directly used as model parameters, and, apart from the price correction discussed, are identical to the data used

---

[28] In a recent meta study, Disdier and Head (2008) report that the mean distance elasticitiy of the 1,467 estimates they analyze is -0.9, very close to our Poisson estimates. Most studies exclude intranational trade but those that include it find estimates of comparable magnitude to ours. For example, Wei (1996) estimates $(1-\sigma_M)\,\delta_{M2} = -2.27$ for a sample of OECD countries between 1982 and 1994, compared to $(1-\sigma_M)\,\delta_{M2} = -1.99$ in our estimation.

[29] Again, also see Appendix C for further details on the ICP data. We have also experimented with using averages across the three or five richest/poorest countries, with similar simulation results. A significant downside of using more countries is, however, that the resulting higher estimates of $\underline{A}$ implied that we needed to drop countries from the data for which the subsistence condition of the model ($\underline{A} < \theta_{Al}$) was violated.

[30] Availability of expenditure and price data from the ICP prevents us from generating data for earlier years. In particular, while the 2005 wave of the ICP has information on up to 191 countries (both developed and developing), this declines to 115 countries in 1996, and to only 56 and 61 countries in 1985 and 1980, respectively. In addition, developed countries are overrepresented in the earlier years, further reducing data availability for the purpose of the study. The availability of productivity and workforce data in agriculture and manufacturing also worsens as we got back in time, although not as dramatically as for the ICP data.

in the regressions from Section 2.[31]

Table 4 presents regression results using our generated data which yield a similar picture as our earlier results using actual data.[32] The coefficient on centrality is positive and significant in all specifications. Likewise, relative productivity is never significantly negative. Similar to Table 2, it has a positive impact on industrialization in columns 1 and 5, but loses its significance as soon as we control for agricultural productivity. Thus, we replicate the basic findings highlighted in the introduction and in Section 2.

Tables 5 and 6 report a number of robustness checks. We first demonstrate that augmenting the model by a third, non-tradable sector (which can be thought of as services) does not change our previous results. We now model the representative individual's preferences from country j as

$$U_j = \alpha \ln(M_j - \underline{M}) + \beta \ln(A_j - \underline{A}) + (1 - \alpha - \beta) \ln S_j,$$

where A_j and M_j are defined as before, and $S_j = s_j$ is the locally produced services good. Similar to agricultural and manufacturing varieties, services are produced using only labor with linear production technology $s_l = \theta_{Sl} L_{Sl}$ (where θ_{Sl} is labor productivity in services in country l). In Appendix D, we provide a more detailed exposition of the model, the resulting equilibrium conditions and analytical examples comparable to Section 4. As we show there, allowing for a service sector in the model complicates the analysis somewhat but the qualitative results go through as before.[33]

Regarding the calibration of this augmented model, note that since the third sector is non-tradable and non-differentiated, we only require new estimates for α, $\underline{A}$, β, and $\underline{M}$ (see Appendix D for the modified procedure for obtaining them). In Table 5, we present the same set of regressions results as in Table 4, this time using the calibrated version of the three-sector model to generate our synthetic dataset. As seen, the results are both qualitatively and quantitatively very similar to before. We conclude that allowing for an additional non-tradable sector does not change our previous conclusions and we work with the initial two-sector model for the rest of this paper.[34]

---

[31] Note that in the regressions on generated data, we use value added per worker as regressors to ensure comparability with the results from Section 2. Price-adjusted productivity levels are only used in the calibration itself, for consistency with the theoretical model in which θ_{Al} and θ_{Ml} represent physical rather than measured productivity. Thus, by using data on price differences, we are effectively allowing for a more general price-setting mechanism in the data than in our model (also see footnote 26).

[32] Note that the set of countries is the slightly different in Tables 1, 2 and 4 because of different data requirements. For generating our artificial data, we require the same independent variables as in Tables 1 and 2, but also employment in agriculture and manufacturing to compute workforce sizes (L_j). On the other hand, we do not need data on manufacturing shares as before. Running regressions on actual and generated data for the 76 countries present in both samples yields very similar results to Tables 1, 2 and 4 (available from the authors).

[33] Specifically, we show that examples 1 and 2 above carry through in the presence of a third nontraded sector. Example 3, for which we do not provide a clear-cut result in the two-sector case, becomes even more difficult in the three-sector case. The added complication arises for the following reason: in a two-sector environment, if manufacturing expands, agriculture shrinks, whereas this need no longer be the case if there is a third sector.

[34] The three-sector model also allows for an interesting extension of our data and results. Since we are now modelling the services sector as well, our model should be better suited to model the sectoral structure of developed countries as well. A priori, there is no reason to believe that comparative advantage or centrality should play a lesser role in determining manufacturing shares for this group of countries. In Appendix Tables A.2 and A.3, we thus present results for the full set of countries for which we have data (both developing and developed). Table A.2 uses actual data, while Table A.3 uses the data just generated by our three sector model. Again, both sets of

In Table 6, we report a number of additional robustness checks for the two-sector model. The first three columns use our alternative set of substitution elasticity estimates ($\sigma_M = 2.0$ and $\sigma_A = 2.6$). In columns 4-6, we use ordinary least squares to estimate equation (27), leading to alternative estimates for δ_{1M}, δ_{1A}, δ_{2M}, and δ_{2A}. Finally, in columns 7-9 we use producer prices rather than consumer prices to deflate relative productivities (see Appendix C). As shown, none of these changes alters the basic message from Table 4. Centrality is positive and significant throughout. Relative productivity is positive and significant when we do not control for agriculture productivity, and it is always insignificant when we do.[35]

In Table 7 we compare our preferred calibration with positive but finite trade cost (see Table 4) to the cases of free trade and autarky (corresponding to zero and infinite international trade costs in terms of the underlying calibration, respectively). As already noted, most of the existing models in the literature are based on one of these two polar scenarios. Our comparison uses two criteria. First, can the model replicate the qualitative correlations found in the data between comparative advantage and centrality, on the one hand, and manufacturing shares, on the other hand? Second, how well do all three parameterizations do in terms of replicating actual manufacturing shares? To evaluate this second criterion, we regress actual on simulated manufacturing shares, and look at the sign and significance of the corresponding regression coefficient, as well as at the associated R^2.

Looking at free trade first, we see that the model's performance in this case is dismal with respect to both criteria (see columns 6-9; columns 1-3 replicate our baseline results for convenience). The coefficient on comparative advantage is, as expected, negative and strongly significant, whereas the one on centrality is slightly negative and insignificant. The regression coefficient from the regression of actual on simulated manufacturing shares is positive but insignificant, and the corresponding R^2 close to zero (see the last two lines of the table). The model's performance with infinitely high trade costs is somewhat better, in the sense that it can replicate the facts related to relative productivity (column 4-6). However, the coefficient on centrality is insignificant throughout.[36] Also, while the correlation between actual and predicted manufacturing shares is positive and highly significant, and the R^2 substantially higher than in the free-trade case, both measures are lower than the ones generated by our baseline parameterization (see column 1-3). We conclude that allowing for positive but finite trade cost is necessary to replicate the stylized facts discussed in the introduction, and also improves the fit of actual and predicted levels of industrialization.

---

results are similar, confirming that the model also performs well when applied to all countries.

[35] For conciseness, we omit the specifications also including population and *per-capita* GDP. Results are again similar to those for our baseline calibration shown in Table 4 (available from the authors upon request).

[36] The fact that the coefficient on centrality is not exactly zero under autarky is of course due to functional form misspecification, given that the true data generating process in the model is more complicated than the simple log-linear relationship postulated in our regression tables throughout. For future research, it would be interesting to investigate whether using functional forms directly implied by the model have higher explanatory power in the actual data as well. We note, however, that this does not invalidate our earlier comparisons based on log-linear specifications as the issue of functional form misspecification applies to both actual and generated data. If the underlying data generating process was similar in both samples, we would expect the same log-linear approximation to yield similar results (as indeed it does).

6 Conclusion

In this paper, we have drawn attention to two cross-sectional facts which, taken together, are not easily explained by existing models of models of industrialization. First, proximity to foreign sources of demand seems to matter for levels of industrialization. That is, there is a positive correlation between manufacturing shares and the 'centrality' of a country, *i.e.* its closeness to foreign markets for its products. By construction, closed-economy models of industrialization are not suited to explain this fact. We also noted that measures of centrality have substantial explanatory power, explaining a comparable share of the cross-country variation in manufacturing shares as one of the central explanatory variables in the literature, *per-capita* income.

While our first stylized fact seems to point to the importance of open-economy models, our second fact suggests the opposite. Specifically, a standard proxy for Ricardian comparative advantage in agriculture (labor productivity in agriculture relative to manufacturing) is not or even positively correlated with manufacturing shares. This contradicts key predictions of open-economy models which predict that countries integrated through trade should specialize according to their comparative advantages.

We have argued that to understand these facts, we need to move beyond the closed-*versus*-open-economy dichotomy prevalent in the literature, and to consider multi-country settings in which countries interact with each other through international trade, but in which this interaction is partly hampered by the fact that trade is not costless. We constructed a simple model along these lines and used analytical examples and a full-scale multi-country calibration to show that it can replicate our stylized facts.

References

[1] Bernard, A.B., J. Eaton, J.B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," American Economic Review, 93(4), 1268-1290.

[2] Breinlich, H. (2006): "The Spatial Income Structure in the European Union – What Role for Economic Geography?," Journal of Economic Geography, Nov. 2006, Vol. 6, No. 5.

[3] Broda, C., J. Greenfield, and D.E. Weinstein (2006): "From Groundnuts to Globalization: A Structural Estimate of Trade and Growth," NBER Working Paper 13041.

[4] Broda, C. and D.E. Weinstein (2006): "Globalization and the Gains from Variety," Quarterly Journal of Economics, Volume 121, Issue 2, May.

[5] Caselli, F. and Coleman, W.J. (2001): "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation," Journal of Political Economy, 109, 584-616.

[6] Crafts, N. F. R. (1980): “Income Elasticities of Demand and the Release of Labor by Agriculture during the British Industrial Revolution: A further appraisal,” Journal of European Economic History, 9, 153-168.

[7] Davis, Donald R. (1998): “The Home Market, Trade, and Industrial Structure,” American Economic Review, 88(5), 1264-1276.

[8] Davis, Donald R. and David E. Weinstein (1996): “Does Economic Geography Matter for International Specialization?,”, NBER Working Paper 5706.

[9] Davis, Donald R. and David E. Weinstein (1998): “Market Access, Economic Geography and Comparative Advantage: An Empirical Assessment,” NBER Working Paper 6787.

[10] Davis, Donald R. and David E. Weinstein (2003): “Market Access, Economic Geography, and Comparative Advantage: An Empirical Test,” Journal of International Economics, 59(1), pp. 1-23.

[11] Deardorff, Alan V. (2004): “Local Comparative Advantage: Trade Costs and the Pattern of Trade”, Discussion Paper No. 500, University of Michigan.

[12] Disdier, Anne-Celia and Keith Head (2008): “The Puzzling Persistence of the Distance Effect on Bilateral Trade,” Review of Economics and Statistics 90, 37-41.

[13] Eaton, J., S. Kortum and F. Kramarz (2008): “An Anatomy of International Trade: Evidence from French Firms,” NBER Working Paper No. 14610.

[14] Echevarria, Christina (1997): “Changes in Sectoral Composition Associated with Economic Growth,” International Economic Review, 38(2), 431-452.

[15] Feenstra, R. (1994): “New Product Varieties and the Measurement of International Prices,” American Economic Review, 84(1), 157-177.

[16] Feenstra, R., J. Romalis, and P. Schott (2002): “U.S. Imports, Exports and Tariff Data, 1989-2001,” mimeo, University of California, Davis.

[17] Feenstra, R., R. Lipsey, H. Deng, A. Ma, and H. Mo (2005): “World Trade Flows, 1962-2000,” NBER Working Paper 11040.

[18] Galor, O. and D. Weil (2000): “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond,” American Economic Review, 90, 806-828.

[19] Golub, Stephen S. and Hsieh, Chang-Tai (2000): “Classical Ricardian Theory of Comparative Advantage Revisited,” Review of International Economics, 8(2), 221-234.

[20] Goodfriend, M. and J. McDermott (1995): “Early Development,” American Economic Review 85, 116-133.

[21] Hanson, G. and C. Xiang (2004): "The Home Market Effect and Bilateral Trade Patterns," American Economic Review, 94 (4), pp. 1108-1129.

[22] Harris, J. (1954): "The market as a factor in the localization of industry in the United States" Annals of the Association of American Geographers, 64, 315-348.

[23] Head, K. and T. Mayer (2006): "Regional Wage and Employment Responses to Market Potential in the EU," Regional Science and Urban Economics 36(5), 573-595.

[24] Kongsamut, P., S. Rebelo and D. Xie (2001): "Beyond Balanced Growth," Review of Economic Studies, 68, 869-882.

[25] Krugman, Paul R. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," American Economic Review, 70, 950-959.

[26] Krugman, Paul R. and E. Helpman (1985): "Market Structure and Foreign Trade," MIT Press, Cambridge, MA.

[27] Laitner, John (2000): "Structural Change and Economic Growth," Review of Economic Studies, 67, 545-561.

[28] Leamer, Edward E. (1987): "Paths of Development in the Three-factor, n-good General Equilibrium Model," Journal of Political Economy, 95(5), 961-999.

[29] Matsuyama, K. (1992): "Agricultural Productivity, Comparative Advantage, and Economic Growth," Journal of Economic Theory, 58, 317-334.

[30] Matsuyama, K. (2009): "Structural Change in an Interdependent World: A Global View of Manufacturing Decline," Journal of the European Economic Association, Papers and Proceedings.

[31] Murphy, Kevin M., A. Shleifer and R.V. Vishny (1989a): "Industrialization and the Big Push," Journal of Political Economy, 97(5), 1003-1026.

[32] Murphy, Kevin M., A. Shleifer and R.V. Vishny (1989b): "Income Distribution, Market Size, and Industrialization," Quarterly Journal of Economics, 104(3), 537-564.

[33] Ngai, L.R. and C.A. Pissarides (2007): "Structural Change in a Multisector Model of Growth," American Economic Review, 97, 429-443.

[34] Parente, Stephen L. and Edward C. Prescott (1994): "Barriers to Technology Adoption and Development," Journal of Political Economy, 102, 298-321.

[35] Parente, Stephen L. and Edward C. Prescott (2000): "Barriers to Riches," MIT Press, Cambridge, MA.

[36] Puga, D. and A.J. Venables (1996): "The spread of industry: spatial agglomeration in economic development," Journal of the Japanese and International Economy, 10, 440-464.

[37] Puga, D. and A.J. Venables (1999): "Agglomeration and Economic Development: Import Substitution vs. Trade Liberalisation," Economic Journal, 109, 292-311.

[38] Redding, S. and A.J. Venables (2004): "Economic Geography and International Inequality," Journal of International Economics, 62, 53-82.

[39] Rosenstein-Rodan, Paul N. (1943): "Problems of Industrialization of Eastern and South-Eastern Europe," Economic Journal, 53, 202-211.

[40] Santos-Silva, João M.C. and Silvana Tenreyro (2005): "The Log of Gravity," Review of Economics and Statistics, 88, 641-658.

[41] Schott, P. (2003): "One Size Fits All? Heckscher-Ohlin Specialization in Global Production," American Economic Review, 686-708.

[42] Syrquin, M. and H. Chenery (1989): "Patterns of Development, 1950 to 1983", World Bank Discussion Paper 41.

[43] Ventura, J. (1997): "Growth and interdependence," Quarterly Journal of Economics, 112, 57-84.

[44] Wei, S.J. (1996): "Intra-National versus International Trade: How Stubborn are Nations in Global Integration?," NBER Working Paper 5531.

[45] Yi, K.M. and J. Zhang (2010): "Structural Change in an Open Economy," mimeo, Federal Reserve Bank of Philadelphia.

A Appendix A: Country Lists and Data used in Cross-Country Regressions

Country List I - Cross-Country Regressions Albania (ALB); Algeria (DZA); Angola (AGO); Argentina (ARG); Armenia (ARM); Azerbaijan (AZE); Bangladesh (BGD); Barbados (BRB); Belarus (BLR); Belize (BLZ); Benin (BEN); Bhutan (BTN); Bolivia (BOL); Botswana (BWA); Brazil (BRA); Bulgaria (BGR); Burkina Faso (BFA); Burundi (BDI); Cambodia (KHM); Cameroon (CMR); Cape Verde (CPV); Central African Republic (CAF); Chad (TCD); Chile (CHL); China (CHN); Colombia (COL); Comoros (COM); Congo, Dem. Rep. (ZAR); Congo, Rep. (COG); Costa Rica (CRI); Cote d'Ivoire (CIV); Croatia (HRV); Czech Republic (CZE); Dominican Republic (DOM); Ecuador (ECU); Egypt, Arab Rep. (EGY); El Salvador (SLV); Equatorial Guinea (GNQ); Eritrea (ERI); Estonia (EST); Ethiopia (ETH); Fiji (FJI); Gabon (GAB); Gambia, The (GMB); Georgia (GEO); Ghana (GHA); Grenada (GRD); Guatemala (GTM); Guinea (GIN); Guinea-Bissau (GNB); Guyana (GUY); Honduras (HND); Hungary (HUN); India (IND); Indonesia (IDM); Iran, Islamic Rep. (IRN); Jamaica (JAM); Jordan (JOR); Kazakhstan (KAZ); Kenya (KEN); Korea, Rep. (KOR); Kyrgyz Republic (KGZ); Lao PDR (LAO); Latvia (LVA); Lesotho (LSO); Lithuania (LTU); Macedonia, FYR (MKD); Madagascar (MDG); Malawi (MWI); Malaysia (MYS); Mali (MLI); Mauritania (MRT); Mexico (MEX); Moldova (MDA); Mongolia (MNG); Morocco (MAR); Mozambique (MOZ); Namibia (NAM); Nepal (NPL); Nicaragua (NIC); Niger (NER); Nigeria (NGA); Oman (OMN); Pakistan (PAK); Panama (PAN); Papua New Guinea (PNG); Paraguay (PRY); Peru (PER); Philippines (PHL); Poland (POL); Romania (ROM); Rwanda (RWA); Saudi Arabia (SAU); Senegal (SEN); Sierra Leone (SLE); Slovak Republic (SVK); South Africa (ZAF); Sri Lanka (LKA); St. Lucia (LCA); St. Vincent and the Grenadines (VCT); Sudan (SDN); Suriname (SUR); Syrian Arab Republic (SYR); Tanzania (TZA); Thailand (THA); Togo (TGO); Trinidad and Tobago (TTO); Tunisia (TUN); Turkey (TUR); Uganda (UGA); Ukraine (UKR); Uruguay (URY); Uzbekistan (UZB); Venezuela, RB (VEN); Vietnam (VNM); Yemen, Rep. (YEM); Zambia (ZMB); Zimbabwe (ZWE).

Country List II - Calibration Sample **Developing countries:** Albania (ALB); Algeria (DZA); Argentina (ARG); Armenia (ARM); Azerbaijan (AZE); Bangladesh (BGD); Barbados (BRB); Belize (BLZ); Bolivia (BOL); Botswana (BWA); Brazil (BRA); Bulgaria (BGR); Cambodia (KHM); Cameroon (CMR); Chile (CHL); China (CHN); Colombia (COL); Costa Rica (CRI); Croatia (HRV); Czech Republic (CZE); Dominican Republic (DOM); Ecuador (ECU); Egypt, Arab Rep. (EGY); El Salvador (SLV); Estonia (EST); Georgia (GEO); Ghana (GHA); Guatemala (GTM); Guyana (GUY); Honduras (HND); Hungary (HUN); Indonesia (IDM); Jamaica (JAM); Jordan (JOR); Kazakhstan (KAZ); Korea, Rep. (KOR); Kyrgyz Republic (KGZ); Latvia (LVA); Lesotho (LSO); Lithuania (LTU); Macedonia, FYR (MKD); Malaysia (MYS); Mexico (MEX); Moldova (MDA); Mongolia (MNG); Morocco (MAR); Namibia (NAM); Nepal (NPL); Nicaragua (NIC); Oman (OMN); Pakistan (PAK); Panama (PAN); Papua New Guinea (PNG); Paraguay (PRY); Peru (PER); Philippines (PHL); Poland (POL); Romania (ROM); Russian Federation (RUS); Saudi Arabia (SAU); Slovak Republic (SVK); South Africa (ZAF); Sri Lanka (LKA); St. Lucia (LCA); St. Vincent and the Grenadines (VCT); Suriname (SUR); Syrian Arab Republic (SYR); Tanzania (TZA); Thailand (THA); Trinidad and Tobago (TTO); Turkey (TUR); Ukraine (UKR); Uruguay (URY); Uzbekistan (UZB); Venezuela, RB (VEN); Vietnam (VNM); Yemen, Rep. (YEM); Zambia (ZMB); Zimbabwe (ZWE).

Developed countries: Australia (AUS); Austria (AUT); Belgium (BEL); Brunei Darussalam (BRN); Canada (CAN); Cyprus (CYP); Denmark (DKF); Finland (FIN); France (FRA);

Germany (DEU); Greece (GRC); Iceland (ISL); Ireland (IRL); Italy (ITA); Japan (JPN); Luxembourg (LUX); Netherlands (NLD); New Zealand (NZL); Norway (NOR); Portugal (PRT); Singapore (SGP); Slovenia (SVN); Spain (ESP); Sweden (SWE); Switzerland (CHE); United Arab Emirates (ARE); United Kingdom (GBR); United States (USA).

Data used in Cross-Country Regressions.

- Share of manufacturing value added in GDP: World Development Indicators (World Bank) and national statistics offices.
- Value added per worker in agriculture and manufacturing (in 2000 USD): World Development Indicators, United Nations Industrial Statistics Database (UNIDO), and national statistical offices.
- GDP, GNP and *per-capita* GDP (2000 USD): World Development Indicators.
- Population size: World Development Indicators.
- Bilateral distances between countries: CEPII Bilateral Distances Database.

B Appendix B: Estimating Substitution Elasticities

The demand side structure in Broda and Weinstein is very similar to ours. In particular, they define a composite imported good M_t which aggregates individual goods in a CES fashion:

$$M_t = \left(\sum_{g \epsilon G} (M_{gt})^{\left(\gamma_g - 1\right)/\gamma_g} \right)^{\gamma_g/\left(\gamma_g - 1\right)}$$

M_{gt} is the subutility derived from the consumption of imported good g at time t. Note that in our setting, we only have two such goods (the manufacturing and agriculture composite good) and that we assume a Cobb-Douglas rather than a CES aggregator. This does not matter in the following because we are interested in substitution elasticities at the next lower level of aggregation only. Similar to us, Broda and Weinstein assume that M_{gt} aggregates varieties differentiated by country of origin and that, in addition, the aggregator takes the following nonsymmetric CES form:

$$M_{gt} = \left(\sum_{c \epsilon C} d_{gct}^{1/\sigma_g} \left(m_{gct}\right)^{(\sigma_g - 1)/\sigma_g} \right)^{\sigma_g/(\sigma_g - 1)} \tag{30}$$

where σ_g is the elasticity of substitution among varieties of good g and d_{gct}^{1/σ_g} denotes a taste or quality parameter for a variety from country c.[37] Associated with this aggregator is the price index:

$$\Phi_{gt}^{M} = \left(\sum_{c \epsilon C} d_{gct} \left(p_{gct}\right)^{1-\sigma_g} \right)^{1/(1-\sigma_g)}$$

---

[37] Feenstra (1994, p.161) shows that allowing for quality differences is important to address the aggregation problem arising from the fact that we only observe unit values rather than prices in the trade data. One problem resulting from this is that we implicitly ignore changes in the number of varieties supplied from each exporting country (we have assumed this away for simplicity in our model but such changes are likely to be an important phenomenon in the data). However, Feenstra demonstrates that changes in the number of varieties are isomorphic to changes in the quality parameters d_{gct}, and thus captured by the error term ε_{gct} in the regression to be estimated below.

where p_{gct} is the price charged by country c for good g at time t. From (30), we can derive the following import demand function (expressed as import shares and in log-differences):

$$\Delta \ln s_{gct} = \varphi_{gt} - (\sigma_g - 1)\,\Delta \ln p_{gct} + \varepsilon_{gct}$$

where $\varphi_{gt} = (\sigma_g - 1)\ln\left(\Phi^M_{gt}/\Phi^M_{gt-1}\right)$ and $\varepsilon_{gct} = \Delta \ln d_{gct}$.

Broda and Weinstein also allow for an upward-sloping supply curve of the form:

$$\Delta \ln p_{gct} = \psi_{gt} + \frac{\omega_g}{1+\omega_g}\Delta \ln s_{gct} + \delta_{gct}$$

where $\omega_g \geq 0$ is the inverse supply elasticity, $\psi_{gt} = \frac{\omega_g}{1+\omega_g}\ln E_{gt}$, E_{gt} is total expenditure on good g at time t in the importing country, and $\delta_{gct} = \frac{1}{1+\omega_g}\Delta \ln \upsilon_{gct}$ captures random changes in the technology factor υ_{gct}. Crucially for the identification strategy below, Broda and Weinstein further assume that demand and supply shocks are independent, implying $E\left(\varepsilon_{gct}\delta_{gct}\right) = 0$.

Supply and demand can be rewritten to eliminate the intercepts φ_{gt} and ψ_{gt} by normalizing with respect to a reference country k:[38]

$$\varepsilon^k_{gct} = \Delta^k \ln s_{gct} + (\sigma_g - 1)\,\Delta^k \ln p_{gct}$$

$$\delta^k_{gct} = \Delta^k \ln p_{gct} - \frac{\omega_g}{1+\omega_g}\Delta^k \ln s_{gct}$$

where $\Delta^k \ln p_{gct} = \Delta \ln p_{gct} - \Delta^k \ln p_{gkt}$, etc. To take advantage of $E\left(\varepsilon_{gct}\delta_{gct}\right) = 0$, we multiply the two normalized equations and obtain:

$$\left(\Delta^k \ln p_{gct}\right)^2 = \theta_1\left(\Delta^k \ln s_{gct}\right)^2 + \theta_2\left(\Delta^k \ln p_{gct}\Delta^k \ln s_{gct}\right) + u_{gct} \tag{31}$$

with $\theta_1 = \frac{\omega_g}{(1+\omega_g)(\sigma_g-1)}$ and $\theta_2 = \frac{\omega_g(\sigma_g-2)-1}{(1+\omega_g)(\sigma_g-1)}$. Although $u_{gct} = \varepsilon_{gct}\delta_{gct}$ is correlated with shares and prices, we can obtain consistent estimates of θ_1 and θ_2 by implementing the following between estimator (averaging across periods t):

$$\left(\overline{\Delta^k \ln p_{gc}}\right)^2 = \theta_1\left(\overline{\Delta^k \ln s_{gc}}\right)^2 + \theta_2\left(\overline{\Delta^k \ln p_{gc}\Delta^k \ln s_{gc}}\right) + \overline{u}_{gc} \tag{32}$$

By the assumption of independence of ε_{gct} and δ_{gct}, we know that $E\left(\overline{u}_{gc}\right) = 0$ and thus $plim\left(\overline{u}_{gc}\right) = 0$ as the number of periods T approaches infinity. So the error term in (32) vanishes, solving the problem of correlation with the regressors. We estimate (32) using weighted least squares to obtain estimates for θ_1 and θ_2.[39] Using the definition of θ_1 and θ_2 above, we then solve for ω_g and σ_g.[40]

---

[38] We choose the reference country so that the number of usable observations is maximised (we need share price data for both country c and the reference country k). We use the U.S. and Canada as reference countries for the cross-country sample and the U.S. import sample, respectively.

[39] We follow Broda and Weinstein (2006, pp. 582-584) in adding an additional term inversely related to the quantity of imports from a given country on the right-hand side of (32) and in weighting the data so that the variances are more sensitive to price movements based on large import quantities than small ones. Broda and Weinstein show that this helps addressing problems arising from measurement error due to the use of unit values (rather than actual prices).

[40] If this approach produces economically infeasible estimates (i.e., $\sigma_g \leq 1$ or $\omega_g < 0$), Broda and Weinstein propose to do a grid search over a large set of feasible values. Fortunately, we did not encounter this problem in our estimation.

C Appendix C: Using ICP Data to Proxy for Prices

In Sections 2 and 5 we use data from the International Comparison Project (ICP) to strip out price variation from measured productivity. To understand this approach, note that the ICP provides data on a number of expenditure categories in both current U.S. dollars and so-called international dollars ($I). One $I is the amount of goods and services one U.S. dollar would purchase in the USA in the base period (2005 in our case as no data were available for 2000). Converting expenditure from current U.S. dollars into $I thus removes any price differences across countries and basically converts expenditures into quantities using implicit aggregators. By comparing local expenditures in U.S. dollars and international dollars, one can derive country-product-specific PPP exchange rates which capture price differences across country. For example, *per capita* expenditure on food in current U.S. dollars was $2,040 in 2005 in the United Kingdom but only 1,586 $I, yielding an implicit price of 1.29 (the price in the USA is normalized to 1). Dividing measured productivity levels ($p_{Ml}m_l/L_{Ml}$ and $p_{Al}a_l/L_{Al}$) by this price converts them into quantities per unit of labor used and thus into appropriate proxies for $\theta_{Ml} = m_l/L_{Ml}$ and $\theta_{Al} = a_l/L_{Al}$. We note that Echevarria (1997) uses a similar procedure, calculating proxies for agricultural and manufacturing prices by dividing expenditures in U.S. dollars by expenditures in international dollars.

One problem with the above approach is that we are implicitly using consumer prices rather than producer prices to deflate production. In terms of our model, ICP prices are proxies for P_{Ml} and P_{Al}, not p_{Ml} and p_{Al}. As a robustness check in section 5, we therefore use the definition of P_{Ml} and P_{Al} to extract information on p_{Ml} and p_{Al} in a model-consistent way. In our model,

$$P_{Mj} = \left[\sum_{l=1}^{R}\left(p_{Ml}T_{lj}^{M}\right)^{1-\sigma_M}\right]^{\frac{1}{1-\sigma_M}} \tag{33}$$

$$P_{Aj} = \left[\sum_{l=1}^{R}\left(p_{Al}T_{lj}^{A}\right)^{1-\sigma_A}\right]^{\frac{1}{1-\sigma_A}}, \tag{34}$$

Together with data on the elasticities of substitution and trade costs which we have obtained independently as part of our calibration strategy, we can solve the above system of equations for p_{Ml} and p_{Al}. In practice, consumer and implied producer prices are almost identical, with a correlation coefficient of above 99% and a level difference of on average less than 4% for manufacturing and less than 1% for agriculture.

D Appendix D: A Three-Sector Model

This appendix works out a three-sector model where the third sector, services, is assumed to be nontraded. We allow for non-homotheticicities in demand to affect the manufacturing sector, too, as this is has been considered in the literature relatively often. (See, for example, Matsuyama (2009).)

D.1 Demand Side

The individual's preferences are now

$$U_j = \alpha \ln(M_j - \underline{M}) + \beta \ln(A_j - \underline{A}) + (1-\alpha-\beta)\ln S_j, \tag{35}$$

with

$$A_j = \left[\sum_{l=1}^{R} a_{lj}^{(\sigma_A-1)/\sigma_A}\right]^{\frac{\sigma_A}{\sigma_A-1}} \tag{36}$$

$$M_j = \left[\sum_{l=1}^{R} m_{lj}^{(\sigma_M-1)/\sigma_M}\right]^{\frac{\sigma_M}{\sigma_M-1}}, \tag{37}$$

$$S_j = s_j, \tag{38}$$

where $\alpha, \beta, \alpha+\beta \in (0,1)$, $\underline{A} < \theta_{Al}$, $\underline{M} < \theta_{Ml}$, $\sigma_A, \sigma_M > 1$. The individual's budget constraint is

$$P_{Aj}A_j + P_{Mj}M_j + P_{Sj}S_j = w_j. \tag{39}$$

As we discuss below, total income equals labor income, as profits are zero. The price indices in the budget constraint are

$$P_{Aj} = \left(\sum_{l=1}^{R} p_{Alj}^{1-\sigma_A}\right)^{\frac{1}{1-\sigma_A}} = \left[\sum_{l=1}^{R} \left(p_{Al}T_{lj}^{A}\right)^{1-\sigma_A}\right]^{\frac{1}{1-\sigma_A}}, \tag{40}$$

$$P_{Mj} = \left(\sum_{l=1}^{R} p_{Mlj}^{1-\sigma_M}\right)^{\frac{1}{1-\sigma_M}} = \left[\sum_{l=1}^{R} \left(p_{Ml}T_{lj}^{M}\right)^{1-\sigma_M}\right]^{\frac{1}{1-\sigma_M}}, \tag{41}$$

$$P_{Sj} = p_{Sj}. \tag{42}$$

where $T_{lj}^{A}, T_{lj}^{M} \geq 1$, $T_{jj}^{A}, T_{jj}^{M} = 1$. Implicit here is the assumption that sector S is non-traded.

Aggregating across all individuals/countries yields the following demands for varieties (net of trade costs):

$$a_l = p_{Al}^{-\sigma_A}\sum_{j=1}^{R}\left(T_{lj}^{A}\right)^{-\sigma_A} P_{Aj}^{\sigma_A-1}E_{Aj}, \tag{43}$$

$$m_l = p_{Ml}^{-\sigma_M}\sum_{j=1}^{R}\left(T_{lj}^{M}\right)^{-\sigma_M} P_{Mj}^{\sigma_M-1}E_{Mj}, \tag{44}$$

$$s_l = p_{Sl}^{-1}E_{Sl}, \tag{45}$$

where

$$E_{Aj} = \left[\underline{A}P_{Aj} + \beta\left(w_j - \underline{A}P_{Aj} - \underline{M}P_{Mj}\right)\right]L_j, \tag{46}$$

$$E_{Mj} = \left[\underline{M}P_{Mj} + \alpha\left(w_j - \underline{A}P_{Aj} - \underline{M}P_{Mj}\right)\right]L_j, \tag{47}$$

$$E_{Sj} = \left[\left(1-\alpha-\beta\right)\left(w_j - \underline{A}P_{Aj} - \underline{M}P_{Mj}\right)\right]L_j. \tag{48}$$

D.2 Production

Goods are produced with linear technologies:

$$a_l = \theta_{Al}L_{Al}, \tag{49}$$

$$m_l = \theta_{Ml}L_{Ml}, \tag{50}$$

$$s_l = \theta_{Sl}L_{Sl}. \tag{51}$$

Perfect competition implies

$$p_{.l} = \frac{w_l}{\theta_{.l}}. \tag{52}$$

D.3 Equilibrium

Equilibrium in the goods markets yields

$$a_l = \theta_{Al}^{\sigma_A} w_l^{-\sigma_A} \left[\sum_{j=1}^{R} \left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A - 1} E_{Aj} \right], \tag{53}$$

$$m_l = \theta_{Ml}^{\sigma_M} w_l^{-\sigma_M} \left[\sum_{j=1}^{R} \left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M - 1} E_{Mj} \right], \tag{54}$$

$$s_l = \frac{\theta_{Sl}}{w_l} E_{Sl}. \tag{55}$$

Labor demand:

$$L_{Al} = \theta_{Al}^{\sigma_A - 1} w_l^{-\sigma_A} \left[\sum_{j=1}^{R} \left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A - 1} E_{Aj} \right], \tag{56}$$

$$L_{Ml} = \theta_{Ml}^{\sigma_M - 1} w_l^{-\sigma_M} \left[\sum_{j=1}^{R} \left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M - 1} E_{Mj} \right], \tag{57}$$

$$L_{Sl} = \frac{E_{Sl}}{w_l}. \tag{58}$$

Full employment requires

$$L_{Al} + L_{Ml} + L_{Sl} = L_l \tag{59}$$

or

$$\theta_{Al}^{\sigma_A - 1} w_l^{-\sigma_A} \left[\sum_{j=1}^{R} \left(T_{lj}^{A}\right)^{1-\sigma_A} P_{Aj}^{\sigma_A - 1} E_{Aj} \right] + \theta_{Ml}^{\sigma_M - 1} w_l^{-\sigma_M} \left[\sum_{j=1}^{R} \left(T_{lj}^{M}\right)^{1-\sigma_M} P_{Mj}^{\sigma_M - 1} E_{Mj} \right] + \frac{E_{Sl}}{w_l} = L_l, \tag{60}$$

These are R non-linear equations in the R wage rates.

D.4 Autarky

It is easy to show that

$$\frac{L_{Al}}{L_l} = \frac{E_{Al}}{w_l L_l} = \beta \left(1 - \frac{\underline{M}}{\theta_{Ml}}\right) + (1-\beta) \frac{\underline{A}}{\theta_{Al}}, \tag{61}$$

$$\frac{L_{Ml}}{L_l} = \frac{E_{Ml}}{w_l L_l} = \alpha \left(1 - \frac{\underline{A}}{\theta_{Al}}\right) + (1-\alpha) \frac{\underline{M}}{\theta_{Ml}}, \tag{62}$$

$$\frac{L_{Sl}}{L_l} = \frac{L_l - (L_{Al} + L_{Ml})}{L_l} = (1 - \alpha - \mu) \left(1 - \frac{\underline{A}}{\theta_{Al}} - \frac{\underline{M}}{\theta_{Ml}}\right). \tag{63}$$

D.5 Free Trade

With costless trade, and assuming $\sigma_A = \sigma_M = \sigma$, it is easy to show that

$$\frac{L_{Ml}/L_{Al}}{L_{Ml'}/L_{Al'}} = \left(\frac{\theta_{Ml}/\theta_{Al}}{\theta_{Ml'}/\theta_{Al'}}\right)^{\sigma-1}. \tag{64}$$

The relative share of the services sector depends positively on the country's wage (as long as the effect of w_l on P_{Al} and P_{Ml} is assumed negligible):

$$\frac{L_{Sl}}{L_l} = \frac{(1-\alpha-\beta)\left(w_l - \underline{A}P_{Al} - \underline{M}P_{Ml}\right)}{w_l}. \tag{65}$$

D.6 Costly Trade

Example 1 with three sectors Consider a three-country world, $R = 3$, and a geographic structure such that country 1 takes a "central" position while countries 2 and 3, which are fully symmetric, are in the "periphery": we model this by assuming that country 1 can trade with both 2 and 3 at positive but finite trade costs ($T_{12} = T_{21} = T_{13} = T_{31} = T > 1$) and that countries 2 and 3 cannot trade with one another ($T_{23} = T_{32} = \infty$). Trade costs are assumed equal across industries. We simplify further by assuming $\sigma_M = \sigma_A = \sigma$. Finally, assume all parameters are identical across countries (except for the bilateral trade costs) and, in particular, that $\theta_{Aj} = \theta_{Mj} = \theta_{Sj} = L_j = 1$ and $\underline{M}= 0$. Profiting from the symmetry we have imposed, let us normalize $w_2 = w_3 = 1$.

The results discussed in example 1 above apply here as well.

Example 2 with three sectors Assume all parameters are identical across countries (except for the bilateral trade costs) and, in particular, that $\theta_{Aj} = \theta_{Mj} = \theta_{Sj} = L_j = 1$, $\sigma_A = \infty$, and $\sigma_M > 1$ but finite. We simplify further by assuming $\underline{A}=\underline{M}= 0$. Again, we consider a geographic structure such that country 1 takes a "central" position while countries 2 and 3 are in the "periphery": here we model this by assuming that country 1 can trade freely with both 2 and 3 ($T_{12} = T_{21} = T_{13} = T_{31} = 1$) and that countries 2 and 3 cannot trade with one another ($T_{23} = T_{32} = \infty$). Trade costs are equal across sectors here. We take the agricultural good as the numéraire.

The results discussed in example 2 above apply here as well with small variations. First, it is easy to show that $L_{Sj} = 1 - \alpha - \beta$ for all countries. Second, it is easy to show as well that $L_{M1} > L_{M2} = L_{M3}$. Third, if parameter values yield $L_{A1} = 0$, then the labor market equilibrium conditions yield $w_1 > w_2 = w_3 = 1$.

D.7 Calibration

Since the third sector is non-tradable and non-differentiated, we do not require new estimates for σ_M, σ_A, δ_{1M}, δ_{1A}, δ_{2M}, and δ_{2A} (note that the expression for manufacturing and agricultural exports in the model is the same as in the two sector version). However, we do require new estimates for α, $\underline{A}$, β, and $\underline{M}$ since β and $\underline{M}$ are new parameters and the meaning of α and $\underline{A}$ has changed due to the introduction of the third sector.

To obtain estimates of these new parameters, we follow our earlier approach to use expenditure shares and food consumption for the richest and poorest country in our data, respectively. To see this, note that expenditure shares in the new model are given by:

$$\begin{aligned}
S_{EMj} &= \frac{E_{Mj}}{w_j L_j} = \alpha - \alpha \underline{A} \frac{P_{Aj}}{w_j} + (1-\alpha)\, \underline{\mathrm{M}} \frac{P_{Mj}}{w_j}, \\
S_{EAj} &= \frac{E_{Aj}}{w_j L_j} = \beta + (1-\beta)\, \underline{A} \frac{P_{Aj}}{w_j} - \beta \underline{\mathrm{M}} \frac{P_{Mj}}{w_j} \\
S_{ESj} &= \frac{E_{Sj}}{w_j L_j} = (1-\alpha-\beta) - (1-\alpha-\beta)\, \underline{A} \frac{P_{Aj}}{w_j} - (1-\alpha-\beta)\, \underline{\mathrm{M}} \frac{P_{Mj}}{w_j}
\end{aligned}$$

As $w_j \to \infty$, $S_{EMj} \to \alpha$, $S_{EAj} \to \beta$ and $S_{ESj} \to (1-\alpha-\beta)$. Thus, a suitable proxy for α and β are the expenditure shares of the richest country in the data.

Likewise, note that as $w_j \to (\underline{A} P_{Aj} + \underline{\mathrm{M}} P_{Mj})$, consumption per head convergence to the agricultural and manufacturing subsistence levels:

$$\begin{aligned}
\lim_{w_j \to \left(\underline{A} P_{Aj} + \underline{\mathrm{M}} P_{Mj}\right)} \frac{E_{Aj}}{L_j P_{Aj}} &= \lim_{w_j \to \left(\underline{A} P_{Aj} + \underline{\mathrm{M}} P_{Mj}\right)} \left(\underline{A} + \alpha P_{Aj}^{-1} w_j - \alpha \underline{A} - \alpha \underline{\mathrm{M}} P_{Mj} P_{Aj}^{-1} \right) = \underline{A}, \\
\lim_{w_j \to \left(\underline{A} P_{Aj} + \underline{\mathrm{M}} P_{Mj}\right)} \frac{E_{Mj}}{L_j P_{Mj}} &= \lim_{w_j \to \left(\underline{A} P_{Aj} + \underline{\mathrm{M}} P_{Mj}\right)} \left(\underline{\mathrm{M}} + \mu P_{Mj}^{-1} w_j - \mu \underline{A} P_{Aj} P_{Mj}^{-1} - \mu \underline{\mathrm{M}} \right) = \underline{\mathrm{M}},
\end{aligned}$$

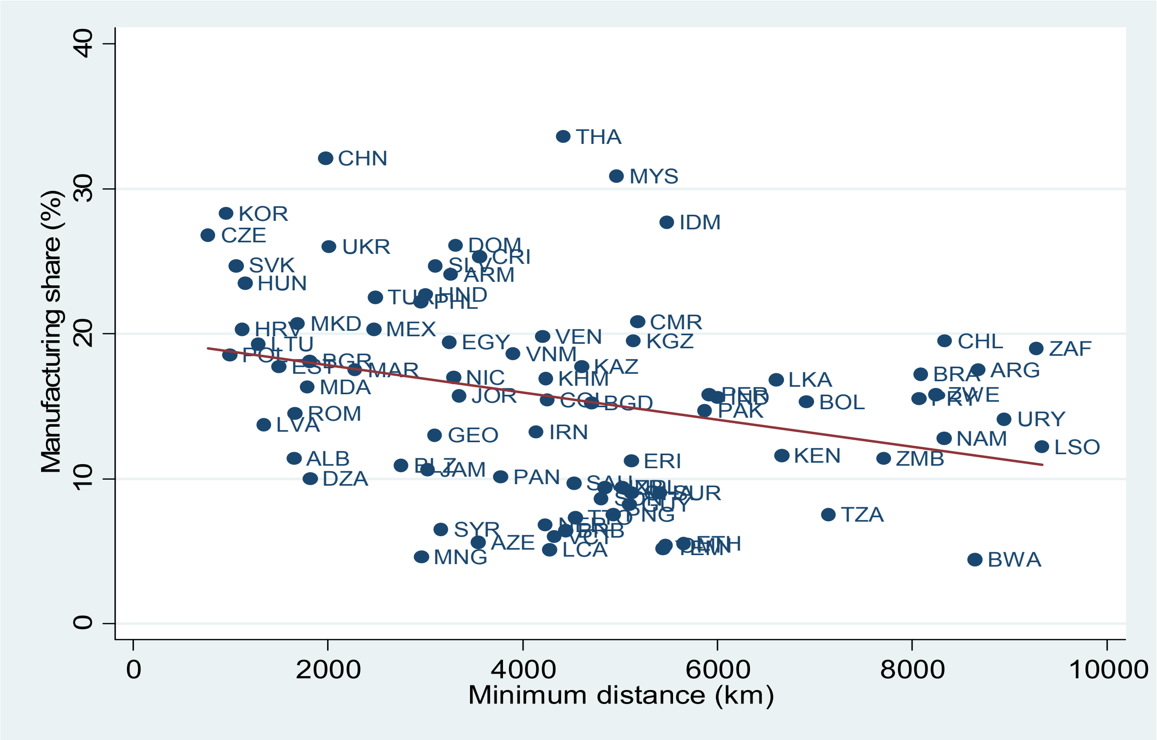
Since $w_j = \underline{A} P_{Aj} + \underline{\mathrm{M}} P_{Mj}$ is the income level which guarantees that the substistence level is just attainable, a suitable proxy for $\underline{A}$ and $\underline{\mathrm{M}}$ are real expenditure per worker in the poorest country in our data (measured in \$I).[41]

---

[41] Note that since we define "rich" and "poor" as total expenditure per worker (which is consistent with our model), the ranking of countries changes slightly with the introduction of a third sector (services expenditure is now taking into account in the definition of income). That is, the poorest country according to a new definition is now Tanzania, explaining the increase in the estimate for $\underline{A}$ in Table 2.

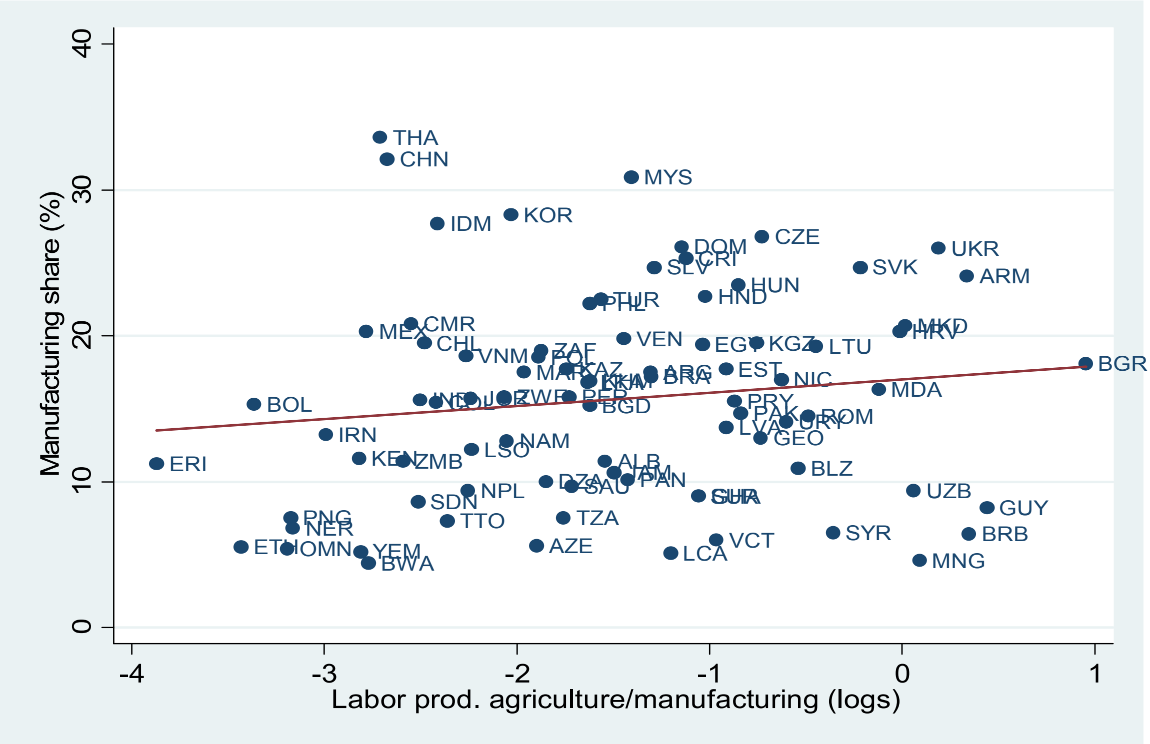
Figures and Tables

Figure 1: GDP manufacturing sharesand minimum distance to main markets (2000)



Notes: Figure plots manufacturing shares in GDP (in %) against the minimum distance (in km) to either of the U.S., the European Union (Netherlands), or Japan. All dataare for 2000. See Appendix A for data sources and country codes.

Figure 2: GDP manufacturing shares and relative productivities (2000)



Notes: Figure plots manufacturing shares in GDP (in %) against the ratio of labor productivity in agriculture and manufacturing. All data are for 2000. See Appendix A for data sources and country codes.

Table 1: Baseline Empirical Results (Developing Countries Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Regressor	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM
log(RELPR)	0.0862			0.0850	0.130**		0.113**	
	(0.0686)			(0.0630)	(0.0633)		(0.0454)	
log(mindist)		-0.332***						
		(0.0754)						
log(CEN)			0.417***			0.462***		0.547***
			(0.126)			(0.119)		(0.133)
Fixed Effects	--	--	--	--	SE-Asia Dummy	SE-Asia Dummy	Year	Year
Years	2000	2000	2000	2000	2000	2000	1980, 1990, 2000	1980, 1990, 2000
Observations	83	83	83	83	83	83	256	256
R-squared	0.023	0.118	0.073	0.026	0.236	0.274	0.045	0.093

Notes: Table displays coefficients and robust standard errors (clustered at the country-level in columns 7-8) for OLS estimations. The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. Mindist is the minimum distance (in km) of a country to either of Japan, the European Union (Netherlands) or the USA. CEN is a country's centrality measure (defined in Section 2). All regressorsare in logs. Results on the included constant are suppressed. For data sources see Appendix A. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table 2: Extended Empirical Results (Developing Countries Only)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Regressor	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM
log(RELPR)	0.130*	-0.0220	0.0698	0.127*	0.140***	-0.0348	0.0279	0.0972**			
	(0.0680)	(0.0805)	(0.0727)	(0.0686)	(0.0449)	(0.0540)	(0.0519)	(0.0459)			
log(CEN)	0.354***	0.355**	0.287**	0.253**	0.413***	0.370**	0.281**	0.298**	0.400***	2.588***	2.930***
	(0.128)	(0.152)	(0.121)	(0.123)	(0.129)	(0.142)	(0.113)	(0.114)	(0.137)	(0.881)	(0.980)
log(POP)	0.177***		0.186***	0.196***	0.171***		0.181***	0.187***	0.186***	0.953***	0.602
	(0.0267)		(0.0279)	(0.0289)	(0.0247)		(0.0235)	(0.0233)	(0.0225)	(0.265)	(0.481)
log(AP)		0.0745	0.124**			0.190***	0.220***				
		(0.0667)	(0.0613)			(0.0655)	(0.0557)				
log(PCGDP)				0.139**				0.197***	0.190***	0.289***	0.0854
				(0.0534)				(0.0523)	(0.0469)	(0.101)	(0.170)
Fixed Effects	--	--	--	--	Year	Year	Year	Year	Year	Year, Country	Long First Difference
Years	2000	2000	2000	2000	1980, 1990, 2000	1980, 1990, 2000	1980, 1990, 2000	1980, 1990, 2000	1980-2005	1980-2005	1980-2005
Observations	83	83	83	83	256	256	256	256	2,977	2,977	73
R-squared	0.319	0.086	0.353	0.380	0.305	0.175	0.399	0.398	0.340	0.877	0.145

Notes: Table displays coefficients and robust standard errors (clustered at the country-level in columns 5-11) for OLS estimations. The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. Mindist is the minimum distance (in km) of a country to either of Japan, the European Union (Netherlands) or the USA. CEN is a country's centrality measure (defined in Section 2). POP is a country's population size, AP its labor productivity in agriculture and PCGDP its per-capita GDP, respectively. All regressorsare in logs. Results on the included constant are suppressed. For data sources see Appendix A. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table 3: Calibrated Parameter Values

Parameter	Value (Baseline)	Value (Robustness)	Outline of Calibration Procedure	Data sources
σ_A	2.3	2.6	Estimated on cross-country (baseline) and U.S. (robustness) data on import quantities and prices for the year 2000, following Broda and Weinstein (2006) and Broda, Greenfield and Weinstein (2006).	UN-NBER (Feenstra et al., 2005)
σ_M	2.3	2.0		US-NBER (Feenstra et al., 2002)
θ_{Al}, θ_{Ml}	country- and sector-specific		Labor productivity in manufacturing and agriculture, corrected for sector specific price differences.	UNIDO, WDI, ICP
$T_{lj}=e^{\delta_{1}*d(int)}dist_{lj}^{\delta_2}$	$\delta_{1A}=2.42, \delta_{1M}=1.53,$ $\delta_{2A}=0.74, \delta_{2M}=0.69$	$\delta_{1A}=3.08, \delta_{1M}=1.17,$ $\delta_{2A}=1.05, \delta_{2M}=1.42$	Coefficients on bilateral distance and internal trade flow dummies from gravity equation estimations (baseline: Poisson QML; robustness: OLS).	NBER, FAO, CEPII
A, α	$A=170$\$I/year, $\alpha=0.81$		Manufacturing expenditure share of richest country (α) and food expenditure per worker of the poorest country (A) in data.	ICP
A, M; α, β	$A=285$\$I/year, $\beta=0.07$ $M=100$\$I/year, $\alpha=0.29$		3-sector model only. Manufacturing and agricultural expenditure share of richest country (α and β) and food and manufacturing expenditure per worker of the poorest country (A and M).	ICP

Notes: Table shows parameter estimates used for the model calibration in Section 5. Also listed are outlines of the calibration procedures and the data sources used (see Section 5 and Appendices B, C and D for details). $I denotes international dollars.

Table 4: Results for Generated Data (Baseline)

Regressor	(1) ltShareM	(2) ltShareM	(3) ltShareM	(4) ltShareM	(5) ltShareM
log(RELPR)	0.206***		-0.027	-0.023	0.319***
	(0.048)		(0.038)	(0.039)	(0.041)
log(CEN)		0.516***	0.187*	0.181*	0.229**
		(0.156)	(0.097)	(0.093)	(0.088)
log(AP)			0.395***	0.399***	
			(0.047)	(0.051)	
log(POP)				0.007	0.187***
				(0.016)	(0.032)
log(PCGDP)					0.665***
					(0.079)
Observations	79	79	79	79	79
R-squared	0.176	0.145	0.721	0.722	0.801

Notes: Table displays coefficients and robust standard errors for OLS estimations using generated data (see Section 5 for details). The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. CEN is a country's centrality measure (defined in Section 2). POP is a country's population size, AP its labor productivity in agriculture, and PCGDP its per-capita GDP, respectively. All regressorsare in logs. Results on the included constant are suppressed. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table 5: Results for Generated Data (Developing Countries, Three-Sector Model)

Regressor	(1) ltShareM	(2) ltShareM	(3) ltShareM	(4) ltShareM	(5) ltShareM
log(RELPR)	0.175***		0.002	0.012	0.227***
	(0.047)		(0.032)	(0.034)	(0.046)
log(CEN)		0.501***	0.194**	0.184**	0.197**
		(0.154)	(0.092)	(0.089)	(0.080)
log(AP)			0.277***	0.283***	
			(0.054)	(0.057)	
log(POP)				0.021	0.150***
				(0.019)	(0.038)
log(PCGDP)					0.478***
					(0.091)
Observations	79	79	79	79	79
R-squared	0.164	0.165	0.535	0.541	0.661

Notes: Table displays coefficients and robust standard errors for OLS estimations using generated data (see Section 5 for details). The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. CEN is a country's centrality measure (defined in Section 2). POP is a country's population size, AP its labor productivity in agriculture, and PCGDP its per-capita GDP, respectively. All regressors are in logs. Results on the included constant are suppressed. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table 6: Results for Generated Data (Robustness Checks for Two-Sector Model)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Regressor	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltshareM	ltShareM	ltShareM
log(RELPR)	0.207***		-0.023	0.231***		0.001	0.206***		-0.027
	(0.047)		(0.038)	(0.048)		(0.036)	(0.048)		(0.038)
log(CEN)		0.502***	0.170*		0.509***	0.139*		0.518***	0.188*
		(0.144)	(0.089)		(0.128)	(0.075)		(0.157)	(0.098)
log(AP)			0.392***			0.397***			0.395***
			(0.046)			(0.048)			(0.047)
Observations	79	79	79	79	79	79	79	79	79
R-squared	0.181	0.149	0.725	0.210	0.180	0.729	0.176	0.145	0.721
σ_A	2.6	2.6	2.6	2.3	2.3	2.3	2.3	2.3	2.3
σ_M	2.0	2.0	2.0	2.3	2.3	2.3	2.3	2.3	2.3
Trade Cost Matrix	Poisson	Poisson	Poisson	OLS	OLS	OLS	Poisson	Poisson	Poisson
Prices Deflators	Consumer	Consumer	Consumer	Consumer	Consumer	Consumer	Producer	Producer	Producer

Notes: Table displays coefficients and robust standard errors for OLS estimations using generated data (see Section 5 for details). The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. CEN is a country's centrality measure (defined in Section 2) and AP its labor productivity in agriculture. All regressors are in logs. Results on the included constant are suppressed. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table 7: Results for Generated Data (Baseline, Free Trade, and Autarky)

	Baseline			Autarky			Free Trade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Regressor	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM	ltShareM
log(RELPR)	0.206***		-0.027	0.232***		0.011	-1.329***		-1.332***
	(0.048)		(0.038)	(0.048)		(0.041)	(0.011)		(0.018)
log(CEN)		0.516***	0.187*		0.038	0.096		-0.023	-0.027
		(0.156)	(0.097)		(0.110)	(0.085)		(0.027)	(0.025)
log(AP)			0.395***			0.417***			0.015
			(0.047)			(0.059)			(0.020)
Observations	79	79	79	79	79	79	79	79	79
R-squared	0.176	0.145	0.721	0.727	0.727	0.727	0.997	0.997	0.997
Coeff. (SE) actual on simulated data	0.777 (0.208)***	0.777 (0.208)***	0.777 (0.208)***	0.722 (0.204)***	0.722 (0.204)***	0.722 (0.204)***	0.175 (0.109)	0.175 (0.109)	0.175 (0.109)
R^2 actual on simulated data	0.147	0.147	0.147	0.133	0.133	0.133	0.0205	0.0205	0.0205

Notes: Table displays coefficients and robust standard errors (clustered at the country-level) for OLS estimations using generated data (see Section 5 for details). The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. CEN is a country's centrality measure (defined in Section 2) and AP its labor productivity in agriculture. All regressors are in logs. Results on the included constant are suppressed. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Appendix Tables

Table A.1: Gravity Equation Estimates for 2000 (Poisson and OLS)

	Manufacturing		Agriculture	
Regressor	(1) Exports	(2) log(Exports)	(3) Exports	(4) log(Exports)
d_{int}	-1.989***	-1.518***	-3.146***	-4.056***
	(0.119)	(0.368)	(0.164)	(0.578)
log(distance)	-0.906***	-1.842***	-0.962***	-1.355***
	(0.052)	(0.029)	(0.055)	(0.067)
Observations	10,170	10,170	3,145	3,145
R-squared	--	0.836	--	0.716
Estimation method	Poisson	OLS	Poisson	OLS

Notes: Table displays coefficients and robust standard errors (clustered by exporter) for OLS and Poisson QML estimations (see Section 5 for details). The dependent variable is the value of bilateral exports. The regressors are a dummy variable (d_{int}) which takes the value one if a trade flow crosses national borders, and the log of bilateral distance. Also included are a full set of exporter and importer fixed effects. Results on the included constant are suppressed. See Table 2 for data sources. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table A.2: Results for Actual Data (All Countries)

Regressor	(1) ltShareM	(2) ltShareM	(3) ltShareM	(4) ltShareM	(5) ltShareM
log(RELPR)	0.104*		-0.00138	0.0825	0.112*
	(0.0596)		(0.0754)	(0.0700)	(0.0621)
log(CEN)		0.316***	0.227**	0.215**	0.202**
		(0.0880)	(0.0974)	(0.0903)	(0.0943)
log(AP)			0.0590	0.0662*	
			(0.0403)	(0.0392)	
log(POP)				0.158***	0.160***
				(0.0242)	(0.0245)
log(PCGDP)					0.0641*
					(0.0375)
Observations	112	112	112	112	112
R-squared	0.032	0.069	0.089	0.296	0.300

Notes: Table displays coefficients and robust standard errors for OLS estimations using generated data (see Section 5 for details). The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. CEN is a country's centrality measure (defined in Section 2). POP is a country's population size, AP its labor productivity in agriculture, and PCGDP its per-capita GDP, respectively. All regressors are in logs. Results on the included constant are suppressed. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.

Table A.3: Results for Generated Data (All Countries, Three-Sector Model)

Regressor	(1) ltShareM	(2) ltShareM	(3) ltShareM	(4) ltShareM	(5) ltShareM
log(RELPR)	0.184*** (0.046)		0.054* (0.029)	0.056* (0.032)	0.164*** (0.036)
log(CEN)		0.338*** (0.099)	0.089* (0.051)	0.089* (0.050)	0.081* (0.048)
log(AP)			0.128*** (0.025)	0.128*** (0.025)	
log(POP)				0.003 (0.015)	0.083*** (0.025)
log(PCGDP)					0.284*** (0.058)
Observations	107	107	107	107	107
R-squared	0.202	0.141	0.411	0.411	0.525

Notes: Table displays coefficients and robust standard errors for OLS estimations using generated data (see Section 5 for details). The dependent variable is the logistic transformation of a country's share of manufacturing in GDP. RELPR is the quotient of a country's agricultural labor productivity and its labor productivity in manufacturing. CEN is a country's centrality measure (defined in Section 2). POP is a country's population size, AP its labor productivity in agriculture, and PCGDP its per-capita GDP, respectively. All regressors are in logs. Results on the included constant are suppressed. *, **, and *** signify statistical significance at the 10%, 5% and 1% levels.