Economic Integration and Structural Change*

Jean Imbs
Paris School of Economics
and CEPR

Claudio Montenegro
The World Bank

Romain Wacziarg
UCLA, NBER and CEPR

May 2012

Abstract

The dynamic evolution of sectoral production - structural change - is associated with systematic changes in the geographic dispersion of activity: in developing countries, sectoral diversification is accompanied by geographic agglomeration, and regions become heterogeneous. In advanced economies, sectoral specialization is accompanied by geographic dispersion, and regions become homogeneous. We argue that developing countries diversify because their regions integrate with each other, and can specialize according to regional comparative advantage. Advanced economies specialize because they integrate internationally and their regions produce according to the global pattern of comparative advantage. We find systematic support for these claims in international data on sectoral production at the regional level, including in the US, Europe, China and India. Consistent with our theory, we find no such evidence once the samples focus on non-traded sectors or relatively closed regions. Economic zones formed by specialized, regionally homogeneous countries, such as Europe, tend to diversify and agglomerate, consistent with their constituent countries integrating with each other.

Keywords: Structural change, international integration, regional integration, sectoral allocation.

JEL Classification: F15, F43, O11, O14, O25, O40

---

*Imbs: Paris School of Economics, 106-112 Boulevard de l'Hôpital, 75013 Paris, (+33) 1 44 07 82 01, jean.imbs@parisschoolofeconomics.eu; Montenegro: The World Bank, 1818 H Street, Washington, DC 20433, cmontenegro@worldbank.org; Wacziarg: UCLA Anderson School of Management, 110 Westwood Plaza. Los Angeles CA 90095, (310) 825 4507, wacziarg@ucla.edu. We are grateful to the Fondation Banque de France for financial support. We thank Caroline Freund, Ann Harrison, Aart Kraay, Ed Leamer and Nico Voigtländer, as well as seminar participants at the University of Lille, the University of Sydney, the Paris School of Economics, ECARES, INSEAD, the Graduate Institute Geneva, the World Bank, Stanford University, UCLA and the University of Michigan for helpful comments. All errors are ours.

1 Introduction

The process of economic integration has a local dimension. Regions of a country integrate with each other progressively, and the country integrates with the rest of the world. We argue that structural change, defined as the evolution of sectoral specialization at the country level, is a reflection of the joint dynamics of local and global integration. To understand structural change at a macroeconomic level, one cannot abstract from relevant information at the sub-national level.

We show empirically that economic diversification early in the development process is accompanied by regional agglomeration, while specialization later in development is associated with dis-agglomeration at the regional level. Agglomeration at early stages of development increases the structural differences between regions: the country diversifies because its constituent regions specialize in different activities. Dis-agglomeration at later stages of development, in turn, occurs homogeneously across regions: the country overall specializes as the sectoral composition of regions becomes more similar.

What explains this empirical pattern of diversification and agglomeration? Our main hypothesis is that a gradual process of economic integration affects the propensity of countries to move through stages of structural transformation. Overcoming barriers to trade across regions allows countries to diversify overall as their regions become able to specialize in specific sectors. The implied regional agglomeration is a manifestation of intra-national trade, which also tends to foster economic convergence between regions. When the process of gradual economic integration reaches the borders of a country, the country overall can start specializing and trading with the rest of the world. This specialization occurs homogeneously across integrated regions, all tending to produce more similar goods for international markets.

We formulate this hypothesis using a simple Ricardian model of an economy composed of distinct regions that are closed to each other, due to a variety of impediments to goods or factor trade. Initially, each regional unit is specialized, for instance in goods of first necessity such as agricultural staples. Thus, a poor country overall is concentrated in a small number of sectors, since its constituent regions produce a small range of similar goods. As regions gradually integrate, due for instance to an improved transportation infrastructure, each regional unit has access to a larger market. Indivisibilities become less binding and regional agglomeration can occur as a result of standard comparative advantage forces. While its regions agglomerate, the country diversifies, as its overall sectoral composition aggregates that of regions specialized in different activities.

In this first stage of development, activity diversifies in the aggregate, and agglomerates locally. Integrated regions become increasingly dissimilar in terms of their production structure. They also become increasingly similar in terms of their income and productivity levels, thanks to the efficiency-enhancing effects of trade, factor movements or the accelerated diffusion of technologies across regions. As regional convergence occurs, it is international comparative advantage rather than regional comparative advantage that ultimately becomes essential.

In the second stage of development, a similar reasoning applies. Integration becomes international rather than intranational.[1] The country as a whole specializes according to its global comparative advantage. Cross-regional differences in factor endowment or productivity have faded away with intranational integration, and the regional location of production has become less relevant. All the regions constituting a country produce a smaller range of sectors, which is determined by international comparative advantage. Sectors dis-agglomerate, as they are produced in an extended range of regions, and regions become increasingly similar in terms of their production structures.[2]

The bulk of this paper is devoted to documenting these patterns empirically using sectoral information on the regions or states that form a country. The evidence builds on three measures. First, we compute a measure of *sectoral* diversification akin to those used in Imbs and Wacziarg (2003), capturing the overall allocation of resources across sectors. Second, we characterize the regional agglomeration of activity, a measure of the *geographic* allocation of production. This requires regional data on sectoral production. Third, we compute a dissimilarity index, akin to the one in Krugman (1991), capturing differences in the patterns of production across regions. Poor, diversifying countries should see integrating economic activities agglomerate geographically and integrated regions become increasingly different. Rich, specializing countries, in contrast, should be domestically integrated and undergo both geographic dis-agglomeration and structural convergence across regions. Of course, only traded goods and open regions are expected to display these patterns.

[1] Obviously, the gradual processes of regional and international integration partially overlap. Our story relies on dometic integration occurring faster than international integration in early stages, with the reverse occurring in later stages.

[2] Dis-agglomeration does not necessarily imply similar regional patterns of production. A sector could dis-agglomerate, but production could still be localized only in a subset of regions. Regions become similar in terms of what sectors they produce precisely because of the preponderance of international comparative advantage. Of course, the regions concerned must be open to trade.

Our claims are established in three steps. First, we focus on three individual countries where structural change is prominent. Two are developing economies, China and India, where the geography and the specialization of economic activity have both recently undergone unprecedented changes. The third is an advanced industrialized country, the US, where the influence of economic integration on the domestic patterns of production is of acute political relevance.

We find that India as a whole diversifies: the allocation of factors becomes homogeneous across sectors as per capita GDP rises. Indian States agglomerate: the geographic allocation of factors becomes increasingly concentrated. Indian States also produce increasingly different goods. The same pattern holds in China. To establish that these patterns result from the process of gradual economic integration described above, we split sectors into traded and non-traded activities, and regions according to the tradeability of the goods produced there.[3] Remarkably, sectoral diversification cannot be detected amongst non-traded sectors. Moreover, geographic agglomeration does not occur in relatively closed regions, which on the contrary tend to become increasingly homogeneous. A similar pattern is observed in China.

In the United States, specialization - defined as rising sectoral concentration - increases with per capita GDP. Simultaneously, the geographic allocation of activity becomes increasingly uniform, and US States produce an increasingly similar set of goods. Geographic agglomeration progressively declines in the data, as all US States increasingly resemble the aggregate production pattern. This mirrors the completion of economic integration between US States. Because of this very integration, little is available in the US by way of sectors or regions that remain unexposed to (intra- or inter-national) trade. The patterns of specialization and geographic homogeneization are therefore predicted to be more widespread there. We do find however that US States where traded production represents a small fraction of State employment are in fact still agglomerating.[4]

The second step is to establish these claims more generally in the laboratory of European integration. European enlargement has created an expanding area of free trade between countries, but each member country has also undergone domestic integration, before or during accession. Regional data on sectoral employment from Eurostat, the European statistical agency, suggest

[3]For instance, an Indian State is classified as relatively closed if Mining, Manufacturing, Transport, Storage, Communication, Banking and Insurance - all sectors commonly considered to be traded - represent less than 15% of regional employment, the median across Indian States.

[4]We identify such States as ones where employment in Agriculture, Mining, Manufacturing, Finance, Insurance and Real Estate represents less than 20% of the State aggregate.

European countries each tend to specialize, as their constituent regions produce a converging set of goods. These are relatively developed and domestically integrated countries. They go through structural dynamics akin to what is observed in the US. Our conjecture is this is a manifestation of European integration. As they enter the European Union, countries specialize according to their comparative advantage within the Union. Since they are domestically integrated, their constituent regions specialize thus, and become structurally more similar.

If our hypothesis is correct, as long as integration among European Union (EU) countries is a stronger force than integration of the EU with the rest of the Worlds, the EU as a whole should diversify, as activity agglomerates at country-level, and EU countries specialize in different goods. European integration is happening between its constituent countries, whereas the degree of openness of the whole zone to world trade remains relatively unchanged. These countries are increasingly trading *with each other*.

We construct a synthetic aggregate of EU-wide employment. Unlike the previous exercises, this can be done with sectoral data at the country level, which we obtain from the International Labor Office (ILO). The data show that the EU diversifies, activity agglomerates at the country level, and countries grow dissimilar. This is particularly true of traded sectors and the countries that produce them, exactly as predicted by our story.[5]

The third step is to consider a sample of 28 countries for which regional data at sectoral level are available. The data include countries at various stages of development, for instance Vietnam, China, Bolivia, Switzerland or the US - but with shorter time coverage, and with a focus on developing economies. They are obtained from the Integrated Public Use Microdata Series (IPUMS) project, which harmonizes census data internationally. The data form a panel where it is possible to focus on the within-country dynamics of specialization, agglomeration and dissimilarity, thanks to country-specific intercepts. The sample reproduces a U-shaped relation between sectoral specialization and per capita GDP, as documented in Imbs and Wacziarg (2003). It is especially prevalent in traded sectors. Geographic agglomeration displays precisely the opposite pattern: it first increases with per capita GDP (as countries diversify because regions specialize), and then reverts to a negative

[5] Since sectoral (but not regional) data are needed for this exercise, information is available for European countries prior to their entry into the EU - i.e. when some non-tradeability still exists in sample. The same is not true of regional data on sectoral activity, whose collection typically starts around the accession date, i.e. when integration is already advanced.

slope (as countries specialize because regions grow similar). The same holds true for an index of regional dissimilarity. These dynamics do not hold when limiting the analysis to to non-traded sectors or to the (closed) regions that produce them.

Our paper is primarily related to the vast literature on structural transformation (Kuznets, 1966, Chenery, Robinson and Syrquin, 1986, Caselli and Coleman, 2001). This literature emphasized a multiplicity of forces underlying structural change in growing economies. Among the most important were: changes in demand patterns due to non-homothetic preferences; sectorally differentiated growth in labor productivity; changing patterns of comparative advantage brought forth by factor accumulation; and economic integration itself. In this paper we focus on the force of economic integration - both at local and global levels.

Our research is also related to the large body of work on economic geography and the location of production (Krugman, 1991a, Krugman, 1991b, Krugman and Venables, 1995, Desmet and Rossi-Hansberg, 2011).[6] In contrast to these contributions, our hypothesized mechanism does not need to rely on demand side externalities, congestion costs or increasing returns to explain the location of production and regional agglomeration patterns. Instead, classical comparative advantage arguments are sufficient to make the point that integrating regions will tend to specialize. This literature is also largely silent about structural change.

Finally, our paper is related to ongoing research on specialization dynamics, both theoretical and empirical. Findlay (1970) examines a model where capital is accumulated and generates predictions about the evolution of comparative advantage from a neoclassical trade model. Ventura (1997) builds on this tradition. Redding (2002) generates country level measures of specialization from a microfounded neoclassical trade model, linking specialization to changes in the abundance of factor endowments. In these contribution, no variation exists in the extent of economic integration across countries, i.e. the interaction between trade and development is analyzed unde rand open trade assumption. Finally, Imbs and Wacziarg (2003) describe two stages of sectoral diversification - at an early stage, countries diversify, while at a later stage of development they specialize. This paper can be understood as providing an explanation for this stylized fact based on the process of gradual regional and international integration.

[6] See Redding (2009) for a survey of the empirical outgrowth of this literature on economic geography. See also Deichmann, Lall, Redding and Venables (2008) on the determinants of industrial location with a specific focus on developing countries.

The paper is organized as follows: Section 2 introduces a simple Ricardian model featuring integration at both the local and international levels. Extensions to nontradable goods and closed regions provide simple tests of our specific mechanism of gradual local and global integration to explain the stages of structural change. Section 3 discusses the measures of sectoral specialization, geographic agglomeration and regional dissimilarity, as well as the data used to compute these measures. Section 4 presents results for India, China and the US. Section 5 focuses on European evidence. Section 6 expands to international data. Section 7 concludes.

2 A Model of Local and International Integration

In this section, we introduce a simple Ricardian model of interregional and international trade. The model has two main goals. The first goal is to identify sufficient conditions under which a process of gradual integration can affect structural change. The model generates predictions on the pattern of diversification, agglomeration and inter-regional sectoral similarity, that are later compared to the patterns found in the data. The second goal is to generate auxiliary predictions from the model that can be evaluated empirically as tests of the specific mechanism emphasized throughout this paper: namely that a process of gradual local and then global integration accounts for the pattern of specialization and agglomeration we uncover empirically. In particular, we examine what happens, in our model, when we introduce nontradable goods and regions that fail to integrate, and then evaluate empirically the predictions that result from these extensions to our basic model, exploiting variation in the degrees of tradability and regional openness.

2.1 Basic Structure

The model features three countries, A, B and C. Each country is composed of 3 regions indexed by $j = 1, 2, 3$, each endowed with L_j units of labor, the sole factor of production. There are three sectors, indexed by $s = 1, 2, 3$ (our analysis focuses on country A). We study a 3 sector, 3 region, 3 country model in order to allow, in extensions to the basic theory, for the possibility that one good is non tradable, that one region remains closed to trade, or that one region integrates globally before integrating with the regions of its own country.

We study the model at three points in time. At time 0, representing countries at early stages of development, every region of every country lives in full autarky. At time 1, the intermediate stage, there is domestic integration for trade in goods. At time 2, meant to capture advanced economies,

there is domestic integration for trade in goods and labor, and international integration for trade in goods. That is, countries are autarkic until time 2. Since countries are initially autarkic, the analysis for time 0 and 1 apply to countries A, B and C identically, while at time 2 we solve for the three-country general equilibrium.

The demand side is as follows: Each region has one identical representative consumer with preferences over all three goods: $U_j(C_{j1}, C_{j2}, C_{j2}) = C_{j1}^{1/3} C_{j2}^{1/3} C_{j3}^{1/3}$, where C_{js} is consumption of good s in region j. Goods 2 and 3, interpreted as manufactured goods (while good 1 can be interpreted as an agricultural staple) can only be produced once a certain threshold level of output can be met. If this indivisibility is binding at time 0, goods 2 and 3 are not produced, and consumer utility is zero in all regions in all countries, a normalization. In other words, indivisibilities initially prevent autarkic regions from moving out of agriculture. Once the size of the market becomes sufficient through local integration, the production of goods 2 and 3 become possible.

The production side is as follows: Exogenous labor productivities are labeled a_{js} where j refers to the region and s refers to the sector. Technology is $Y_{js} = a_{js} L_{js}$ where L_{js} is the amount of labor employed in sector s. So for all j, $L_j = L_{j1} + L_{j2} + L_{j3}$.

2.2 Analysis at Time 0

The analysis if particularly simple at time 0. If there were no indivisibilities, each region would be diversified, and solving for each region's autarky general equilibrium in this case is particularly simple. We can introduce a system of prices p_{js} (price of good s in region j), and wages w_{js}.

The representative consumer in each of regions $j = 1, 2, 3$ solves:

$$\begin{aligned} \underset{C_{j1}, C_{j2}}{Max}\ U_j(C_{j1}, C_{j2}, C_{j3}) &= C_{j1}^{1/3} C_{j2}^{1/3} C_{j3}^{1/3} \\ \text{subject to } p_{j1}C_{j1} + p_{j2}C_{j2} + p_{j3}C_{j3} &= w_{j1}L_{j1} + w_{j2}L_{j2} + w_{j3}L_{j3} \end{aligned}$$

The first order conditions lead to:

$$\frac{C_{js}}{C_{jr}} = \frac{p_{jr}}{p_{js}} \text{ for any sector } s,\ r.$$

On the producer side, the producer of good s in region j, operating under perfect competition (taking p_{js} as given), solves:

$$\underset{L_{js}}{Max} \Pi_{js} = p_{js} Y_{js} - w_{js} L_{js} = p_{js} a_{js} L_{js} - w_{js} L_{js}$$

The familiar first order conditions, for $j = 1, 2, 3$ and $s = 1, 2, 3$, are:

$$p_{js} a_{js} = w_{js}$$

In addition to these conditions, there are resource constraints, namely $L_j = L_{j1} + L_{j2} + L_{j3}$ for $j = 1, 2, 3$. The arbitrage conditions for wages within each region are $w_{j1} = w_{j2} = w_{j3} \equiv w_j$ for $j = 1, 2, 3$ (wages can be normalized to 1 in each region). Then:

$$\begin{aligned} C_{js} &= Y_{js} = \frac{a_{js}}{3} L_j \\ \frac{p_{js}}{p_{jr}} &= \frac{a_{jr}}{a_{js}} \end{aligned}$$

for any sector s, r.

We now introduce indivisibilities. We assume that sector 1 (agriculture) can be opened without cost, while for sectors 2 and 3 (manufacturing), a one time threshold level of production F_{js} (for $j = 1, 2, 3$ and $s = 2, 3$) expressed in terms of sector j's output, must be reached at time 0 in order for the sector to operate (F is a fixed cost). We assume that $F_{js} > a_{js} L_j / 3$, so each region remains fully specialized in good 1 at time 0. This captures the idea that regions in autarky have markets that are too small to support the opening of manufactured goods sectors, consistent with Adam Smith's idea that the size of the market limits the extent of specialization. Then, region 1 produces $Y_{11} = a_{11} L_1$, region 2 produces $Y_{21} = a_{21} L_2$ and region 3 produces $Y_{31} = a_{31} L_3$ in the regional autarky equilibrium. Since goods 2 and 3 are not produced, $Y_{12} = Y_{22} = Y_{32} = Y_{13} = Y_{23} = Y_{33} = 0$. Moreover, since there is no interregional trade, $C_{js} = Y_{js}$ for all j, s.

From these equilibrium values of production we can compute indices of specialization, agglomeration and interregional sectoral similarity that were discussed in the introduction. We will say more about these indices in Section 3, where we define them in the context of the empirical analysis. For now it suffices to mention that in what follows we employ the exact theoretical counterparts to the indices used in the empirical section.

We can calculate the countrywide Herfindahl index of diversification and find:

$$S^H = \sum_s \left(\frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} \right)^2 = 1$$

This is obvious since only one sector is being produced: the country is fully concentrated in the production of that sector.

Geographic agglomeration, a Herfindahl index of the regional concentration of sector 1, is:

$$A_1^H = \sum_j \left(\frac{Y_{j1}}{\sum_j Y_{j1}} \right)^2 = 1 - 2 \left(\frac{a_{11}L_1 a_{21}L_2 + a_{11}L_1 a_{31}L_3 + a_{21}L_2 a_{31}L_3}{(a_{11}L_1 + a_{21}L_2 + a_{31}L_3)^2} \right) < 1$$

The degree of dissimilarity between regions is:

$$D_1 = \frac{1}{3} \sum_{j<k} \left| \frac{Y_{js}}{\sum_s Y_{js}} - \frac{Y_{ks}}{\sum_s Y_{ks}} \right| = 0$$

This is also rather obvious since both regions have identical shares of the production of good 1 (shares equal to 1).

2.3 Analysis at Time 1

At time 1, the country is still autarkic but the regions can now trade. Labor remains immobile across regions. Thus, there is goods price arbitrage across regions (by free trade) and wage arbitrage across sectors (but not across regions).

We assume that in all countries, region 1 will produce and export good 1, region 2 will produce and export good 2, and region 3 will produce and export good 3. For tractability we consider only assigments of a single sector to a single region. It is well-known since Jones (1961) that in the n-sector, n-region Ricardian model, focusing on the class of assignments of a single sector to be produced in a single region, an assignment will be the efficient assignment and hence the Ricardian equilibrium if and only if it maximizes the product of sectoral productivities across all other possible assignments of that class. Formally:

$$\prod_i a_{ii} > \prod_i a_{ij(i)} \text{ for all } j(i)$$

where $j(i)$ is any *other* assignment of sector j to region i. Without loss of generality, we assume that this condition holds, so that the pattern of specialization described above results.

Under free regional trade region 1 will produce only good 1, region 2 will produce only good 2 and region 3 will produce only good 3. Then output in each region is:

$$\begin{aligned} Y_{11} &= a_{11}L_1; Y_{12} = 0; Y_{13} = 0 \\ Y_{21} &= 0; Y_{22} = a_{22}L_2; Y_{23} = 0 \\ Y_{31} &= 0; Y_{32} = 0; Y_{33} = a_{33}L_3 \end{aligned}$$

This is an equilibrium if the threshold level of production for goods 2 and 3 in regions 2 and 3 is now met, i.e. if $F_{ii} < a_{ii}L_i$ (for $i = 2, 3$). We assume this is the case, so that the size of the (countrywide) domestic market is sufficient to support the operation of the manufacturing sectors.

At time 1, the country overall is diversified (each region produces each of the three goods), production is geographically agglomerated, and regions are sectorally dissimilar. In principle this is enough to calculate indices of diversification, agglomeration and similarity either for labor per sector or output per sector.[7] The countrywide Herfindahl index of diversification is:

$$\begin{aligned} S^H &= \sum_s \left(\frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} \right)^2 \\ &= 1 - \frac{2(a_{11}L_1 a_{22}L_2 + a_{11}L_1 a_{33}L_3 + a_{22}L_2 a_{33}L_3)}{(a_{11}L_1 + a_{22}L_2 + a_{33}L_3)^2} < 1 \end{aligned}$$

The geographic agglomeration Herfindahl index of sector 1 is:

$$A_1^H = \sum_j \left(\frac{Y_{j1}}{\sum_j Y_{j1}} \right)^2 = 1$$

Similarly, $A_2^H = 1$ and $A_3^H = 1$. Since production of each sector occurs in a different region, the country is fully agglomerated.

The average degree of sectoral dissimilarity between regions is:

$$D_1 = \frac{2}{J(J-1)} \sum_{j<k} \left| \frac{Y_{js}}{\sum_s Y_{js}} - \frac{Y_{ks}}{\sum_s Y_{ks}} \right| = \frac{2}{3}$$

This is the maximal sectoral dissimilarity across regions for a given sector when $J = S = 3$. Similarly, $D_2 = 2/3$ and $D_3 = 2/3$, so:

$$D = \frac{1}{S} \sum_s D_s = \frac{2}{3}$$

[7] For completeness we can use the consumer's and producers' first order conditions, the consumer's budget constraint, the resource constraints, the market clearing conditions and the arbitrage conditions on wages and prices to fully characterize the Ricardian equibrium of this simple system:

$$\begin{aligned} C_{11} &= C_{21} = C_{31} = \frac{a_{11}L_1}{3} \\ C_{22} &= C_{12} = C_{32} = \frac{a_{22}L_2}{3} \\ C_{33} &= C_{13} = C_{23} = \frac{a_{33}L_3}{3} \\ \frac{p_1}{p_2} &= \frac{a_{22}L_2}{a_{11}L_1}; \frac{p_2}{p_3} = \frac{a_{33}L_3}{a_{22}L_2} \\ \frac{w_1}{w_2} &= \frac{L_2}{L_1}; \frac{w_2}{w_3} = \frac{L_3}{L_2} \end{aligned}$$

This too is intuitive: since each region is fully specialized, each is maximally dissimilar from the other.

2.4 Analysis at Time 2

We focus the analysis on country A, but the results are symmetric for countries B and C. Time 2 is meant to capture advanced economies, and accordingly we assume that three things have changed. To generate clear predictions, we purposefully make these assumptions stark. Firstly, each country's trade with the other has become fully open. Secondly, labor can move freely across regions within a country (but not across countries). The wage arbitrage condition then implies that $L_1^i = L_2^i = L_3^i$ $\forall i \in \{A, B, C\}$. From now on, we denote by $L^i = L_1^i + L_2^i + L_3^i$ total labor in country i. Thirdly, regional productivities have converged and are now equalized: $a_{1s}^i \to a_{2s}^i \to a_{3s}^i \equiv a_s^i$ for $s = 1, 2, 3$ and $i \in \{A, B, C\}$. In other words, as regions integrate, regional comparative advantage disappears - there is convergence in labor productivities. A possible interpretation of this assumption is that with open interregional trade, technologies for the production of different goods diffuse across regions, resulting in the same sectoral productivities across regions of a given country.[8] Thus, each country is now composed of three identical regions.

Assume without loss of generality that country A will produce sector 1, B produces sector 2 and C produces sector 3 (this requires an assumption on the product of a_1^i, a_2^i and a_3^i, for $i \in \{A, B, C\}$, that is analogous to the one made at time 1 with respect to regional comparative advantage). Then by a reasoning exactly identical to the one for time 1 (but applied to countries rather than regions), we have (for country A):

$$\begin{aligned}
Y_{11}^A &= a_1^A L_1^A = \frac{1}{3} a_1^A L^A \\
Y_{21}^A &= a_1^A L_2^A = \frac{1}{3} a_1^A L^A \\
Y_{31}^A &= a_1^A L_3^A = \frac{1}{3} a_1^A L^A \\
Y_{12}^A &= 0; Y_{22}^A = 0; Y_{32}^A = 0 \\
Y_{13}^A &= 0; Y_{23}^A = 0; Y_{33}^A = 0
\end{aligned}$$

[8] The regional convergence of aggregate productivity across regions for rich countries is a well-documented fact in the growth literature. See for instance Barro and Sala-i-Martin (chapter 11) for an empirical investigation across US states, Japanese prefectures and European regions. Our assumption is that such convergence carries over at the sectoral level as well.

(similarly in countries B and C).

The rest of the equilibrium (relative prices, consumption levels, relative wages) can be derived in a way analogous to the equilibrium at time 1. Countries are perfectly specialized (A in good 1, B in good 2 and C in good 3), and there is complete regional disagglomeration within each country, which each country's respective sector of production located uniformly across regions. Regions within each country are identical. The various indices we are interested in take on the following values:

The country-wide Herfindahl index of diversification is:

$$S^H = \sum_s \left(\frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} \right)^2 = 1$$

This follows from each country now producing in only one sector: each country is now fully specialized.

The geographic agglomeration of sector 1 is:

$$A_1^H = \sum_j \left(\frac{Y_{j1}}{\sum_j Y_{j1}} \right)^2 = \frac{1}{3}$$

This follows from sector 1 being produced in all three identical regions of country A.

The degree of dissimilarity between regions is:

$$\begin{aligned} D_1 &= \frac{2}{J(J-1)} \sum_{j<k} \left| \frac{Y_{js}}{\sum_s Y_{js}} - \frac{Y_{ks}}{\sum_s Y_{ks}} \right| = 0 \\ D_2 &= 0 \text{ (sector 2 not produced anywhere)} \\ D_3 &= 0 \text{ (sector 3 not produced anywhere)} \\ D &= \frac{1}{S} \sum_s D_s = \frac{1}{3}(D_1 + D_2 + D_3) = 0 \end{aligned}$$

2.5 Summary: The Evolution of Specialization, Agglomeration and Similarity

In the 3 goods, 3 regions, 3 countries model, the evolution of sectoral diversification, regional agglomeration and interregional sectoral similarity is summarized in the following table:

	Country-level Diversification (S^H)
Time 0 (initial)	1
Time 1 (intermediate)	$1 - \frac{2(a_{11}L_1a_{22}L_2+a_{11}L_1a_{33}L_3+a_{22}L_2a_{33}L_3)}{(a_{11}L_1+a_{22}L_2+a_{33}L_3)^2}$
Time 2 (developed)	1
	Regional Agglomeration (A^H)
Time 0 (initial)	$1 - 2\left(\frac{a_{11}L_1a_{21}L_2+a_{11}L_1a_{31}L_3+a_{21}L_2a_{31}L_3}{(a_{11}L_1+a_{21}L_2+a_{31}L_3)^2}\right)$
Time 1 (intermediate)	1
Time 2 (developed)	$\frac{1}{3}$
	Regional Dissimilarity (D)
Time 0 (initial)	0
Time 1 (intermediate)	$\frac{2}{3}$
Time 2 (developed)	0

Our simple Ricardian model points to a pattern of diversification going hand in hand with agglomeration and increasing regional differences between time 0 and time 1, and a pattern of specialization, dis-agglomeration and convergence in sectoral structure between time 1 and time 2. These predicted patterns are confronted to the data in sections 4, 5 and 6.

2.6 Extension 1: A nontradable good

We first assume that good 3 is nontradable in all periods (sector 3 can be reinterpreted as the service sector). The goal of this extension is to examine what happens to the dynamics of structural change when focusing on samples that exclude nontradable goods. We now have 2 tradable goods, 3 regions. and 3 countries. With a closed sector, the analysis changes in interesting ways, but does not differ much analytically from the $3 \times 3 \times 3$ case. The analysis for time 0 does not change at all, as goods 2 and 3 are not produced.

At time 1, all regions produce good 3, the nontradable good. Goods price arbitrage across regions holds for all but good 3. We continue to assume that region 1 has a comparative advantage in good 1 and region 2 in good 2 ($a_{11}/a_{12} > a_{31}/a_{32} > a_{21}/a_{22}$). The main issue is whether region 3 produces good 1, good 2 or both. This depends on demand, relative productivities and relative sizes. In either case, our results concerning patterns of specialization, agglomeration and structural similarity do not change qualitatively. Assume that parameters are such that region 3 produces both goods 1 and 2. Appendix 1 solves for the general equilibrium in this case. Each region j

will produce good 3 in quantity $Y_{j3} = a_{j3}L_j/2$. Regions 1 and 2 will, in addition, produce only goods 1 and 2, respectively, while region 3 will produce both. Solving for output quantities, we can compute indices of specialization, agglomeration and similarity.

The difference between countrywide specialization in this case and in the generic case with 3 tradable goods cannot be signed in general. The country overall is still diversified in the sense of producing all three goods, but it does so in different locations compared to the full-tradability case. What we do know for sure is that a time 1, the country is less agglomerated than in the generic case: A^H is strictly smaller than 1. The main reason is that good 3 is produced in all three regions, since it is nontradable. Another reason is that region 3 might produce both goods 1 and 2.[9] For the same reasons, interregional sectoral dissimilarity tends to be lower in period 2 when there are nontradables.

Finally, at time 2, we continue to assume that country A has a comparative advantage in sector 1 while country B has a comparative advantage in sector 2: $a_1^A/a_2^A > a_1^C/a_2^C > a_1^B/a_2^B$. Countries are now imperfectly specialized, because of the presence of good 3 (the nontradable good): $S^H < 1$. However, there is again complete regional disagglomeration ($A^H = 1/3$), and all regions are sectorally similar ($D = 0$).

The bottom line prediction of this extension is that, when including non-tradables goods in the sample: 1) the pattern of agglomeration at time 1 should be less pronounced (flatter) than when excluding nontradable goods and 2) structural change toward more countrywide specialization should be less pronounced (flatter) a time 2 than when excluding nontradables. These predictions will be tested in the empirical sections.

2.7 Extension 2: A closed region

We now assume that a region, region 3, remains closed at time 1, instead of integrating with the rest of country A.[10] The analysis for time 0 is again unchanged. At time 1, region 3 continues

[9] In the case where region 3 only produces one of goods 1 and 2, reduced agglomeration compared to the generic case still obtains due to the nontradable good.

[10] We could also assume that region 3 remains closed at time 2, although advanced economies rarely display regions that remain closed domestically. Predictions on the dynamics of structural change in this case depend on whether the closed region specializes in the sector for which the country overall has an international comparative advantage. If this is the case, nothing clanges. If not, the country is predicted to remain more diversified than in the generic case, and more agglomerated. In other words, allowing region 3 to remain closed at time 2 again weakens the dynamics of

to look as it did at time 0, i.e. $Y_{31} = a_{31}L_3$; $Y_{32} = Y_{33} = 0$. For the other two regions, we have to consider a Ricardian model with 3 sectors and 2 regions. We can write a chain of comparative advantage, assuming the following:

$$\frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

With this assumption, 5 distinct cases arise:

1) Region 1 produces good 1, region 2 produces goods 1, 2, 3:

$$\frac{w_1}{w_2} = \frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

2) Region 1 produces good 1, region 2 produces goods 2 and 3:

$$\frac{a_{11}}{a_{21}} > \frac{w_1}{w_2} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

3) Region 1 produces goods 1 and 3, region 2 produces goods 2 and 3 (perhaps the focal case):

$$\frac{a_{11}}{a_{21}} > \frac{w_1}{w_2} = \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

4) Region 1 produces goods 1 and 3, region 2 produces good 2:

$$\frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{w_1}{w_2} > \frac{a_{12}}{a_{22}}$$

5) Region 1 produces goods 1, 2 and 3, region 2 produces good 2:

$$\frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}} = \frac{w_1}{w_2}$$

Cases 4 and 5 are essentially the same as cases 1 and 2, respectively, as far as predicted indices of specialization, agglomeration and dissimilarity are concerned, so we derived these indices in cases 1-3. Appendix 1 contains the details, but the bottom line is that, whatever the case under consideration, at least one region produces more than one good. Thus, at time 1 predicted indices of agglomeration and dissimilarity are unambiguously lower than in the generic case. On the other hand, the predicted index of countrywide specialization bears an ambiguous relation with its counterpart in the generic case, depending on the relative sizes of the regions.

The bottom line is that, compared to the full specialization equilibrium of the baseline model, where we had perfect agglomeration, maximal interregional dissimilarity and country-wide diversification, we still have countrywide diversification (the change in the extent of diversification is

change in specialization, agglomeration and sectoral dissimilarity, changing little to the overall conclusions from the analysis of this extension.

ambiguous), but less agglomeration and less interregional dissimilarity across all 5 cases. Thus, introducing closed regions weakens the dynamics of regional agglomeration at intermediate stages of development.

3 Measurement and Data

We first introduce the three indices central to this paper. We then turn to describing our data sources.

3.1 Indices

Sectoral specialization and geographic agglomeration are captured using conventional concentration indices. Let Y_{ijst} denote a measure of economic activity in sector s of region j in country i at time t. A simple Herfindahl index of sectoral specialization is first computed, as:

$$S_{it}^{H} = \sum_{s} \left(\frac{\sum_j Y_{ijst}}{\sum_s \sum_j Y_{ijst}} \right)^2$$

The index S_{it}^{H} reflects the time pattern of sectoral specialization for country i. The numerator sums sectoral activity across all regions; the denominator represents aggregate country-level economic activity in year t. Alternatively, a Gini index of sectoral specialization is computed according to:

$$S_{it}^{G} = \frac{1}{2} - \frac{1}{S} \left(CSS - \frac{1}{2} \right)$$

where CSS denotes cumulated sectoral shares, $\left(\sum_j Y_{ijst} \right) / \left(\sum_s \sum_j Y_{ijst} \right)$, and S is the total number of sectors. Both S_{it}^{H} and S_{it}^{G} can be readily obtained from sectoral data, i.e. no regional decomposition is necessary. These are the measures used in Imbs and Wacziarg (2003) to describe the dynamics of structural change.

Geographic agglomeration is captured similarly. Define:

$$A_{ist}^{H} = \sum_{j} \left(\frac{Y_{ijst}}{\sum_j Y_{ijst}} \right)^2$$

A_{ist}^{H} is an index of the regional agglomeration of sector s in country i at time t. The denominator represents activity in sector s across all the regions of country i. The summation is performed on the (squared) shares of each region j in overall sectoral activitiy. The Herfindahl index captures the

regional lumpiness of activity in sector s across the regions j that constitute country i. Analogously, define the Gini measure of regional agglomeration:

$$A_{ist}^{G} = \frac{1}{2} - \frac{1}{S}\left(CRA - \frac{1}{2}\right)$$

where CRA denotes cumulated regional shares $Y_{ijst} / \left(\sum_j Y_{ijst}\right)$.

Both of the latter two indices require a sectoral breakdown of economic activity at the sub-national level. Both capture the sector-specific geographic agglomeration of activity. They must therefore be aggregated up to the country level. We do so using the share of each sector in the overall economy, $\left(\sum_j Y_{ijst}\right) / \left(\sum_s \sum_j Y_{ijst}\right)$ as weights. That is, we compute:

$$A_{it}^{H} = \sum_s \frac{\sum_j Y_{ijst}}{\sum_s \sum_j Y_{ijst}} A_{ist}^{H}$$

and:

$$A_{it}^{G} = \sum_s \frac{\sum_j Y_{ijst}}{\sum_s \sum_j Y_{ijst}} A_{ist}^{G}$$

Given the well-known cardinal properties of Herfindahl and Gini indices, high values of these indicators indicate a high degree of sectoral agglomeration across regions.

The degree of dissimilarity between regions is captured by an average of bilateral differences in sectoral shares. For all pairs of regions j and k in country i, we compute:

$$D_{ist} = \frac{2}{J(J-1)} \sum_{j<k} \left| \frac{Y_{ijst}}{\sum_s Y_{ijst}} - \frac{Y_{ikst}}{\sum_s Y_{ikst}} \right|$$

where J is the total number of regions in country i. The index D_{ist} captures the average dissimilarity in sectoral allocation between any two regions that constitute country i. Sectoral information at regional level is of course of the essence to obtain D_{ist}.

This measure is obtained sector by sector, so it again must be aggregated up to the country level. That can be done either arithmetically, or using sector weights, by defining, respectively:

$$D_{it} = \frac{1}{S} \sum_s D_{ist}$$

and:

$$D_{it}^{W} = \sum_s \frac{\sum_j Y_{ijst}}{\sum_s \sum_j Y_{ijst}} D_{ist}$$

Both D_{it} and D_{it}^{W} take high values when a country is constituted of regions with heterogeneous sectoral activities. The latter gives high weight to sectors that are economically important *in the*

aggregate. But a measure of heterogeneity at regional level should not necessarily reflect country-wide production patterns. Two regions can be dissimilar even if they both specialize in (different) sectors that carry little weight in the country as a whole, and an index of dissimilarity should capture such heterogeneity. For this reason D_{it} is the preferred measure since this index makes use of no country-wide information when computing regional dissimilarity.[11]

All of these indices can be computed on the universe of available sectors and regions. To test our conjecture, however, it is important to obtain counter-factual properties, arising from sectors and regions that are sheltered from economic integration. This raises empirical as well as computational issues. Given a set of non-traded sectors, it is possible to identify a corresponding set of relatively closed regions, defined as those where non-traded sectors represent a large share of activity. At the one-digit sector classification level, it is for instance customary to consider the production of energy, construction, retail trade, restaurants, hotels or community and government services as non-traded goods. The classification is undoubtedly coarse, but the inclusion of traded goods in non-traded categories creates if anything a bias against finding any differences across the two sub-samples.

Given such sectoral and geographic splits, the paper computes versions of all five indices sub-samples that reflect varying degrees of tradability or openness. For the specialization indices S_{it}^H and S_{it}^G, focused on the sectoral dimension of the data, we report values that correspond to traded or non-traded industries. For the agglomeration indices A_{it}^H and A_{it}^G, and dissimilarity index D_{it}, which are all computed from the regional dimension of the data, we report values arising from closed or open regions separately.

3.2 Data

Given the sectoral and regional indices computed in this paper, it is important that the data be balanced across all four dimensions, within each country considered. Changes in the regional or sectoral coverage translate in unstable estimates of indices of agglomeration or specialization, that can reflect mere changes in data cpverage. Sectors or regions were excluded to ensure that the number of regions and sectors remains constant over time within each country. This required taking a stance in cases a given region was missing some sectoral information, in that a choice had to be made in terms of eliminating the sector or the region altogether. Preservation of data coverage guided these decisions, summarized in the data appendix.

[11] The index D_{ist} closely resembles the measure proposed by Krugman (1991), chapter 3.

The same treatment has to be applied to artificial multi-national economic areas, constructed from country data, such as the European data. There, the number of sectors must be homogeneous across the countries constituting the area. Regional data are not necessary: country-level, sectoral information is sufficient, such as thatcirculated by the International Labor Office (ILO). International data on sectoral activity are more naturally balanced, especially when focused on European economies. Still, a few sectors or countries were omitted, as listed in the data appendix.

The five indices computed in this paper require some aggregation across sectors, across regions, and across countries. Given the homogeneity and comparability demands this imposes on the data, our preference went toward using employment as a measure of economic activity. Employment is measured in universal units, which alleviates issues of exchange rates and the measurement of sectoral price indices. In addition, international information on sectoral activity at regional level almost invariably focuses on employment rather than gross output or value added. However, whenever possible and for individual countries in particular, some production data were also used.

US data were obtained from the Bureau of Economic Analysis (GSP). We have information on State-level employment and output in 78 sectors, at the 4-digit SIC level, covering all sectors in each State's Gross State Product (GSP). For India, only output data are available, for the 28 States and Union Territories, observed for 13 sectors. Indian employment data are not available at this level of disaggregation.[12]

4 Case Studies

This section presents results for India, China and the US. Sectoral data on Indian real value added are available at the one-digit aggregation level, between 1980 and 1995. Figure 1A reports the dynamics of S_t^G, A_t^G and D_t for India, as against PPP-adjusted real per capita GDP in the country as a whole. The first row reports indices computed on the universe of available regions and sectors. India overall is unambiguously diversifying, and this is accompanied with economic activity agglomerating across its states. States also become increasingly dissimilar in terms of their production patterns.

The lower row in Figure 1A reports the same indices, now computed on sub-samples that are relatively sheltered from economic integration. Sectoral specialization S_t^G is computed on

[12] The Indian data is available at http://mospi.nic.in/dwh/index.htm

non-traded sectors only, defined as Commerce, Construction, Public Administration, Real Estate, Trade, Hotels, Restaurants and Other Services. As the first panel confirms, these activities do not diversify in India between 1980 and 1995. Output in non-traded sectors specializes weakly, although the positive slope is close to zero.

The second graph reports the dynamics of A_t^G, computed over closed regions. A region is closed if traded sectors represent a small share of regional output. Most Indian states produce agricultural goods. The share of Agriculture, Forestry and Fishing exceeds 30 percent of regional output in all Indian States except five: Delhi, Goa, Maharashtra, Pondicherry and Tamil Nadu. If traded sectors were to include agricultural production, then these five regions would be closed. But the opposite is probably true, as they are the richest and most urban in India. Agricultural production is therefore omitted from the criterion used to define regional openness for India. We identify closed Indian States as ones where the production in non-traded sectors exceeds 85 percent of regional output, net of Agriculture, Forestry and Fishing. Open states have high output shares in Mining, Manufacturing, Transport, Storage, Communication, Banking and Insurance.

Figure 1A shows unambiguously that activity does not agglomerate in closed Indian States. In fact, the third panel in Figure 1A shows these closed regions are characterized by increasingly similar production in non-traded sectors. Non-traded activities dis-agglomerate in closed regions. The contrast is complete with what happens in the rest of the country, where economic integration prevails.

Figure 1B plots the analogous graphs for China. From 2003, the Chinese statistical agency started collecting sectoral employment data at regional level for urban units only. A sample focused only on urban employment is problematic for our purposes, as it omits the main employment reserve and source of structural change in China. The sample therefore covers only years between 1995 and 2002. The upper panel presents plots of S_t^G, A_t^G and D_t for China as against per capital GDP in 2005 PPP dollars. China unambiguously diversifies, and that is accompanied with employment agglomerating at regional level. But the index D_t displays a significantly negative trend, so that Chinese regions actually seem to produce an increasingly similar set of goods.

The second and third rows for Figure 1B plot the dynamics of the three indices, computed over sub-samples. As in India, the share of agricultural employment is high in all Chinese regions, and is only low in heavily urbanized regions. So employment in Agriculture, Forestry, Animal Husbandry and Fishing is omitted when determining what regions are relatively closed - doing

otherwise would have all Chinese regions open, except for Shanghai or Beijing. Instead, we define a relatively closed region as one where the average share of employment in Mining, Manufacturing, Transport, Storage, Post, Information Transmission, Computer Service, Software, and Financial Intermediation represents less than 32% of regional employment. The threshold is chosen as the median value across Chinese regions.

In the second row of Figure 1B, specialization is computed on non-traded sectors, agglomeration is computed on closed regions, and dis-similarity is obtained for non-traded sectors between closed regions. As is apparent, non-traded sectors actually diversify in China, while closed regions de-agglomerate. Finally, non-traded activities are the reason why regions become more similar. The last row of Figure 1B focuses on the complementary samples, traded sectors and open regions. It reproduces the dynamics observed in the full sample, with one notable difference: open regions are actually becoming increasingly *dis*-similar in terms of their production of traded goods. In other words, the homogeneization of Chinese regions apparent from the full sample is a reflection of dynamics experienced in activities that are relatively sheltered to trade.

Employment data on the United States comes from the BEA. Coverage includes 63 sectors at the three-digit NAICS level, observed in each of the 51 States from 1980 to 2000. The upper panel of Figure 2 plots the dynamics of S_t^G, A_t^G and D_t against PPP-adjusted real per capita GDP in the US as a whole. The US specializes, in the sense that employment concentrates in few sectors, and the Gini coefficient computed across sectors increases with per capita GDP. Activity dis-agglomerates, and US States produce an increasingly similar set of indutries. US States are completely integrated with each other, with more than a century of free trade in goods and factors. The well-documented specialization of US states in different activities - which corresponds to diversification at the country level - has been achieved and is beginning to reverse itself: since 1980, our findings suggest the US has been specializing as a whole, because the same sectors now tend to be produced across all its States. We conjecture this parallels the well-documented convergence in labor productivity across US states, i.e. cross-state patterns of comparative advantage become less relevant. International productivity differences, i.e. cross-country comparative advantage, now becomes central to explaining both regional and national specialization dynamics. The US - and its constituent States - are specializing according to their global comparative advantage.

Once again such phenomena are expected to prevail in integrating activities and regions of the US. In contrast with developing economies, it is unlikely many activities remain effectively non-

traded at the level of each US state. So the counter-factual that consists in computing S_t^G, A_t^G and D_t for closed sectors and regions is elusive. Such a split is attempted nonetheless, with non traded sectors comprised of Construction, Transport, Storage, Communication, Trade, Personal Services and the Government.[13] The lower panel of Figure 2 reports the results. Diversification also prevails amongst non-traded sectors, with a trend perhaps slightly less pronounced. Closed States are defined as ones where non-traded activities represent more than 80% of State employment. There are ten such States, and employment does in fact agglomerate there, contrary to what happens in the rest of the country.[14] Yet, these closed States do not grow dissimilar in terms of their allocation of non-traded activity. They show similar trend to what is observed elsewhere. Such ambiguous results point to the difficulty in identifying non-traded sectors in a Federation where trade has been free for more than a century. Alternatives where integration is less achieved because of more recent economic or political changes are likely to offer more contrast.

5 European Integration

5.1 Overall Sample

Regional information on the sectoral allocation of employment is available for 15 European countries, comprising historical members such as France or Belgium and accession countries such as Bulgaria, Poland or Slovakia. Figure 3 plots the dynamics of S_{it}^G, A_{it}^G and D_{it} for the available cross-section of European countries, against PPP-adjusted real per capita GDP expressed in 2005 US dollars. In this sample, per capita GDP ranges between $\$5,800$ in Romania (1998) and $\$51,000$ in Norway (2008). With such a broad range, there is a possibility that the dynamics of specialization, agglomeration or similarity between regions vary within sample. Thus, Figure 3 plots the raw panel data (without removing country means), along with fitted regression lines allowing for non-linearities. The fitted estimations isolate within-country variation in the panel, and they correspond to spline regressions over three quantiles. The upper panel in Figure 3 represents the measures of S_{it}^G, A_{it}^G and D_{it} obtained for all available sectors and regions. An upward trend is apparent in S_{it}^G, so that European countries tend to specialize on average as they grow richer. This

[13] Most of these activities are almost certainly traded, at least between US States. So they do not represent an adequate counter-factual.

[14] The ten closed States are: Alaska, DC, Florida, Hawaii, Maryland, Montana, Nevada, New Mexico, North and South Dakota.

is clearly apparent at high income levels - much less so in low income countries. In fact, lowest income countries tend to diversify if anything. The between-country evidence is strongly reminiscent of the U-shaped pattern of specialization documented in Imbs and Wacziarg (2003).

The dynamics of A_{it}^G are less marked, but a weak trend reversal is apparent: low income countries tend to go through geographic agglomeration, while high income countries are characterized by the opposite. And at the same time it is quite clear that the regions constituting these countries become increasingly similar. The within-country dynamics of D_{it} display a downward trend, which is clearest amongst high income (specializing) countries. Once again, European countries are specializing, just as the regions forming each member country produce an increasingly similar set of industries. Our preferred interpretation for this finding is that European countries integrate with each other, and specialize in activities that are different from one country to the other. The regions that constitute each country tend to resemble each other, as it is international comparative advantage (within the EU) that increasingly determines what member countries produce.

Are these trends significant? The upper panel in Table 1 reports the estimated values of β in spline regressions of the type:

$$X_{it} = \alpha_i + \beta\ Y_{it} + \varepsilon_{it}$$

where X_{it} denotes one of the three indices of interest, Y_{it} is real per capita GDP in country i and year t. Each regression is performed on three quantiles of per capita income, defined as "low", "middle" and "high" income samples. The results confirm significant U-shaped dynamics for sectoral specialization, with high income countries specializing. In contrast, agglomeration displays a significant hump shaped relation with per capita GDP: activity agglomerates in low income countries, but dis-agglomerate in middle income samples. Finally, regional dissimilarity falls in middle income countries. The heterogeneity in the significance of the results does partly reflect the coarseness of the data splits. For instance, specialization is still significant in a dataset that would pool middle and high income countries. The same is true of agglomeration amongst low and middle income economies, or of the increase in regional dissimilarity amongs low and mid income countries. Using a non-parametric kernel estimator that eschews such arbitrary data splits, we confirm the generality of our results.

5.2 Splits between Open and Closed Regions or Sectors

Are these dynamics most pronounced in open sectors or regions? As is the case in India and China, it is important to exclude agriculture from the definition of open regions, else all European regions would be classified as open. Consider the share of Agriculture, Mining, Manufacturing, Transport, Telecommunication, Finance and Business Services in total regional employment. In Bulgaria, it ranges from 0.44 to 0.52, and from 0.39 to 0.55 in the Czech Republic. The ranges are virtually identical in France (0.32 - 0.46) or in Belgium (0.32 - 0.43). Of course, it is employment in agricultural activities that explains such high shares in accession countries, whereas it is the rest of conventional traded-goods that does so in historical member countries.

Open regions are therefore defined as ones where the share of Mining, Manufacturing, Transport, Telecommunication, Finance and Business Services in total regional employment exceeds 35%. The threshold corresponds to the median value across the Eurostat sample. Non-traded sectors, in turn, include Construction, Retail, Hotels, Other Services, Public Services and Public Administration. The lower panel of Figure 3 plots the dynamics of the three indices computed on restricted samples. In particular, the Figure illustrates S_{it}^G computed over non-traded sectors, A_{it}^G for closed regions, and D_{it} for non-traded sectors in closed regions.

Specialization is slightly less pronounced amongst non-traded sectors, and no observable differences appear in the agglomeration index. Dissimilarity (between closed regions and non-traded sectors) tends if anything to increase amongst high income countries. The lower panel of Table 1 confirms these results. The non-monotonic relation between sectoral specialization and per capita GDP is confirmed in non-traded sectors. Agglomeration, in turn, is significantly hump-shaped: increasing with per capita GDP at low income levels, and then decreasing in middle income countries. Finally, dissimilarity (in closed regions and in non-traded sectors) increases amongst high income countries. In short, with the exception of the dynamics of D_{it}, little is different across the two samples over which S_{it}^G, A_{it}^G, and D_{it} are computed.

In fact, the contrast between the two panels in Figure 3 and Table 1 is obscured by the very nature of the available information on European regions. Eurostat effectively starts gathering data as a country is well-advanced along the accession path. For instance data are collected from 2003 in Bulgaria, and from 1998 in the Czech Republic or in Slovakia. These do not represent actual accession dates, but they do reflect an ongoing integration process between Eastern European economies and the rest of the European Union (EU). By construction therefore, the sample put

together by Eurostat is biased against autarkic regions in the EU. The very definition of closed regions or sectors is therefore elusive in such a selection of country and year coverage. In that sense, the European case is similar to the US. But here it is because of the bias in the data collection process, rather than because of deep economic integration, although it is likely also the case that European regions, even for relatively recent entrants into the EU, are more regionally integrated than their counterparts in developing countries like India and China. For all these reasons, splits between open and closed regions in Europe are perhaps less relevant than they are for China and India.

5.3 An Alternative Test

Fortunately, an alternative test may help evaluate this paper's major conjecture that economic integration drives observed patterns of specialization, agglomeration and inter-regional similarity. The claim is that European countries are formed of increasingly homogeneous regions, because each country specializes in different activities as it integrates with the rest of the Union. As a result, the economic area formed by an aggregate of European countries should in fact diversify - exactly the opposite from what each country goes through in isolation. Moreover, economic activity in this European aggregate should agglomerate at country level, while each country becomes increasingly dissimilar from the rest of the Union. No regional information is necessary to establish these claims. We can investigate them in *country*-level data on sectoral activity, where the sampling problem present in Eurostat data does not exist.

The question is asked within employment data collected by the International Labor Office (ILO), with information on one-digit employment from 1992 to 2008 for 13 European countries.[15] Coverage now includes sectoral employment in accession countries substantially before the process of entry into the EU, so that meaningful sample splits are possible. Figure 4 reports the dynamics of S_t^G, A_t^G and D_t for Europe as a whole, against European per capita GDP. The upper panel plots the measures computed over the full sample of sectors and countries. It is clear that Europe as a whole diversifies, even though its member countries actually specialize. Table 2 reports the actual estimated slope of the relation, which is negative and significant at 1% confidence level. It must therefore be the case that different European countries specialize in different sectors. This is

[15] Relative to the Eurostat sample, France and Romania were discarded. ILO reports no data for France prior to 2003. Data for the Agriculture sector in Romania display implausible volatility, with yearly growth rates in excess of 25%.

confirmed by the positive trend in A_{it}^{G} in Europe as a whole, apparent in Figure 4, and significant at 1% confidence level in Table 2. European countries specialize, but in different activities, so that Europe overall diversifies.

Do European countries become more dissimilar? The trend in D_t is downward, and significant at the 5% confidence level. This suggests European countries, if anything, become more similar. But the measure corresponds to a sample where some countries are relatively closed to trade, in Eastern Europe at the beginning of the 1990's. The lower panel of Figure 4 focuses on non-traded sectors and/or relatively closed countries. A country is categorized as closed if average employment in Mining, Quarrying, Manufacturing, Transport, Storage, Communication, Finance, Insurance, Real Estate and Business Services is below its median value, 45%. According to this definition, four European countries are on average closed to trade over the period considered: Bulgaria, Greece, Poland and Portugal. Our conjecture is that D_t trends downward in the full sample because non-traded sectors tend to develop similarly, which obscures increased dissimilarity amongst traded activities in open countries. Figure 4 confirms that D_t does actually trend upwards in an integrating subsample, with a coefficient that is significantly positive in Table 2. When focused on traded sectors, open European countries actually become dissimilar (consistent with results in Midelfart-Knarvik et al., 2000). Interestingly, specialization prevails amongst traded sectors, and agglomeration is significantly more pronounced amongst open countries, with a coefficient estimate four times larger in Table 2.

Europe provides a laboratory that confirms this paper's conjecture. Structural change is associated with both local and global economic integration. Taken individually, European countries specialize - and activity is allocated increasingly equally across their regions. As a result, European regions become increasingly homogeneous within each country. This happens as each country integrates outside of its borders, exactly in the same way as the US as a whole did. Contrary to the US, however, it is possible to construct a sample formed of the very countries that integrate with each other, i.e. an integrating area formed of specializing countries. This European free trade area does in fact diversify as a whole, as activity agglomerates at the country level, and member countries become dissimilar - all of which prevails most significantly in traded sectors amongst open countries. For Europe as a whole, therefore, integration *within its borders* remains the dominant force behind structural change relative to integration with the rest of the world.

6 Global Integration

Data on sectoral employment at regional level are available from the Integrated Public Use Microdata Series (IPUMS) project, for a maximum of 28 countries, of which 9 are developed.[16] The data focus on developing economies, with two thirds of the observations on per capita GDP below $\$15,000$ at 2005 PPP exchange rates. Regional data at such stages of development are the main benefit of using IPUMS data. This comes at the expense of the time series dimension, however, with an average of 3.7 years available for each region, and 3 years in low income countries. The upper row in Figure 5 reports the raw panel data for the values of S_{it}^G, A_{it}^G, and D_{it} computed on the full available sample. Each graph also plots the fitted value corresponding to spline regressions on three quantiles of per capita GDP. The coefficients are estimated on the within-country variation, as country fixed effects are included. Diversification is pronounced at values of per capita GDP below $\$15,000$, which is also where activities agglomerate regionally. An upward trend in D_{it} is also observable below $\$15,000$ per capita GDP, but less markedly. At higher income levels, these trends tend to revert. But fewer observations are available, and within-country estimates are less precise.

The upper panel of Table 3 investigates the significance of these patterns. The three quantiles of per capita GDP used to perform the spine regressions are labeled "low", "middle" and "high" income groups. Diversification is significant in low and middle income groups, as is agglomeration in low income countries. The increase in dissimilarity is insignificant in the low income sample. At high income levels, specialization is insignificant, but A_{it}^G and D_{it} both display significantly negative trends.

Per capita GDP ranges from \$870 to $\$41,900$ in the IPUMS sample. The dispersion in stages of development - and in the degree of domestic versus international integration - is considerable. Given such dispersion, it is important to impose a definition of relatively closed regions in a country-specific manner. An absolute threshold would ignore the possibility that some regions are relatively closed in an open country, or relatively open in a closed one. A region is defined as relatively closed if the share of employment in traded sectors is below the country average. As before, agriculture production is omitted from the definition of regional openness, as doing otherwise would classify autarkic, agricultural regions as open. An open region is one where employment in Mining, Manufacturing,

[16] The complete list of countries, along with geographic, year and industry coverage, is included in the data Appendix.

Transport, Telecommunications, Finance and Business Services represents a larger share of regional employment than in the country as a whole.

The lower panel of Figure 5 plots S_{it}^G, A_{it}^G, and D_{it} against per capita GDP, for integrating sub-samples. Specialization S_{it}^G is now measured on traded sectors only, and diversification is now more pronounced at all income levels.[17] Agglomeration A_{it}^G, and especially dissimilarity D_{it} are now markedly hump-shaped. The lower panel of Table 3 confirms that these trends are significant. In particular, relative to the full sample, integrating regions in low income countries are now clearly becoming more dissimilar, consistent with our theory.[18]

7 Conclusion

This paper documents a robust pattern of structural change through the development process. Poor countries are sectorally concentrated, but diversify their sectoral base in early stages of development. This sectoral diversification goes hand in hand with regional agglomeration, so that the regions of a given developing economy look increasingly dissimilar from each other. Such regional specialization is what affords aggregate sectoral diversification, through the force of (regional) comparative advantage. With regional specialization comes regional convergence in productivity, i.e. a withering of regional comparative advantage. A turning point then occurs, as international comparative advantage becomes of the essence. The country then starts to specialize overall, as do all of its constituent regions. Regional agglomeration falls, as regions produce increasingly similar goods for the international market. Regions become increasingly similar in terms of what they produce.

We interpret these broad patterns as resulting from a process of gradual economic integration, first among a country's constituent regions and later among countries themselves. We find strong evidence in favor of this interpretation by contrasting the patterns of specialization, agglomeration and structural similarity between tradable and non tradable sectors, and between open and closed regions. The patterns are systematicaly absent from relatively sheltered regions and activities. The results are obtained in a variety of datasets, a multiplicity of specialization measures, and

[17] As before, Agriculture is considered to be a traded good when it comes to splitting traded and non-traded activities (but not when it comes to defining closed regions).

[18] The IPUMS data leave somewhat to be desired, because of their limited time dimension. An alternative is provided by the Luxembourg Income Survey, which offers more time variation. We plan to make use of these data in future versions of this paper.

a multiplicity of approaches. We present individual country case studies, examine multi-country zones of economic integration such as the EU, where a region is in fact a country. We also present systematic international evidence, for which we assembled a comprehensive and unique dataset of sectoral data at the regional level over time.

If structural change is driven by economic integration, it is a symptom, rather than a policy tool. It reflects the joint influences of local and global economic integration. Sectoral diversification is often considered a desirable goal of economic policy, because it limits aggregate volatility, and dilutes the aggregate consequences of terms of trade shocks. Our evidence suggests that sectoral diversification builds on domestic economic integration. Irrespective whether diversification is desirable from a welfare standpoint, it results from infrastructure investments, roads, railroads, inter-regional connectivity, and the reduction in inter-regional impediments to trade and factor movements. To the extent that sectoral diversification is a policy objective, it is probably best achieved through domestic integration rather than through the heavy hand of industrial policy, with well known drawbacks.

Our findings also imply that sectoral specialization is a natural outcome of regional convergence in productivity and factor endowments. As inter-regional differences in comparative advantage disappear, international integration takes front stage, and countries specialize. At that intermediate stage of development, preserving sectoral diversification would imply protectionist measures, necessary to counteract the importance of international relative to intranational comparative advantage. This paper is silent on welfare, but it underlines the key tradeoff represented by such distortive policies in upper middle income countries constituted of integrated regions.

References

Barro, Robert J. and Xavier Sala-i-Martin (2004), *Economic Growth*, MIT Press.

Caselli, Francesco and John W. Coleman III (2001), "The US Structural Transformation and Regional Convergence: A Reinterpretation," *Journal of Political Economy*, vol. 109, no. 3. pp. 584-616.

Chenery, Hollis, Sherman Robinson and Moshe Syrquin (1986), *Industrialization and Growth: A Comparative Study*. Oxford: Oxford University Press (for the World Bank).

Deichmann, Uwe, Somik V. Lall, Stephen J. Redding and Anthony J. Venables (2008), "Industrial Location in Developing Countries", *The World Bank Research Observer*, vol. 23, no. 2, pp. 219-246.

Desmet, Klaus and Esteban Rossi-Hansberg (2011), "Spatial Development", *working paper*, Universidad Carlos III de Madrid.

Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography, and Trade," *Econometrica*, vol. 70: pp. 1741-1779.

Findlay, Ronald (1970), "Factor Proportions and Comparative Advantage in teh Long Run", *Journal of Political Economy*, vol. 78, no. 1, pp. 27-34.

Ghosh, B., Marjit, S. and Neogi, C. (1998), "Economic growth and regional divergence in India, 1960 to 1995," *Economic and Political Weekly*, vol. 33, no. 26, pp. 1623–1630.

Grossman, G. and E. Helpman, *Innovation and Growth in the Global Economy*, Cambridge: MIT Press, 1991.

Imbs, Jean and Romain Wacziarg (2003), "Stages of Diversification," *American Economic Review*, vol. 93, no. 1, p. 63-86.

Jones R. W. (1961) "Comparative Advantage and the Theory of Tariffs: A Multi-Country, Multi-Commodity Model," *Review of Economic Studies*, vol. 28, pp. 161-175.

Kim, Sukkoo (1995), "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860-1987", *Quarterly Journal of Economics*, vol. 110, no. 4, pp. 881-908.

Koren, Miklós and Silvana Tenreyro (2007), "Volatility and Development," *Quarterly Journal of Economics*, vol. 122, no. 1, pp. 243–287.

Krugman (1991a), "Increasing Returns and Economic Geography", *Journal of Political Economy*, vol. 99, no. 3. pp. 483-499.

Krugman, Paul (1991b), *Geography and Trade*, Cambridge: MIT Press.

Krugman, Paul and Anthony Venables (1995), "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, vol. 110, no. 4, pp. 858-880

Kuznets, Simon S. (1966), *Modern Economic Growth: Rate, Structure, and Spread*, New Haven, CT: Yale University Press.

Melitz, Marc J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, vol. 71(6), p. 1695-1725, November.

Midelfart-Knarvik, K. H., H. G. Overman, S. J. Redding, A. J. Venables (2000), "The Location of European Industry", *Report prepared for the Directorate General for Economic and FInancial Affairs, European Commission.*

Puga, Diego and Anthony Venables (1998), "Agglomeration and Economic Development: Import Substitution vs. Trade Liberalization," *working paper*, Center for Economic Performance, London School of Economics.

Redding, Stephen (2002), "Specialization Dynamics," *Journal of International Economics*, vol. 58, pp. 299-334.

Redding, Stephen (2009), "The Empirics of New Economic Geography," *working paper*, London School of Economics, May.

Ventura, Jaume (1997), "Growth and Interdependence," *Quarterly Journal of Economics*, vol. 112, no. 1, pp. 57–84.

Table 1: Eurostat - Regional Data

	Specialization	Agglomeration	Dissimilarity
Low	7.22 (1.31)	8.39^{***} (3.76)	-1.13 (-0.88)
Middle	-5.73^{*} (-1.82)	-3.10^{**} (-2.42)	-5.38^{***} (-7.30)
High	10.70^{***} (2.90)	-2.42 (-1.61)	-0.50 (-0.58)
Obs.	166	166	166
	Specialization (NT)	Agglomeration (Closed)	Dissimilarity (ClosNT)
Low	0.75 (0.17)	-1.15 (-0.28)	-2.38 (-0.39)
Middle	-13.00^{***} (-5.16)	-0.31 (-0.06)	1.40 (0.53)
High	6.61^{**} (2.24)	1.42 (0.48)	6.29^{*} (1.91)
Obs.	166	113	113

Notes: The Table reports the estimates of spline regressions performed on three quantiles of the data, labeled "Low", "Middle" and "High". All estimations include country-specific fixed effects. Coefficients are multiplied by 10^7. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 10% (5%, 1%) confidence level.

Table 2: ILO - Sectoral Data

	Specialization	Agglomeration	Dissimilarity
EU per capita GDP	-9.36^{***} (-3.39)	6.10^{***} (3.59)	-1.20^{**} (-2.33)
Obs.	17	17	17
	Specialization (T)	Agglomeration (Open)	Dissimilarity (OpenT)
EU per capita GDP	-9.88^{***} (-9.18)	25.00^{***} (8.15)	16.90^{***} (11.20)
Obs.	17	17	17

Notes: Coefficients are multiplied by 10^7. Student's t-statistics are reported between parentheses. *** (**,*) denote significance at 10% (5%, 1%) confidence level.

Table 3: IPUMS - Regional Data

	Specialization	Agglomeration	Dissimilarity
Low	-228.23^{***} (-9.29)	47.30^{***} (3.32)	10.00 (1.44)
Middle	-75.30^{***} (-4.94)	4.70 (0.53)	-18.30^{***} (-4.32)
High	0.69 (0.14)	-5.97^{**} (-2.15)	-8.03^{***} (-5.91)
Obs.	103	103	103
	Specialization (T)	Agglomeration (Open)	Dissimilarity (OpenT)
Low	-207.15^{***} (-5.98)	45.32^{***} (2.71)	118.20^{***} (6.06)
Middle	-118.21^{***} (-5.48)	-5.24 (-0.50)	-51.20^{***} (-4.24)
High	-12.80^{*} (-1.90)	-5.95^{*} (-1.82)	-6.79^{*} (-1.79)
Obs.	103	103	103

Notes: The Table reports the estimates of spline regressions performed on three quantiles of the data, labeled "Low", "Middle" and "High". All estimations include country-specific fixed effects. Coefficients are multiplied by 10^7. Student's t-statistics are reported between parentheses. *** (**,*) denote significance at 10% (5%, 1%) confidence level.

Appendix 1 - Theoretical results for extensions 1, 2 and 3.

Extension 1. A nontradable good

Suppose good 3 is nontradable at all times. Then we have 2 goods and 3 countries. At time 0, nothing changes compared to the generic $3 \times 3 \times 3$ case. Region 1 produces $Y_{11} = a_{11}L_1$, region 2 produces $Y_{21} = a_{21}L_2$ and region 3 produces $Y_{31} = a_{31}L_3$:

$$\begin{aligned} S^H &= 1 \\ A^H &= 1 - 2\left(\frac{a_{11}L_1a_{21}L_2 + a_{11}L_1a_{31}L_3 + a_{21}L_2a_{31}L_3}{(a_{11}L_1 + a_{21}L_2 + a_{31}L_3)^2}\right) < 1 \\ D &= D_1 = 0 \end{aligned}$$

At time 1, the country is still autarkic but the regions can now trade, except in good 3. Goods are produced where autarky relative prices are lowest. In particular, assume:

$$\frac{a_{11}}{a_{12}} > \frac{a_{31}}{a_{32}} > \frac{a_{21}}{a_{22}}$$

There are various possibilities, but region 1 will tend to produce good 1, region 2 will tend to produce good 2, and region 3 may produce either good 1 or 2 or both. Assume that demand, productivities and sizes are such that region 3 produces goods 1 and 2. So region 1 will produce good 1, region 2 produces good 2 and region 3 produces good 1 and 2. All regions produce good 3 since it is nontraded. Then region 3 pins down the relative price of good 1 in terms of good 2:

$$\frac{p_1}{p_2} = \frac{a_{32}}{a_{31}}$$

Relative wages immediately follow:

$$\frac{w_1}{w_2} = \frac{a_{32}}{a_{31}}\frac{a_{11}}{a_{22}}$$

Then we can solve for consumption and remaining relative prices:

$$\begin{aligned} C_{11} &= \frac{a_{11}L_1}{3}; C_{12} = \frac{a_{11}L_1}{3}\frac{a_{32}}{a_{31}}; C_{13} = \frac{a_{13}L_1}{3} \\ C_{21} &= \frac{a_{22}L_2}{3}\frac{a_{31}}{a_{32}}; C_{22} = \frac{a_{22}L_2}{3}; C_{23} = \frac{a_{23}L_2}{3} \\ C_{31} &= \frac{a_{31}L_3}{3}; C_{32} = \frac{a_{32}L_3}{3}; C_{33} = \frac{a_{33}L_3}{3} \\ \frac{p_{13}}{p_1} &= \frac{a_{11}}{a_{13}}; \frac{p_2}{p_{23}} = \frac{a_{23}}{a_{22}}; \frac{p_{33}}{p_2} = \frac{a_{32}}{a_{33}} \end{aligned}$$

Finally we can derive production:

$$\begin{aligned} Y_{11} &= \frac{a_{11}L_1}{3} + \frac{a_{22}L_2}{3}\frac{a_{31}}{a_{32}}; Y_{12} = 0; Y_{13} = \frac{a_{13}L_1}{3} \\ Y_{21} &= 0; Y_{22} = \frac{a_{11}L_1}{3}\frac{a_{32}}{a_{31}} + \frac{a_{22}L_2}{3}; Y_{23} = \frac{a_{23}L_2}{3} \\ Y_{31} &= \frac{a_{31}L_3}{3}; Y_{32} = \frac{a_{32}L_3}{3}; Y_{33} = \frac{a_{33}L_3}{3} \end{aligned}$$

This is enough to calculate indices of diversification, agglomeration and similarity either for labor per sector or output per sector:

$$\begin{aligned} S^H &= \sum_s \left(\frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} \right)^2 \\ &= 1 - \frac{2\left(a_{11}L_1 + a_{31}L_3 + a_{22}L_2\frac{a_{31}}{a_{32}}\right)\left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right)}{\left(\left(a_{11}L_1 + a_{31}L_3 + a_{22}L_2\frac{a_{31}}{a_{32}}\right) + \left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right) + (a_{13}L_1 + a_{23}L_2 + a_{33}L_3)\right)^2} \\ &\quad - \frac{2\left(a_{11}L_1 + a_{31}L_3 + a_{22}L_2\frac{a_{31}}{a_{32}}\right)(a_{13}L_1 + a_{23}L_2 + a_{33}L_3)}{\left(\left(a_{11}L_1 + a_{31}L_3 + a_{22}L_2\frac{a_{31}}{a_{32}}\right) + \left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right) + (a_{13}L_1 + a_{23}L_2 + a_{33}L_3)\right)^2} \\ &\quad - \frac{2\left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right)(a_{13}L_1 + a_{23}L_2 + a_{33}L_3)}{\left(\left(a_{11}L_1 + a_{31}L_3 + a_{22}L_2\frac{a_{31}}{a_{32}}\right) + \left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right) + (a_{13}L_1 + a_{23}L_2 + a_{33}L_3)\right)^2} \\ &\quad - \frac{2\left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right)(a_{13}L_1 + a_{23}L_2 + a_{33}L_3)}{\left(\left(a_{11}L_1 + a_{31}L_3 + a_{22}L_2\frac{a_{31}}{a_{32}}\right) + \left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right) + (a_{13}L_1 + a_{23}L_2 + a_{33}L_3)\right)^2} \end{aligned}$$

So $S^H < 1$

The country is now diversified. Is it more diversified than when one good was not tradable? This S^H must be compared to $1 - \frac{2(a_{11}L_1a_{22}L_2 + a_{11}L_1a_{33}L_3 + a_{22}L_2a_{33}L_3)}{(a_{11}L_1 + a_{22}L_2 + a_{33}L_3)^2}$. In the fraction here, both the numerator and the denominator are unambiguously larger than in the full free trade case, so the answer is ambiguous.

Geographic agglomeration is:

$$\begin{aligned}
A_1^H &= \sum_j \left(\frac{Y_{j1}}{\sum_j Y_{j1}}\right)^2 = 1 - \frac{2\left(a_{11}L_1 + a_{22}L_2\frac{a_{31}}{a_{32}}\right)(a_{31}L_3)}{\left(a_{11}L_1 + a_{22}L_2\frac{a_{31}}{a_{32}} + a_{31}L_3\right)^2} < 1 \\
\text{Similarly, } A_2^H &= \sum_j \left(\frac{Y_{j2}}{\sum_j Y_{j2}}\right)^2 = 1 - \frac{2\left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2\right)(a_{32}L_3)}{\left(a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{32}L_3\right)^2} < 1 \\
A_3^H &= \sum_j \left(\frac{Y_{j3}}{\sum_j Y_{j3}}\right)^2 \\
&= 1 - \frac{2(a_{13}L_1)(a_{23}L_2) + 2(a_{13}L_1)(a_{33}L_3) + 2(a_{23}L_2)(a_{33}L_3)}{(a_{13}L_1 + a_{23}L_2 + a_{33}L_3)^2} < 1 \\
\text{Thus, } A^H &= \sum_s \frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} A_s^H < 1
\end{aligned}$$

The presence of nontradables makes the country look less agglomerated than when tradables were present. There are two reasons: 1) the nontradable good is produced in every region and 2) goods 1 and 2 are produced in each of regions 1 and 2, respectively, but also in region 3. The country would still be less agglomerated than in the generic case if region 3 were not diversified, but produced only one of goods 1 or 2.

The degree of dissimilarity between regions is:

$$\begin{aligned}
\text{If } s_{11} &> s_{31}, \text{ then } D_1 = \frac{1}{3}\left[2 - \frac{2a_{13}a_{32}L_1}{a_{11}a_{32}L_1 + a_{22}L_2a_{31} + a_{13}a_{32}L_1}\right] < \frac{2}{3} \\
\text{If } s_{11} &< s_{31}, \text{ then } D_1 = \frac{1}{3}\left[\frac{2Y_{31}}{\sum_s Y_{3s}}\right] = \frac{1}{3}\left[\frac{a_{31}L_3}{a_{31}L_3 + a_{32}L_3 + a_{33}L_3}\right] < \frac{1}{3} \\
\text{If } s_{22} &> s_{32}, \text{ then } D_2 = \frac{2}{3}\left[\frac{a_{11}L_1a_{32} + a_{22}a_{31}L_2}{a_{11}L_1a_{32} + a_{22}a_{31}L_2 + a_{23}a_{31}L_2}\right] < \frac{2}{3} \\
\text{If } s_{22} &< s_{32}, \text{ then } D_2 = \frac{2}{3}\left[\frac{a_{32}L_3}{a_{31}L_3 + a_{32}L_3 + a_{33}L_3}\right] < \frac{2}{3} \\
D_3 &= \frac{1}{3}\left[\left|\frac{a_{13}L_1}{a_{11}L_1 + a_{22}L_2\frac{a_{31}}{a_{32}} + a_{13}L_1} - \frac{a_{23}L_2}{a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{23}L_2}\right|\right. \\
&\quad + \left|\frac{a_{13}L_1}{a_{11}L_1 + a_{22}L_2\frac{a_{31}}{a_{32}} + a_{13}L_1} - \frac{a_{33}L_3}{a_{31}L_3 + a_{32}L_3 + a_{33}L_3}\right| \\
&\quad \left. + \left|\frac{a_{23}L_2}{a_{11}L_1\frac{a_{32}}{a_{31}} + a_{22}L_2 + a_{23}L_2} - \frac{a_{33}L_3}{a_{31}L_3 + a_{32}L_3 + a_{33}L_3}\right|\right] \\
&\quad \text{This is likely to be small due to demand being identical} \\
&\quad \text{across regions and good 3 being produced everywhere} \\
&\quad \text{(i.e. shares of good 3 in regional economies are relatively balanced),} \\
&\quad \text{but we can't say anything definitive without more assumptions.} \\
\text{So } D &= \frac{1}{S}\sum_s D_s \text{ is likely to be smaller than 2/3.}
\end{aligned}$$

Average dissimilarity between regions will tend to be lower than in the generic case because of the presence of nontradables, which are produced everywhere, and because region 3 may produce at least 2 goods. However, it is hard to get definitive analytical results on this point without additional assumptions.

At time 2, we now consider three countries, A, B and C. Assume that country A has a comparative advantage in sector 1, B in sector 2 and C is intermediate:

$$\frac{a_1^A}{a_2^A} > \frac{a_1^C}{a_2^C} > \frac{a_1^B}{a_2^B}$$

(these assumptions on a_1^i, a_2^i and a_3^i, $\forall i \in \{A, B, C\}$ are analogous to the ones made for time 1 with respect to regional comparative advantage). We again assume that country C will produce all

3 goods. Then by a reasoning exactly identical to the one for time 1, we have:

$$\begin{aligned} Y_1^A &= \frac{a_1^A L^A}{3} + \frac{a_2^B L^B}{3}\frac{a_1^C}{a_2^C}; Y_2^A = 0; Y_3^A = \frac{a_3^A L^A}{3} \\ Y_1^B &= 0; Y_2^B = \frac{a_1^A L^A}{3}\frac{a_2^C}{a_1^C} + \frac{a_2^B L^B}{3}; Y_3^B = \frac{a_3^B L^B}{3} \\ Y_1^C &= \frac{a_1^C L^C}{3}; Y_2^C = \frac{a_2^C L^C}{3}; Y_3^C = \frac{a_3^C L^C}{3} \end{aligned}$$

Regions produce identical amounts of each goods within each country (1/3 of the aggregate amounts in the array immediately above):

Country A:

$$\begin{aligned} Y_{11}^A &= Y_{21}^A = Y_{31}^A = \frac{a_1^A L^A}{9} + \frac{a_2^B L^B}{9}\frac{a_1^C}{a_2^C} \\ Y_{12}^A &= Y_{22}^A = Y_{32}^A = 0 \\ Y_{13}^A &= Y_{23}^A = Y_{23}^A = \frac{a_3^A L^A}{9} \end{aligned}$$

Country B:

$$\begin{aligned} Y_{11}^B &= Y_{21}^B = Y_{31}^B = 0 \\ Y_{12}^B &= Y_{22}^B = Y_{32}^B = \frac{a_1^A L^A}{9}\frac{a_2^C}{a_1^C} + \frac{a_2^B L^B}{9} \\ Y_{13}^B &= Y_{23}^B = Y_{23}^B = \frac{a_3^B L^B}{3} \end{aligned}$$

Country C (case of diversification):

$$\begin{aligned} Y_{11}^C &= Y_{21}^C = Y_{31}^C = \frac{a_1^C L^C}{9} \\ Y_{12}^C &= Y_{22}^C = Y_{32}^C = \frac{a_2^C L^C}{9} \\ Y_{13}^C &= Y_{23}^C = Y_{23}^C = \frac{a_3^C L^C}{9} \end{aligned}$$

Because of the non-tradable good 3, countries are imperfectly specialized. Among tradable goods, A is specialized in good 1, B in good 2 and C may or may not be specialized in either good (in the equilibrium above, it is assumed to produce both - that depends on assumptions on model parameters). There is complete regional disagglomeration within each country, which each country's respective sector of production located uniformly across regions. Regions within each country are identical.

The country used to produce 3 goods now it only produces 2. We can compute the countrywide Herfindahl index of diversification as:

$$S^H = 1 - \frac{2\left(a_1^A L^A + a_2^B L^B \frac{a_1^C}{a_2^C}\right)\left(a_3^A L^A\right)^2}{\left(a_1^A L^A + a_2^B L^B \frac{a_1^C}{a_2^C} + a_3^A L^A\right)^2} < 1$$

Compared to the generic case, the country is less concentrated, i.e. more diversified than before (country C is even more diversified, unless it produces only one of the tradables).

Geographic agglomeration is:

$$\begin{aligned} A_1^H &= \frac{1}{3} \\ A_2^H &= \text{undefined since sector 2 is not produced in country A} \\ A_3^H &= \frac{1}{3} \\ A^H &= \frac{1}{3} \end{aligned}$$

Compared to the generic case, the country is equally dis-agglomerated.

The degree of dissimilarity between regions is:

$$\begin{aligned} D_1 &= 0 \\ D_2 &= \text{undefined (sector 2 not produced anywhere)} \\ D_3 &= 0 \\ D &= \frac{1}{S}\sum_s D_s = \frac{1}{3}(D_1 + D_2 + D_3) = 0 \end{aligned}$$

Compared to the generic case, the country has experienced the same amount of structural convergence (complete).

Extension 2: A closed region

We now assume that region 3 remains closed throughout times 0, 1 and 2.

Time 0

Nothing changes. Region 1 produces $Y_{11} = a_{11}L_1$, region 2 produces $Y_{21} = a_{21}L_2$ and region 3 produces $Y_{31} = a_{31}L_3$. Since goods 2 and 3 are not yet available to be produced, $Y_{12} = Y_{22} = Y_{32} = Y_{13} = Y_{23} = Y_{33} = 0$. Moreover, since there is no interregional trade, $C_{js} = Y_{js}$ for all j, s. Welfare is $U_1(C_{11}, C_{12}, C_{13}) = U_2(C_{21}, C_{22}, C_{23}) = U_2(C_{31}, C_{32}, C_{33}) = 0$. Only one good is available for consumption so welfare is zero (normalization).

Time 1

At time 1, region 3 remains closed and therefore features the autarky allocation. We assume region 3 retains its specialized structure since it is more in line with the idea of indivisibilities preventing new sectors from opening up at time 1 in closed regions. Then region 3 looks the way it does at time 0: $Y_{31} = a_{31}L_3; Y_{32} = Y_{33} = 0$ (if region 3 instead diversifies, then $Y_{31} = a_{31}L_3/3$, and $Y_{32} = a_{32}L_3/3$ and $Y_{33} = a_{33}L_3/3$).

For regions 1 and 2, we have a Ricardian model with 3 sectors and 2 regions. We can write a chain of comparative advantage:

$$\frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

With this assumption, 5 distinct cases arise: As is well-known, the equilibrium relative wage w_1/w_2 determines which case we are in. Cases are as follows:

1) Region 1 produces good 1, region 2 produces goods 1, 2, 3:

$$\frac{w_1}{w_2} = \frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

2) Region 1 produces good 1, region 2 produces goods 2 and 3:

$$\frac{a_{11}}{a_{21}} > \frac{w_1}{w_2} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

3) Region 1 produces goods 1 and 3, region 2 produces goods 2 and 3 (perhaps the focal case):

$$\frac{a_{11}}{a_{21}} > \frac{w_1}{w_2} = \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}}$$

4) Region 1 produces goods 1 and 3, region 2 produces good 2:

$$\frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{w_1}{w_2} > \frac{a_{12}}{a_{22}}$$

5) Region 1 produces goods 1, 2 and 3, region 2 produces good 2:

$$\frac{a_{11}}{a_{21}} > \frac{a_{13}}{a_{23}} > \frac{a_{12}}{a_{22}} = \frac{w_1}{w_2}$$

Note that from the viewpoint of our measures of specialization, agglomeration and similarity, regions 1 and 2 are interchangeable, as are sectors 1 and 2. Thus, cases 1 and 2 are essentially the same as cases 4 and 5 from the viewpoint of predicted indices. So we limit ourselves to solving for cases 1, 2, 3.

Good one has the same price p_1 in regions 1 and 2 and so do good 2 (p_2) and good 3 (p_3) (arbitrage): $p_{11} = p_{21} \equiv p_1$; $p_{12} = p_{22} \equiv p_2$; $p_{13} = p_{23} \equiv p_3$. Together with the first order

conditions, budget constraints, etc, this allows us to solve fully for all C_{ij}, relative prices, and relative wages.

Case 1: Region 1 produces good 1, region 2 produces goods 1, 2, 3. Solving:

$$\begin{aligned}
C_{11} &= \frac{a_{11}L_1}{3}; C_{12} = \frac{a_{11}a_{22}L_1}{3a_{21}}; C_{13} = \frac{a_{11}a_{23}L_1}{3a_{21}}; \\
C_{21} &= \frac{a_{21}L_2}{3}; C_{22} = \frac{a_{22}L_2}{3}; C_{23} = \frac{a_{23}L_2}{3} \\
\frac{w_1}{w_2} &= \frac{a_{11}}{a_{21}}; \frac{p_1}{p_2} = \frac{a_{22}}{a_{21}}; \frac{p_2}{p_3} = \frac{a_{23}}{a_{22}} \\
Y_{11} &= a_{11}L_1 \\
Y_{21} &= \frac{a_{21}L_2 - 2a_{11}L_1}{3} \\
Y_{22} &= a_{22}\left[\frac{a_{11}L_1 + a_{21}L_2}{3a_{21}}\right] \\
Y_{23} &= a_{23}\left[\frac{a_{11}L_1 + a_{21}L_2}{3a_{21}}\right]
\end{aligned}$$

Note that prices and wages are pinned down only in relative terms.

Case 2: Region 1 produces good 1, region 2 produces goods 2 and 3

$$\begin{aligned}
C_{11} &= \frac{a_{11}L_1}{3}; C_{12} = \frac{a_{22}L_2}{6}; C_{13} = \frac{a_{23}L_2}{6} \\
C_{21} &= \frac{2a_{11}L_1}{3}; C_{22} = \frac{a_{22}L_2}{3}; C_{23} = \frac{a_{23}L_2}{3} \\
\frac{w_1}{w_2} &= \frac{L_2}{2L_1}; \frac{p_1}{p_2} = \frac{a_{22}L_2}{2a_{11}L_1}; \frac{p_1}{p_3} = \frac{a_{23}L_2}{2a_{11}L_1} \\
Y_{11} &= a_{11}L_1 \\
Y_{22} &= \frac{a_{22}L_2}{2}; Y_{23} = \frac{a_{23}L_2}{2}
\end{aligned}$$

Case 3: Region 1 produces goods 1 and 3, region 2 produces goods 2 and 3 (a focal case):

$$\begin{aligned}
C_{11} &= \frac{a_{11}L_1}{3}; C_{12} = \frac{a_{13}a_{22}L_1}{3a_{23}}; C_{13} = \frac{a_{13}L_1}{3} \\
C_{21} &= \frac{a_{11}a_{23}L_2}{3a_{13}}; C_{22} = \frac{a_{22}L_2}{3} C_{23} = \frac{a_{23}L_2}{3} \\
\frac{w_1}{w_2} &= \frac{a_{13}}{a_{23}}; \frac{p_1}{p_2} = \frac{a_{13}}{a_{11}}\frac{a_{22}}{a_{23}}; \frac{p_2}{p_3} = \frac{a_{23}}{a_{22}} \\
Y_{11} &= a_{11}\left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{13}}\right); Y_{13} = a_{13}\left(\frac{2a_{13}L_1 - a_{23}L_2}{3a_{13}}\right) \\
Y_{22} &= a_{22}\left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{23}}\right); Y_{23} = a_{23}\left(\frac{2a_{23}L_2 - a_{13}L_1}{3a_{23}}\right)
\end{aligned}$$

What is left is computing indices of diversification, agglomeration and similarity in all 3 cases, and comparing them to the generic case. To save space, we limit ourselves to doing so for case 3:

We can compute the countrywide Herfindahl index of diversification as:

$$\begin{aligned} S^H &= \sum_s \left(\frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} \right)^2 \\ &= \frac{\left[a_{11} \left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{13}} \right) + a_{31}L_3 \right]^2 + \left[a_{22} \left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{23}} \right) \right]^2 + \left[a_{13} \left(\frac{2L_1}{3} - \frac{a_{23}L_2}{3a_{13}} \right) + a_{23} \left(\frac{2L_2}{3} - \frac{a_{13}L_1}{3a_{23}} \right) \right]^3}{\left(a_{11} \left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{13}} \right) + a_{13} \left(\frac{2L_1}{3} - \frac{a_{23}L_2}{3a_{13}} \right) + a_{23} \left(\frac{2L_2}{3} - \frac{a_{13}L_1}{3a_{23}} \right) + a_{31}L_3 \right)^2} \end{aligned}$$

The country still produces all three goods and is thus diversified, but whether it is more or less diversified than under full open trade is ambiguous.

Geographic agglomeration is:

$$\begin{aligned} A_1^H &= \sum_j \left(\frac{Y_{j1}}{\sum_j Y_{j1}} \right)^2 = \frac{\left[a_{11} \left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{13}} \right) \right]^2 + (a_{31}L_3)^2}{\left[a_{11} \left(\frac{a_{13}L_1 + a_{23}L_2}{3a_{13}} \right) + a_{31}L_3 \right]^2} < 1 \\ A_2^H &= \sum_j \left(\frac{Y_{j2}}{\sum_j Y_{j2}} \right)^2 = 1 \\ A_3^H &= \sum_j \left(\frac{Y_{j3}}{\sum_j Y_{j3}} \right)^2 = \frac{\left(a_{13} \left(\frac{2L_1}{3} - \frac{a_{23}L_2}{3a_{13}} \right) \right)^2 + \left(a_{23} \left(\frac{2L_2}{3} - \frac{a_{13}L_1}{3a_{23}} \right) \right)^2}{\left[a_{13} \left(\frac{2L_1}{3} - \frac{a_{23}L_2}{3a_{13}} \right) + a_{23} \left(\frac{2L_2}{3} - \frac{a_{13}L_1}{3a_{23}} \right) \right]^2} < 1 \\ \text{Thus, } A^H &= \sum_s \frac{\sum_j Y_{js}}{\sum_s \sum_j Y_{js}} A_s^H < 1 \end{aligned}$$

Two states produce good 1 and two states produce good 3, while only one state produces good 2, so agglomeration is unambiguously weaker than in the generic case. By a similar reasoning, dissimilarity is lower than the maximum since two goods are produced in more than one region. The degree of dissimilarity between regions is :

$$\begin{aligned} D_1 &= \frac{2}{J(J-1)} \sum_{j<k} \left| \frac{Y_{js}}{\sum_s Y_{js}} - \frac{Y_{ks}}{\sum_s Y_{ks}} \right| = \frac{2}{3} \\ D_2 &= \frac{2}{3} - \frac{2a_{23} \left(2a_{23}L_2 - a_{13}L_1 \right)}{3a_{22} \left(a_{13}L_1 + a_{23}L_2 \right) + 3a_{23} \left(2a_{23}L_2 - a_{13}L_1 \right)} < \frac{2}{3} \\ \text{if } s_{13} &> s_{23} : D_3 = \frac{2}{3} - \frac{2a_{11} \left(a_{13}L_1 + a_{23}L_2 \right)}{3a_{13} \left(2a_{13}L_1 - a_{23}L_2 \right) + 3a_{11} \left(a_{13}L_1 + a_{23}L_2 \right)} < \frac{2}{3} \\ \text{if } s_{13} &< s_{23} : D_3 = \frac{2}{3} - \frac{2a_{22} \left(a_{13}L_1 + a_{23}L_2 \right)}{3a_{23} \left(2a_{23}L_2 - a_{13}L_1 \right) + 3a_{22} \left(a_{13}L_1 + a_{23}L_2 \right)} < \frac{2}{3} \\ \text{So } D &= \frac{1}{S} \sum_s D_s < \frac{2}{3} \end{aligned}$$

Can do similarly for each case, but the bottom line is that some regions produce more than one good, so at time 1 agglomeration is lower and dissimilarity lower also, while diversification can change in either direction.

Compared to the full specialization equilibrium of the baseline model, where we had perfect agglomeration, maximal interregional dissimilarity and country-wide diversification, here we still have countrywide diversification (the change in the extent of diversification is ambiguous), but less agglomeration and less interregional dissimilarity across all 5 cases (really 10 cases, depending on how we treat region 3). That is because a region produces at least 1 sector as opposed to exactly one sector. In many cases a region can produce two or more sectors.

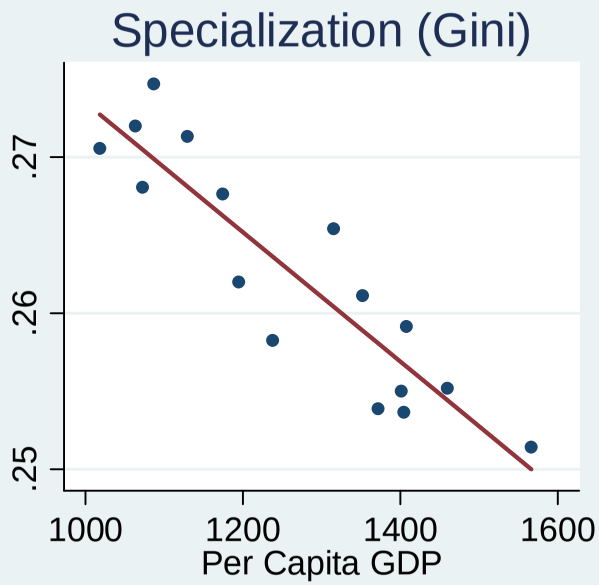
Time 2

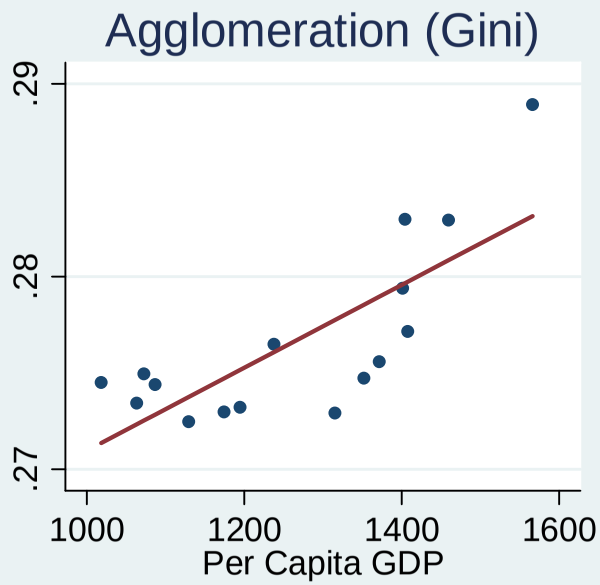
Two assumptions can be made: a simple one and a complicated one. The simple one is that closed regions of each country have finally opened up at time 2. Then, the focal outcome is the same as that in the generic 3×3 case: full specialization at the country level. Considering a closed region simply affects time 1 analysis, namely the degree of agglomeration and dissimilarity are lower when there are closed regions (the nonmonotonicities is weakened when allowing for closed regions).

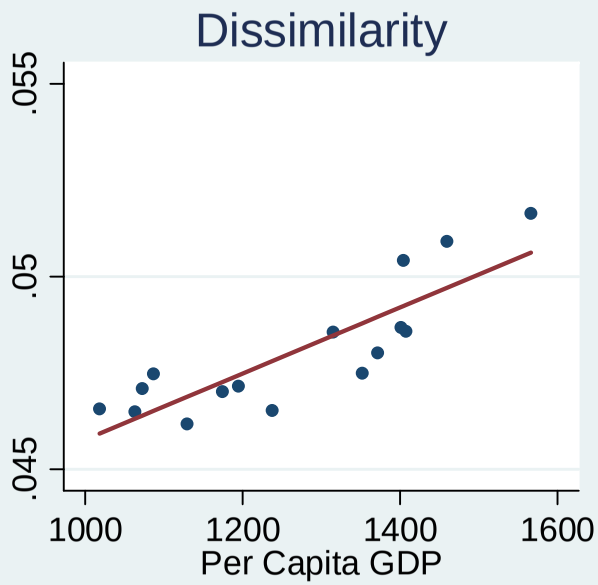
The more complicated assumption is the (perhaps less realistic - there aren't really closed regions in advanced economies) assumption that once a region is enclaved, it stays enclaved. So, while the two open regions integrate with each other and the rest of the world, the third (closed) region of each country remains closed. All our previous assumptions on labor mobility, international integration and productivity convergence hold for the two regions that are open in each country - the closed regions are taken out of the picture (but their sectoral composition has to be included for the calculation of the indices). We again have a 3 country model with 3 goods, and the analysis will be similar to that in the generic case, but the trading units are smaller (in each country k, a mass of population L_3^k is in regions that remain autakic, perfectly specialized in producing good 1). Assuming again that country k produces good k, for all $k = 1, 2, 3$, and focusing again on the class of assignments of one country to one good, it is easy to see that predictions on the dynamics of structural change in this case depend on whether the closed region remains specialized in the sector for which the country overall hapens to have an international comparative advantage. If this is the case, nothing changes compared to the generic case. If not, the country is predicted to remain more diversified than in the generic case, and more agglomerated: open regions produce a

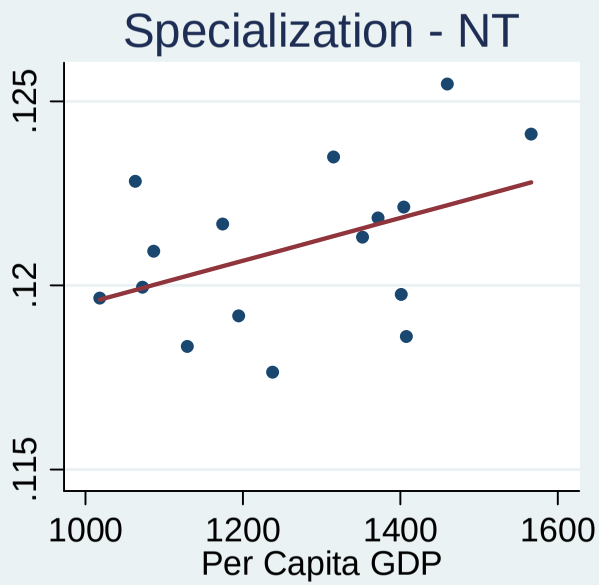
single good according to the cloutry's pattern of international comparative advantage, while the closed region produces a different good. In other words, allowing region 3 to remain closed at time 2 again weakens the dynamics of change in specialization, agglomeration and sectoral dissimilarity, this time at time 2 as well.

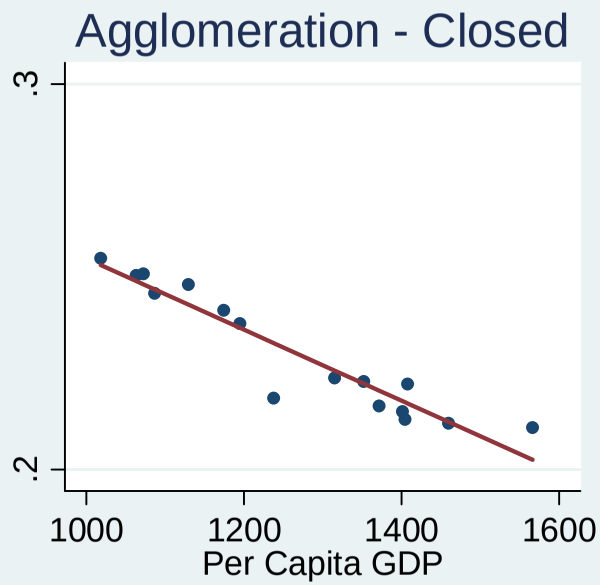
Figure 1A: India











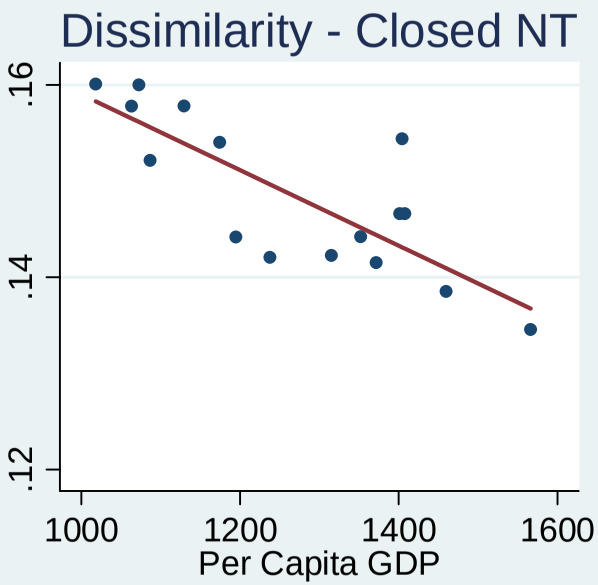
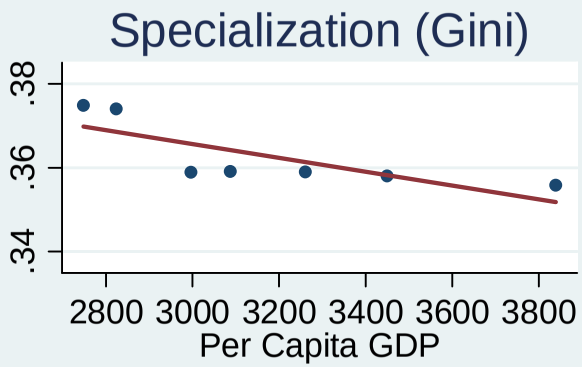
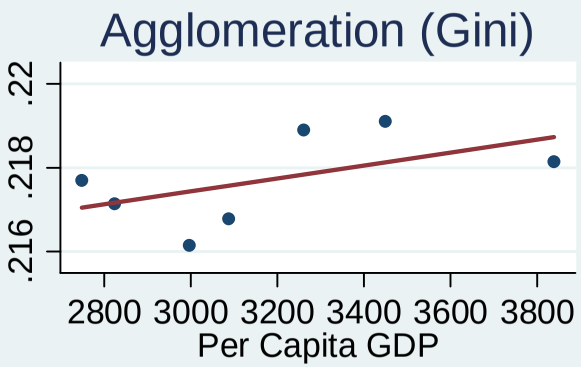
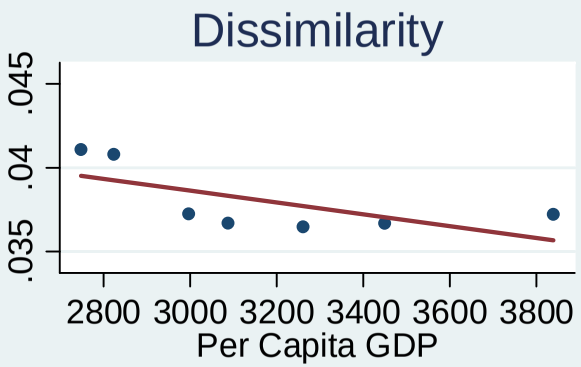
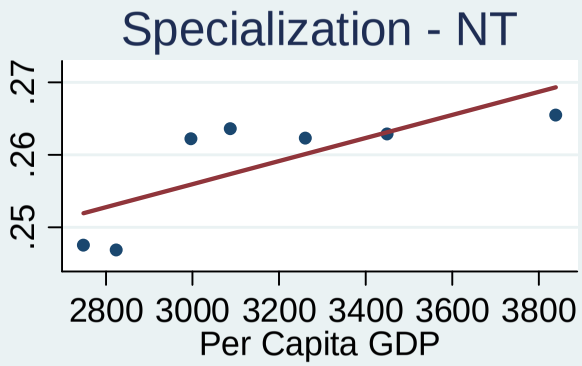


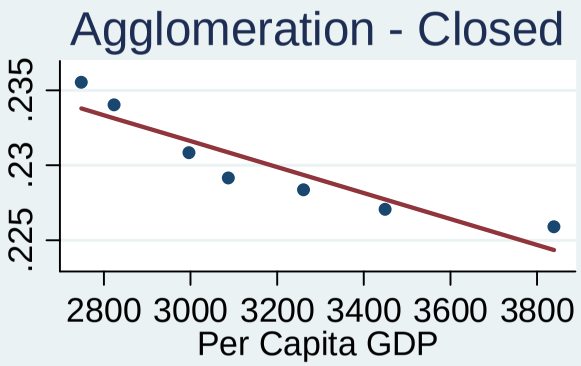
Figure 1B: China

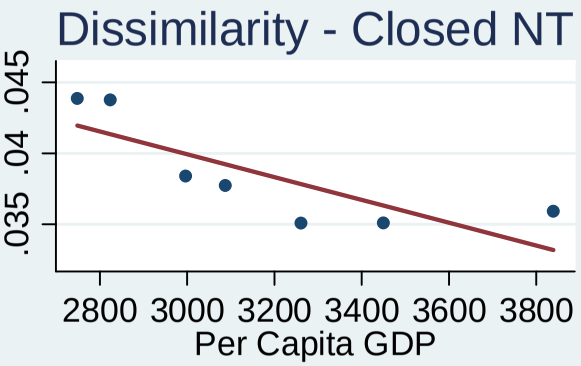


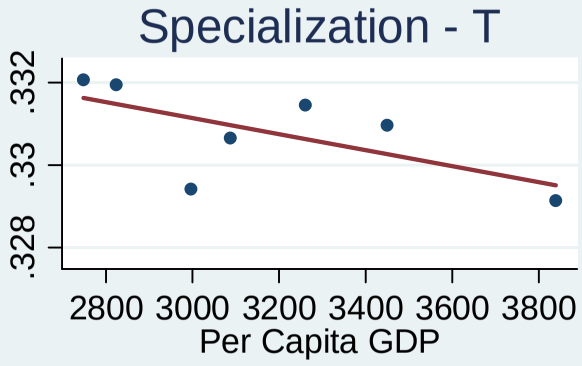


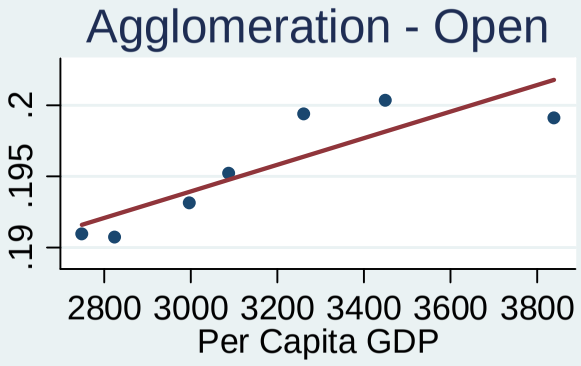












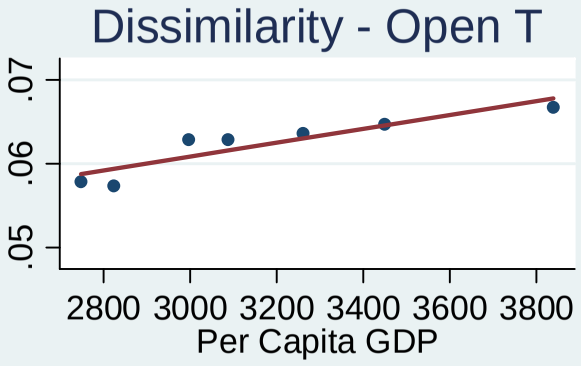
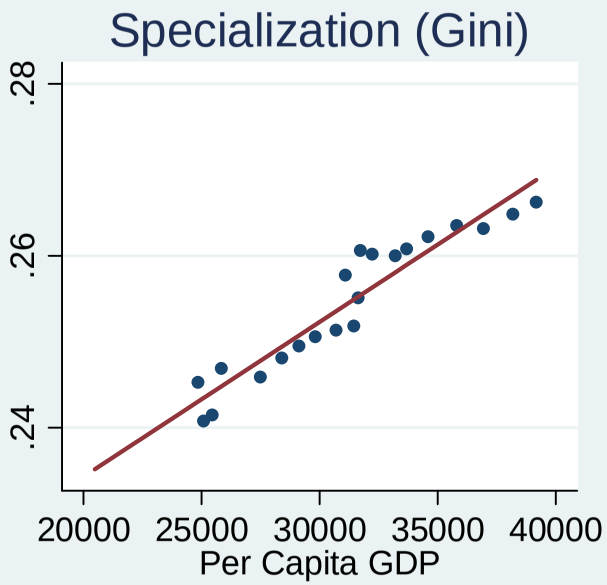
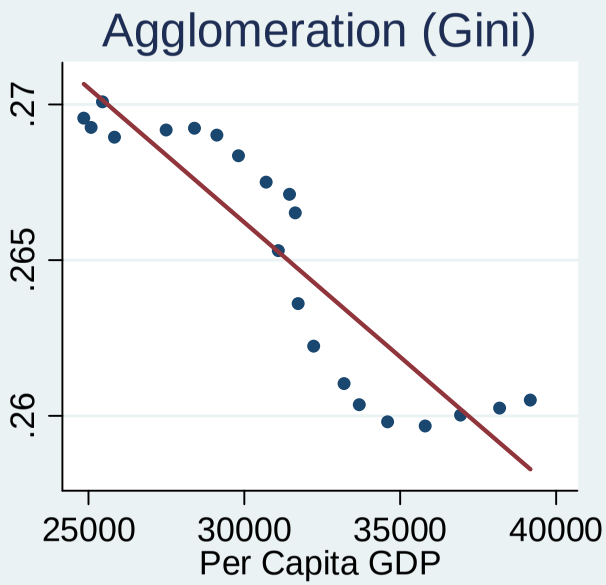
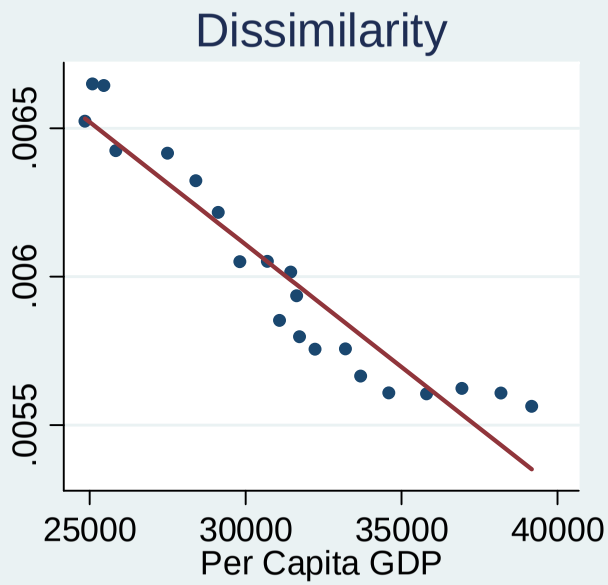
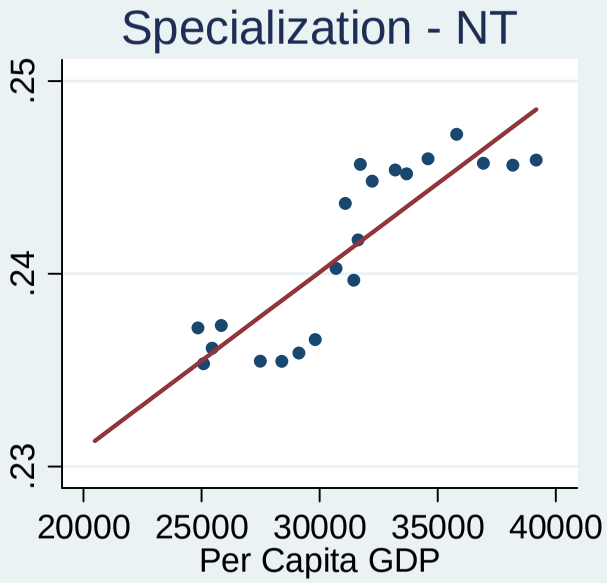


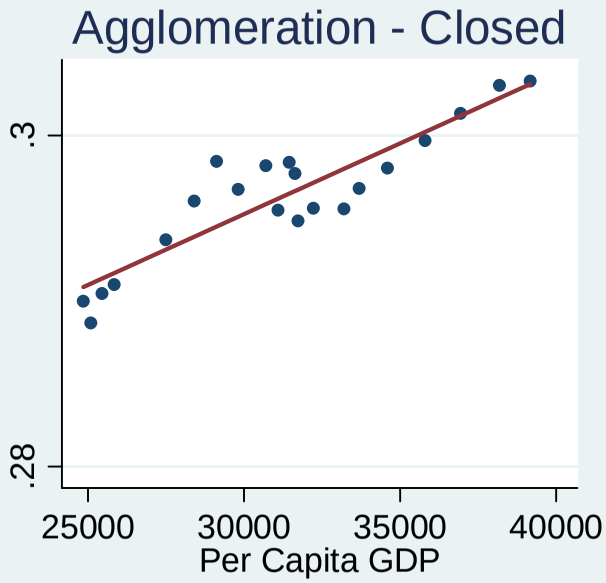
Figure 2: USA


Specialization (Gini)
.24
.26
.28
20000
25000
30000
35000
40000
Per Capita GDP


Agglomeration (Gini)
.26
.265
.27
25000
30000
35000
40000
Per Capita GDP


Dissimilarity
.0055
.006
.0065
25000
30000
35000
40000
Per Capita GDP


Specialization - NT
.23
.24
.25
20000
25000
30000
35000
40000
Per Capita GDP


Agglomeration - Closed
.28
.3
25000
30000
35000
40000
Per Capita GDP

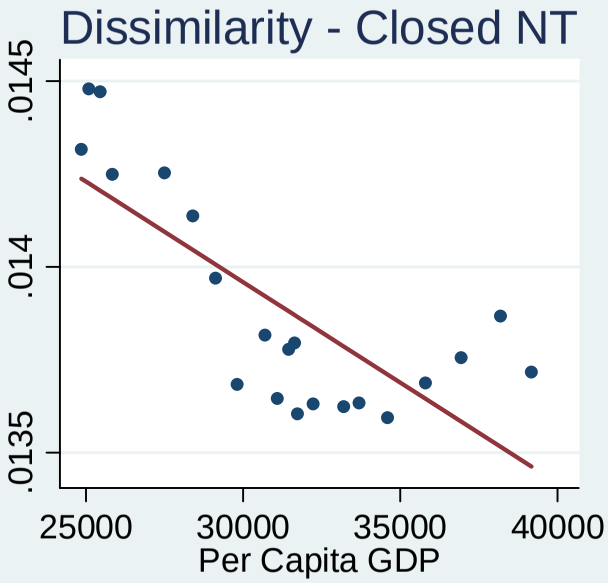

Dissimilarity - Closed NT
.0135
.014
.0145
25000
30000
35000
40000
Per Capita GDP

Figure 3: Eurostat

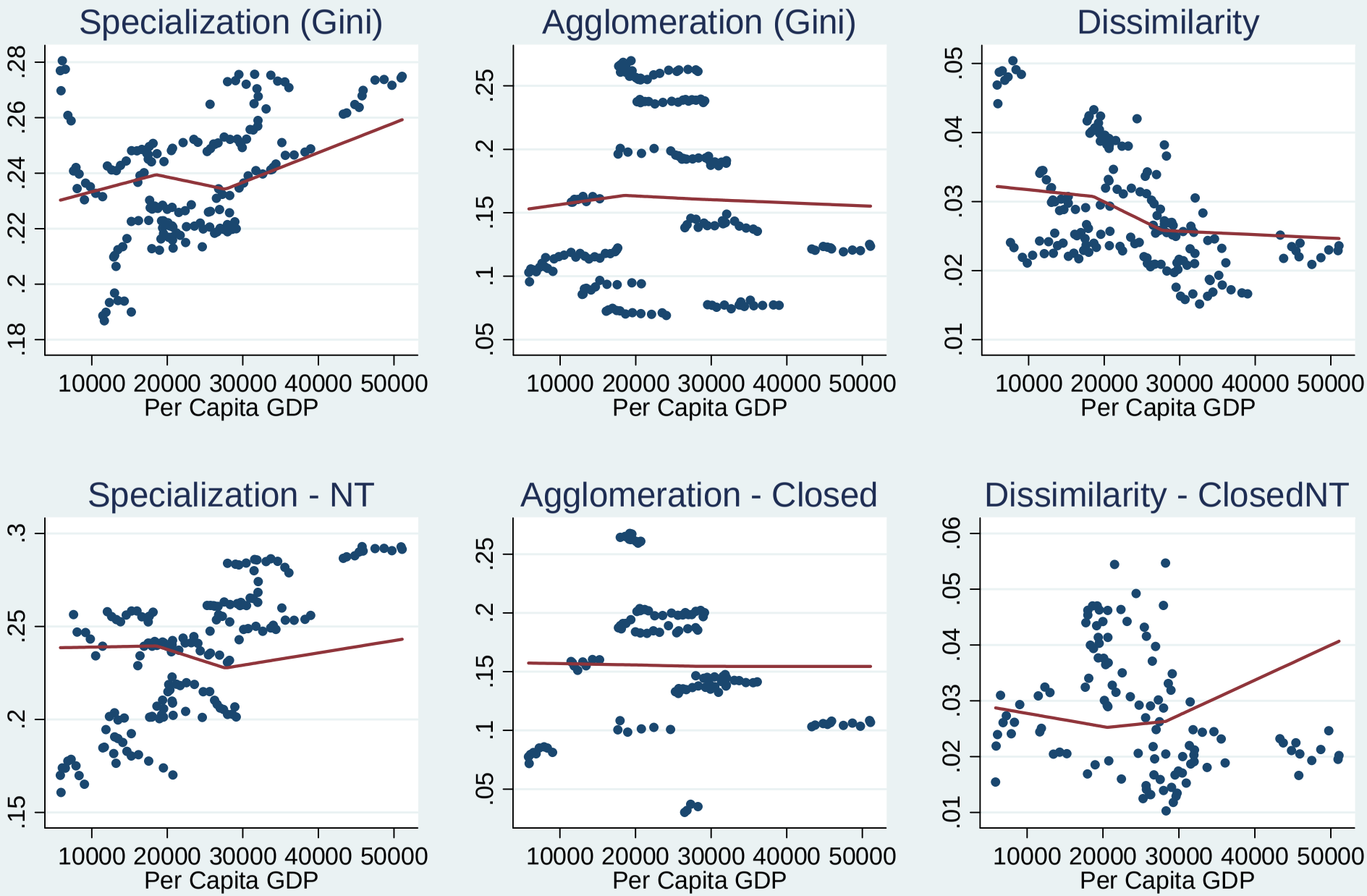


Figure 4: ILO

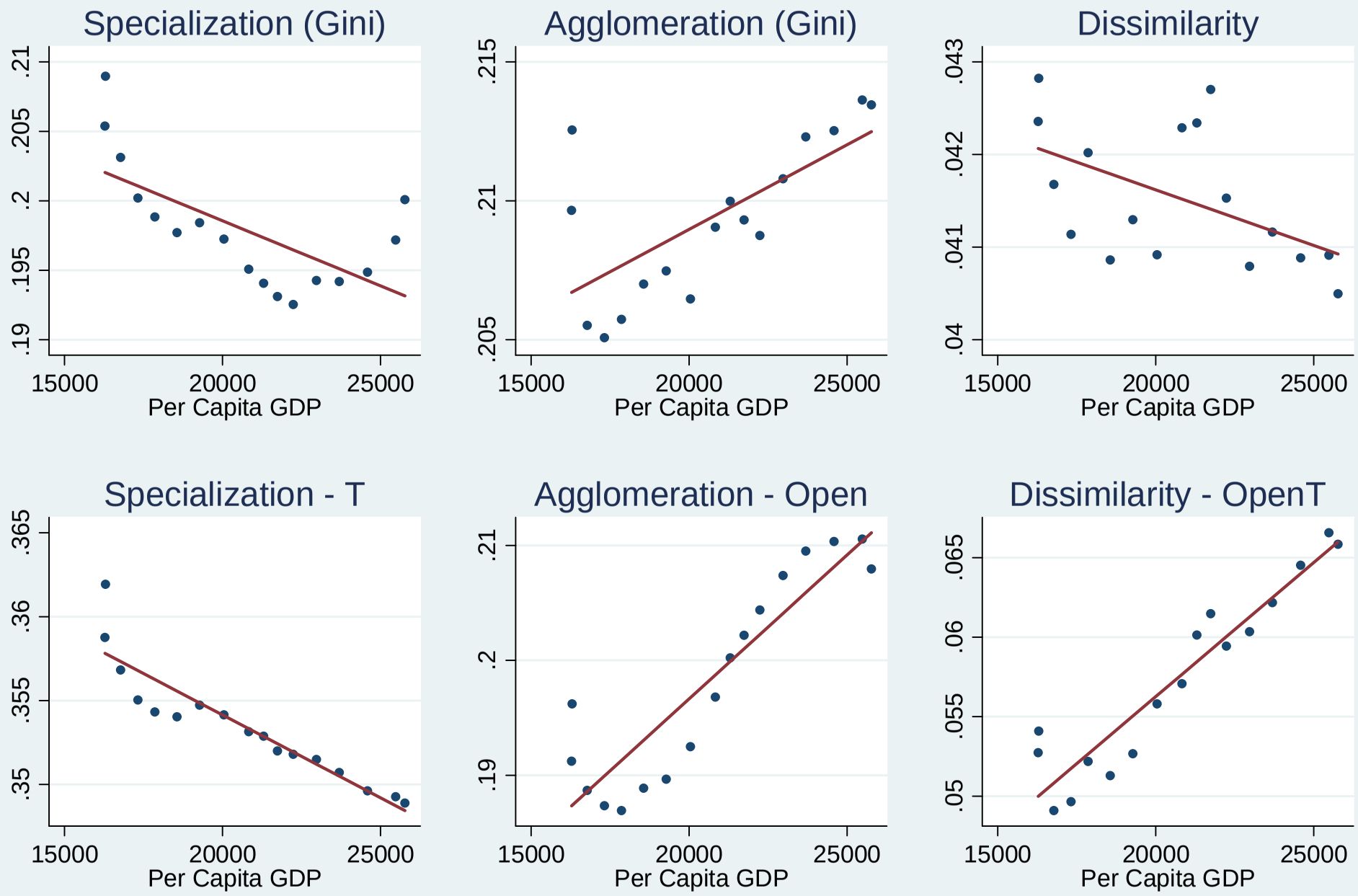

Specialization (Gini)
.19
.195
.2
.205
.21
15000
20000
25000
Per Capita GDP
Agglomeration (Gini)
.205
.21
.215
15000
20000
25000
Per Capita GDP
Dissimilarity
.04
.041
.042
.043
15000
20000
25000
Per Capita GDP
Specialization - T
.35
.355
.36
.365
15000
20000
25000
Per Capita GDP
Agglomeration - Open
.19
.2
.21
15000
20000
25000
Per Capita GDP
Dissimilarity - OpenT
.05
.055
.06
.065
15000
20000
25000
Per Capita GDP

Figure 5: IPUMS