Economic Integration and Structural Change*

Jean Imbs
Paris School of Economics
and CEPR

Claudio Montenegro
The World Bank

Romain Wacziarg
UCLA, NBER and CEPR

June 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. 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 University of Vienna, 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, reflects 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 tend 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. Only traded goods are expected to display these patterns.

Our claims are established in three steps. First, we examine the laboratory of European inte-

[1]Obviously, the gradual processes of regional and international integration partially overlap. Our story relies on domestic integration being the dominant force in early stages, and international integration the dominant force 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.

gration. The successive phases of European economic integration have created a wider and deeper area of free trade between countries. Each member country has also undergone domestic integration, before or during accession. Regional data on sectoral employment from EUROSTAT suggest European countries taken individually tend to specialize, and their constituent regions produce a converging set of goods. Thus, European countries behave like the advanced economies of our model. They are, indeed, relatively developed and domestically well-integrated countries. Our conjecture is that the European dynamics of specialization and agglomeration are largely a manifestation of European integration. As they enter the European Union (EU), and as the degree of integration within the Union deepens, countries increasingly specialize according to their comparative advantage within the Union. Since they are domestically integrated, their constituent regions produce goods in which the country overall has an EU-wide comparative advantage, and they become structurally more similar.

The European setting offers an additional, unique test of our story. If our hypothesis is correct, as long as integration among EU countries is a stronger force than the integration of the EU with the rest of the World, the EU *as a whole* should diversify. We should see activity agglomerating at the country-level, and EU countries specializing 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*. To examine whether this holds in the data, we construct an EU-wide aggregate. Unlike the previous exercises, this can be done with sectoral data at the country level, which we obtain from the International Labor Office (ILO). Countries are now effectively treated as the regions of a wider economically integrated zone - the EU. The data show that the EU diversifies, activity agglomerates at the country level, and countries grow dissimilar. This is particularly true for traded sectors, exactly as predicted by our story. The European data, both at the level of individual countries' regions and at the level of the EU as a whole, offer strong evidence for an integration-driven theory of structural change.

Second, we focus on three individual countries, i.e. case studies, 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. Remarkably, sectoral diversification cannot be detected amongst non-traded sectors. Moreover, geographic agglomeration is especially prevalent in traded activities. A similar pattern is observed in China.

The opposite patterns are observed in the United States. There, 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 dominance of international trade in the US, relative to economic integration between States, which is complete in our sample. Because of this very integration, little is available in the US by way of regions that remain unexposed to (intra- or inter-national) trade. The patterns of specialization and geographic homogeneization are therefore predicted to hold for most sectors there.

The third and final step in our empirical analysis is to seek more systematic cross-national data where the different dynamics of structural change can be explored in a sample that includes both developing and advanced economies. We consider two samples for which regional data at sectoral level are available. The first, based on census data, includes countries at various stages of development, for instance Vietnam, China, Bolivia, Switzerland or the US - but with short 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 second, taken from national statistical agencies, reports information on sector-level real value added for 14 economies. The time coverage is longer, but countries are mostly from the developed world. In both panels it is possible to focus on the within-country dynamics of specialization, agglomeration and dissimilarity, thanks to country-specific intercepts. Both datasets reproduce a U-shaped relation between sectoral specialization and per capita GDP, as documented in Imbs and Wacziarg (2003). This pattern is especially pronounced 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 slope (as countries specialize because regions grow similar). The same holds true for an index of regional dissimilarity. These dynamics prevail in traded activities, but do not hold when limiting the analysis to to non-traded sectors.

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).[3] 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 at the economywide level.

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 contributions, no variation exists in the extent of economic integration across countries, i.e. the interaction between trade and development is analyzed under an 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.

[3]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 focuses on European evidence. Section 5 presents results for India, China and the US. Section 6 explores international panel 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 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: 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. We then evaluate empirically the predictions that result from these extensions to our basic model.

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. 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 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 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, 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 the 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.

The countrywide 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 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:

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

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 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 efficient, 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.[4] 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:

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

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

[4] 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:

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

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. Three things are assumed to 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.[5] 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):

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

(similarly in countries B and C).

The rest of the equilibrium 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

[5] 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.

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 subsequent sections.

2.6 Extension 1: A nontradable good

We now assume that good 3 remains 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 in the presence of 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.[6] 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.[7] The analysis for time 0 is again unchanged. At time 1, region 3 continues to

[6] 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.

[7] 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

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: 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.

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 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.

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

3 Measurement and Data

The three indices central to this paper are first introduced. Data sources and empirical specifications are then described.

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.[8] S_{it}^{H} can be readily obtained from sectoral data, i.e. no regional decomposition is necessary. It is the same measure as the one 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 activity. The Herfindahl index captures the regional lumpiness of activity in sector s across the regions j that constitute country i.[9]

[8] 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. Results using the Gini index were very similar to those using the Herfindahl index. The empirical results are only presented using the latter.

[9] Analogously, define the Gini measure of regional agglomeration:

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

This Herfindahl requires a sectoral breakdown of economic activity at the sub-national level. It measures the sector-specific geographic agglomeration of activity, and must 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:

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

A high value indicates 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}.[10]

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.

---

where CRA denotes cumulated regional shares $Y_{ijst}/\left(\sum_j Y_{ijst}\right)$. Again, results using the Gini index were similar to those using the Herfindahl index.

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

All three 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. At the one-digit sector classification level, it is 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 splits, the paper computes versions of all three indices over sub-samples that reflect varying degrees of tradability. For the specialization index S_{it}^{H}, which focuses on the sectoral dimension of the data, we report values that correspond to traded or non-traded industries. For the agglomeration index A_{it}^{H}, and dissimilarity index D_{it}, which are all computed from the regional dimension of the data, we also report values when splitting the data between tradable and non-tradable sectors. As the theory made clear, the dynamics we identified above should hold more strongly for indices computed over tradable sectors only.

3.2 Data

A major challenge of this paper is to find data on sectoral composition at the subnational level, across time, for as many countries as possible. To our knowledge, a comprehensive attempt to gather such data in a cross-national panel was never undertaken. The task can be particularly challenging for developing countries, and for countries that do not have a federal structure. It is somewhat easier for advanced economies and/or countries or zones that have a federal or quasi-federal structure (including Europe).

Additional difficulties arise from the need to compute indices that are comparable across time. The data has four dimensions: time, sector, countries and regions. Given the sectoral and regional indices computed in this paper, it is important that the data be balanced across all dimensions, within each country considered (on the other hand, country fixed effects can be used to soak up differences in sectoral nomenclatures and the definition of regional units across countries). Changes in the regional or sectoral coverage across time lead to unstable estimates of indices of agglomeration or specialization, reflecting mere changes in data coverage. Thus, sectors or regions were excluded to ensure that the number of regions and sectors remained constant over time within each country.

This required taking a stance in cases a given region was missing some sectoral information: a choice had to be made in terms of eliminating the sector or the region altogether. The goal of preserving as much data as possible guided these decisions. The rule we followed was that if dropping a sector resulted in fewer lost observations than dropping a region, we dropped the sector (and vice versa).[11] There is no reason to believe this would lead to anything else than classical measurement error in the indices of specialization, agglomeration and dissimilarity, i.e. raising the standard error of our regressions.

The three main 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.

For Europe, we use two datasets. The first, from EUROSTAT, features yearly data on regional employment at the one-digit level of disaggregation for 14 countries over the period 1992-2008 (the data is gathered starting at the date of accession into the EU, so for some more recent entrants the timespan is shorter).[12] This dataset allows us to examine the dynamics of specialization, agglomeration and dissimilarity for each EU country in isolation. The second European dataset, from the ILO, also features yearly data on regional employment at the one digit level, for the same 14 countries over the same period (here, the accession date to the EU is irrelevant to determine the timespan of available data). This is the dataset we use to examine the sectoral and regional allocation dynamics for the EU as a whole. Each country is treated as the analog of a "region" and we can examine the EU wide pattern of specialization, agglomeration and dissimilarity, without requiring sub-national data. Analogously to the country-level data, the number of sectors must be homogeneous across the countries constituting the area, and countries must be available continuously across time. Thankfully, international data on sectoral activity are more naturally balanced

[11] In some cases where fewer than 4 consecutive observations were missing, the data were interpolated linearly, rather than dismissing the whole sector or region.

[12] The countries are: Austria, Belgium, Czech Republic, Spain, France, Greece, Hungary, Ireland, Norway, Poland, Portugal, Romania, Sweden, and Slovakia.

than sub-national data, especially for European economies. However, a few sectors or countries were still omitted to ensure that the coverage of countries and sectors was constant across time.

Case studies evidence is based on subnational sectoral data for three individual countries: The US, India and China. US data were obtained from the Bureau of Economic Analysis, with 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). The dataset is available from 1969 to 2000, but many observations are missing prior to 1980 so we focus on the 1980-2000 time period. The requirement of consistent sectoral coverage over time means we are left with 63 sectors after dropping those for which not all years are observed in all states.

For India, real state domestic product by sector is available from National Accounts sources, for the 28 States and Union Territories, observed for 13 sectors, i.e. slightly more disaggregated than the 1-digit level of disaggregation. The time coverage is 1980-1995. While longer series are available for India, a change in the base year for sectoral relative prices in 1980 leads to a discrete change in some sector's shares in that year compared to preceding decades, so we restrict attention to data for which real sectoral output correspond to a unique base year.

For China, we obtained yearly data on employment at the provincial level, covering 12 sectors from 1995 to 2009. Thus, we focus on 1995-2002. The primary sector (Farming, Forestry, Animal Husbandry and Fishery) displays suspiciously high volatility, with year-on-year changes of up to 600%, and an *average* of 39%. This is more than ten times more volatile than other sectors in China. While some large changes in employment in the primary sector in China are possible, such cataclysmic changes are implausible. Herfindahl indices are notoriously sensitive to measurement error. Given the large share of agriculture in some Chinese regions, measurement error is likely to carry over to the aggregate. The primary sector in China is therefore omitted.[13] Unfortunately, from 2003 onwards, the data only covers urban employment, whereas it includes both rural and urban employment from 1995 to 2002. This creates an unavoidable break in 2002, and makes comparisons difficult between the two periods, not least because urban employment reflects only part of economic activity. Thus, we focus on 1995-2002.

Our third source of evidence is international data from the Integrated Public Use Microdata Series (IPUMS) project, which harmonizes (decennial) census data internationally between 1960

[13] But in panel analyses all available sectors are kept. The censoring is only performed for the single country case study on China.

and 2007. Data on sectoral employment at the regional level are available for 28 countries, of which 9 are developed.[14] This dataset allows us to examine, in a sample with many poor countries, the dynamics of specialization, agglomeration and dissimilarity in a more systematic fashion, including country fixed effects. Regional data at initial stages of development are the main benefit of using IPUMS data, but this comes at the expense of the time series dimension, with an average of 3.7 years available for each region, and 3 years for developing countries. In other words, the time coverage here is more sparse than for some of our other datasets, but the geographic coverage is broader and developing countries are well-represented.

The fourth and final sample is constructed from data released by national statistical agencies. Its main appeal is the use of a measure of sector-level, regional activity based on real value added, rather than employment. If sector-level productivities diverge, specialization or agglomeration can prevail in employment, but not in production. Output data provides a desirable check on the paper's main results. We have gathered regional data on sector-level real value added for 14 countries, mostly focused on developed economies.[15] The second appeal of these data is their time coverage, much more substantial than census data, with an average of 15 years per country. Within-country dynamics are easier to identify in such a panel.

3.3 Specification

For each dataset, we examine the evolution of our three main indices against per capita GDP, a measure of the level of development. The goal is descriptive: we seek to examine whether the predictions of our model at various stages of development hold in the data. We present plots of the indices against per capita GDP, and we run the simple corresponding regressions to establish statistical significance. For single-country datasets (India, China, US and EU-ILO), the specifications

[14] The countries are: Argentina, Austria, Bolivia, Brazil, Canada, Chile, China, Colombia, Costa Rica, Ecuador, France, Greece, India, Malaysia, Mali, Mexico, Panama, Peru, Portugal, Romania, South Africa, Spain, Switzerland, Thailand, United Kingdom, United States, Venezuela, and Vietnam.

[15] The countries are: Australia, Austria, Canada, Chile, Colombia, Denmark, India, Indonesia, Japan, New Zealand, Peru, Portugal, Slovenia, and the United Kingdom. The sources are detailed in the Appendix.

are of the form:

$$\begin{aligned} S_t^H &= \alpha_1 + \beta_1 YPC_t + \varepsilon_t^1 \\ A_t^H &= \alpha_2 + \beta_2 YPC_t + \varepsilon_t^2 \\ D_t &= \alpha_3 + \beta_3 YPC_t + \varepsilon_t^3 \end{aligned}$$

where YPC_t is per capita income at time t. For the panel datasets (EUROSTAT, IPUMS, and Real Value Added datasets), the specifications are as follows:

$$\begin{aligned} S_{it}^H &= \delta_1 YPC_{it} + \nu_i^1 + \eta_{it}^1 \\ A_{it}^H &= \delta_2 YPC_{it} + \nu_i^2 + \eta_{it}^2 \\ D_{it} &= \delta_3 YPC_{it} + \nu_i^3 + \eta_{it}^3 \end{aligned}$$

where ν_i^1, ν_i^2 and ν_i^3 are country fixed effects. These regressions are estimated separately for samples of low and high income countries, in order to capture the potentially different dynamics of sectoral change in the countries composing each sample. Country-specific intercepts are also specific to each sub-sample, to account for the fact that the average level of specialization (agglomeration, dissimilarity) can depend on the countries included in the sample considered.[16]

Finally, we consider sample splits by tradability. We recomputed S_{it}^H, A_{it}^H and D_{it} over sectors that are commonly considered to be traded sectors, and separately for non-traded goods sectors. Our theory has different predictions regarding the dynamics of these indices whether nontraded goods are included when computing the indices.

4 Integration Across Regions and Countries of Europe

4.1 Structural change in the EUROSTAT data

We start by characterizing the pattern of structural transformation within countries of Europe, exploiting subnational panel data on sectoral composition. The EUROSTAT data comprises both historical members such as France or Belgium and accession countries such as Bulgaria, Poland or Slovakia, i.e. a wide range of development levels. Figure 1 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

[16] See Imbs and Wacziarg (2003).

(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 1 plots the raw panel data (without removing country means), along with fitted fixed-effects regression lines corresponding to fitted relationships across two subsamples of relatively rich and relatively poor European countries. The subsamples are separated at the median value of per capita GDP. The fitted estimation isolates the within-country variation in the panel.

A U-shaped relationship with income is apparent in S_{it}^H, so that European countries tend to first diversify and later specialize as they grow richer. This pattern corresponds exactly to the U-shaped pattern of specialization documented in Imbs and Wacziarg (2003). Turning to the regression results of Table 1, column 1, we confirm that for the least wealthy among this European sample, the pattern is one of diversification (albeit it is not statistically significant at conventional levels), while for rich countries a more pronounced and highly significant pattern of specialization is apparent. Turning to the dynamics of A_{it}^H, we see that in this sample, agglomeration falls, consistent with the predictions of our model for rich countries (Table 1, column 2). At the same time, for the least wealthy countries in this sample, the pattern of agglomeration is not statistically significant. Finally, 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 again is clearest amongst high income (specializing) countries. Overall, the results are consistent with the main predictions of our story: a relatively rich set of countries should be observed to specialize, deagglomerate regionally and become regionally more homogeneous. Also consistent with our model, this pattern is more pronounced for the richer countries in the sample than for the poorer countries. In Europe, countries integrate mainly with each other and therefore specialize according to their comparative advantage. Since it is international comparative advantage (especially within the EU) that determines what member countries produce, regions become more homogeneous and activity dis-agglomerates.

4.2 Structural change in the ILO Data

An additional test may help evaluate this paper's major conjecture that economic integration drives observed patterns of specialization, agglomeration and inter-regional similarity. Our 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 of what each country goes through in isolation. Moreover, economic activity in this European aggregate should agglomerate at the country level, and 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, using ILO data. Indices are now computed on Europe overall, for the 17 years of available data

Figure 2 reports the dynamics of S_t^H, A_t^H and D_t for Europe as a whole, against aggregate European per capita GDP, with corresponding regressions appearing in Table 2. The first column plots the measures computed over the full sample of sectors and countries. It is clear that Europe as a whole diversifies, as its member countries actually specialize in different sectors. Table 2 reports the actual estimated slope of the relation, which is negative and significant at 1% confidence level. To summarize, European countries specialize, but in different activities, so that Europe overall diversifies.

The patterns for A_t^H and D_t in Europe as a whole are less clear. Both coefficients on per capita income are negative, and statistically significant for D_t. The second and third columns of Figure 2 demonstrate that this is due in part to non-traded sectors, and in part to agriculture. The second column of Figure 2 focuses on traded sectors: A_t^H and D_t continue to display weakly negative slopes.[17] But both feature a clear break at the middle of the sample, corresponding to the effective entry into the Union of the Czech Republic, Hungary, Poland, and Slovakia. Both A_t^H and D_t display positive trends prior to this event. The third colum in Figure 2 reveals the break pertains to agriculture. When omitting the primary sector from the sample of traded sectors, the EU as a whole is clearly and significantly agglomerating and its constituent countries become increasingly dissimilar: the trends in A_t^H and D_t are now positive and significant, consistent with the diversification apparent from column 1 of the Figure.

When focusing only on traded sectors (but not agriculture), European countries do become structurally more dissimilar from each other, as they specialize in different goods and activity agglomerates at country level. In fact, the fourth column of Figure 2 shows that, for nontraded goods in isolation, if anything Europe overall is specializing and its constituent countries are becoming similar. The patterns predicted by our theory hold mostly for the subsample of traded sectors, exactly as expected.

[17] Traded sectors are defined as Agriculture, Mining, Quarrying, Manufacturing, Transport, Storage, Communication, Finance, Insurance, Real Estate and Business Services.

Why would agriculture obscure European-wide phenomena of sectoral diversification accompanied by geographic agglomeration and national divergence? First, the Common Agricultural Policy create distortions in the production of primary goods across member countries, which is liable to obscure the consequences of trade in primary goods within the Union. Second, it is possible that employment in agriculture actually falls everywhere as new countries enter the Union, thanks to productivity gains.[18] In countries like Poland or Hungary, rising productivity frees up labor, that moves from primary activities to manufactures or services. New entering countries become similar to older, industrial, member countries: D_t falls just as new countries enter the Union. But agricultural employment also falls in large, incumbent members like France, because some of the production is now located further East: Agglomeration A_t^H also falls just as new countries enter the Union. This possibility emphasizes the importance of using production, as well as employment data, to account for changes in productivity. This is an important issue we will come back to.

To summarize, Europe provides an ideal laboratory to empirically assess this paper's conjectures. 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 with the rest of Europe. 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 (in turn, when the process of European integration will be more advanced, we should expect Europe to start specializing *vis-à-vis* the rest of the world). These patterns are strongly consistent with the idea that a gradual process of local and then global economic integration can account for the observed dynamic patterns of specialization, agglomeration and structural similarity.

[18] In fact, the CAP was heavily reformed in 2005, precisely as new entrants came into the Union.

5 Three Case Studies

This section presents results for India, China and the US. These case studies offer tests of our model for two developing countries and one advanced country, jointly representing a large share of the world economy and population.

Figure 3 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, for the 16 years of available data. The top row displays the path of indices computed on all available regions and sectors. India overall is diversifying, and this is accompanied with economic activity agglomerating across States and Union Territories. These regions also become increasingly dissimilar in terms of their production structure. All three patterns are statistically significant (Table 4). These three facts are highly consistent with our model.

The following rows use measures of sectoral change computed over subsamples of traded and nontraded sectors, respectively (non-traded sectors are defined as Commerce, Construction, Public Administration, Real Estate, Trade, Hotels, Restaurants and Other Services). Patterns found for the economy overall are more pronounced and more significant for traded sectors, as we expect. They are less pronounced (and in the case of specialization, reversed) when isolating nontraded sectors.

Figure 4 plots analogous graphs for China. 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.

It is possible that the latter pattern is driven by the presence of nontraded sectors in the sample. To investigate this, the second and third rows for Figure 4 plot the dynamics of the three indices, computed over sub-samples with tradable and nontradable sectors, respectively.[19] Indeed, dissimilarity is rising between regions of China when focusing on traded sectors, and Table 4 shows that this finding is statistically significant, despite the paucity of observations (only 7 years of available

[19] Traded sectors are defined as Mining, Manufacturing, Transport, Storage, Post, Information Transmission, Computer Service, Software, and Financial Intermediation. Agriculture is omitted from the sample altogether, because it displays anomalous volatility relative to the other sectors.

data). We see again that the pattern of diversification, agglomeration and growing dissimilarity exists strongly among traded sectors, while nontraded sectors feature neither diversification nor growing dissimilarity.

The final case study concerns an advanced industrialized country, the United States. Figure 5 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, and Table 5 mirrors these figures with regressions. The US specializes, in the sense that employment concentrates in few sectors, and the Herfindahl indexcomputed across sectors increases with per capita GDP. Activity dis-agglomerates, and US States produce an increasingly similar set of indutries. Arguably, 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, and 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: 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. In other words, he US - and its constituent States - are specializing according to their global comparative advantage.

6 Global Integration

To generalize the results found so far, we turn to the two datasets with a panel dimension. Since we observe several countries at various levels of development, we revert to the practice of separating our sample between low and high income countries, and running separate regressions on each subsample to evaluate the predictions of the model at different stages of development. The census data collected by IPUMS has short time coverage, so the subsamples are once again separated at the median value of per capita GDP. Instead, the panel with output data has an average of 15 years of data, ranging between 1955 and 2007. Since low income developing countries represent less than half the observations there, the range of per capita GDP is split three ways.

IPUMS census data focus on developing economies, with two thirds of the observations on per capita GDP below about $\$15,000$ at 2005 PPP exchange rates, and very few observations above $\$25,000$. Hence, we expect the data to mostly reflect patterns prevalent during the first stage of diversification, among poorer countries. This is indeed what we find. The upper row in Figure 6 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 fixed effects regressions for the subsamples of low and high income countries, as previously defined. Corresponding regressions appear in the top row of Table 6. We find that diversification is pronounced at values of per capita GDP below $\$10,000$, and in that range activities agglomerate regionally. At higher income levels, these trends tend to flatten out, consistent with the model. Dissimilarity, however, does not display any significant trend at low income levels, and falls in rich countries.

The second rows of Figure 6 and of Table 6 examine the time path of measures computed over samples of traded sectors (defined here as Agriculture, Mining, Manufacturing, Transport, Telecommunication, Finance and Business Services). The dynamics are sharper than when including all sectors. In particular, the upward trend in agglomeration for poor countries is larger and still statistically significant, and the extent of diversification at early stages becomes larger. Dissimilarity displays a striking and significant hump shaped pattern when focusing on tradable goods only. The lowest panels of Figure 6 and Table 6 show the dynamics are the reverse in a sample focused on non-traded sectors.

Table 7 conduct a similar exercise on the sample of 14 countries with real value added data. Figure 7 provides three illustrative examples taken from this source: a developing country, Colombia, that actually goes through the stages of diversification in sample. And two developed countries with especially long coverage, Japan and the United Kingdom. The Figure focuses on traded sectors, and illustrates the paper's main results. Table 7 explores their robustness in output data in more systematic fashion. Thanks to larger time coverage, per capita GDP ranges from $\$1,000$ to more than $\$40,000$ in this sample. The total of 208 observations is split three ways, rather than using a median value that would obscure phenomena specific to developing economies.[20] The upper panel of Table 7 confirms that diversification prevails in this sample at low and medium levels of per capita GDP. It is accompanied by weakly significant agglomeration for low income countries,

[20] We could go the extra step and reproduce the lowess estimation in Imbs and Wacziarg (2003), running within-country regressions over rolling windows constituted by sub-samples of the data. That is made difficult by the limited country coverage. Regional output data are only available for 14 countries: the window length has to be high to ascertain enough cross-country dispersion is preserved in each sub-sample. The end results are little different from Table 7 and Figure 7. In contrast, we had more than 60 countries when all that was needed was sector-level information.

and an overall convergence in regional production patterns.

The middle panel of Table 7 displays the dynamics implied by traded sectors only (defined here as primary goods, manufactures and financial services). The results are striking: output is going through strong and significant stages of diversification followed by specialization at low and medium per capital GDP levels. The exact opposite dynamics hold for agglomeration and for regional dissimilarity. No such patterns are observable amongst non-traded goods, where agglomeration is insignificant, and regional production patterns converge significantly at all levels of income.[21] Figure 8 reproduce these fitted regressions against the backdrop of raw data.

7 Conclusion

This paper documents a robust pattern of structural change throughout 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 contributes to 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 essential. 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 in the laboratory of European integration, and by contrasting the patterns of specialization, agglomeration and structural similarity between tradable and non tradable sectors. The patterns are usually absent from nontraded activities, and stronger among traded ones. The results are obtained in a variety of datasets, a multiplicity of specialization measures, and a multiplicity of approaches - including individual country case studies, multi-country zones of economic integration such as the EU, where a region is in fact a country, and systematic

[21] This last result is to be expected as non-traded sectors are produced everywhere.

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 results at least in part from domestic economic integration. Irrespective whether diversification is desirable from a welfare standpoint, it can be understood as the result of infrastructure investment, 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, which has 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 center stage, and countries specialize. At that more advanced 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 comprised 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.

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:

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

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:

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

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 :

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

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.

Table 1: Eurostat - Regional Data

	Specialization	Agglomeration	Dissimilarity
Low	-0.987 (-1.40)	-0.173 (-0.61)	-0.265^{***} (-2.76)
High	0.970^{***} (6.07)	-0.229^{**} (-2.39)	-0.175^{**} (-2.18)
Obs.	81	81	81

Notes: The Table reports coefficient estimates in a regression of a sectoral Herfindahl index on real per capita GDP estimated for "low" and "high" income sub-samples. The mid-point corresponds to median per capita GDP. The number of observations refers to each sub-sample. All estimations include country-specific fixed effects. Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels.

Table 2: ILO - Sectoral Data

	Specialization (All)	Agglomeration (All)	Dissimilarity (All)
EU per capita GDP	-1.21^{***} (-4.86)	-0.272 (-1.64)	-0.505^{***} (-3.28)
Obs.	17	17	17
	Specialization (T w/AGR)	Agglomeration (T w/AGR)	Dissimilarity (T w/AGR)
EU per capita GDP	-2.78^{***} (-6.72)	-1.74^{***} (-5.33)	-0.217 (-0.56)
Obs.	17	17	17
	Specialization (T w/oAGR)	Agglomeration (T w/oAGR)	Dissimilarity (T w/oAGR)
EU per capita GDP	-7.60^{***} (-13.11)	0.826^{***} (3.84)	1.38^{***} (9.51)
Obs.	17	17	17
	Specialization (NT)	Agglomeration (NT)	Dissimilarity (NT)
EU per capita GDP	0.901^{***} (3.31)	2.81^{***} (10.79)	-1.81^{***} (-7.37)
Obs.	17	17	17

Notes: The Table reports coefficients estimates in regressions of various indexes on EU-wide per capita GDP. Specialization and agglomeration are measured using Herfindhal indexes. The upper panel computes indexes on the full sample of sectors. The two middle panels focuse on traded sectors (Mining and Quarrying, Manufacturing, Transport, Storage and Communication, and FIRE), to which Agriculture is or is not added. The lower panel focuses on non-traded sectors (Electricity, Gas and Water, Construction, Wholesale and Retail Trade, Restaurants and Hotels, Community, Social and Personal Services). Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels.

Table 3: India

	Specialization (All)	Agglomeration (All)	Dissimilarity (All)
Per capita GDP	-80.10^{***} (-10.34)	19.40^{***} (4.42)	8.57^{***} (6.11)
Obs.	16	16	16
	Specialization (T)	Agglomeration (T)	Dissimilarity (T)
Per capita GDP	-174.10^{***} (-13.98)	25.01^{***} (4.09)	15.70^{***} (5.46)
Obs.	16	16	16
	Specialization (NT)	Agglomeration (NT)	Dissimilarity (NT)
Per capita GDP	10.20^{***} (3.56)	9.83^{***} (5.98)	14.40^{***} (5.39)
Obs.	16	16	16

Notes: The Table reports coefficients estimates in regressions of various indexes on India per capita GDP. Specialization and agglomeration are measured using Herfindhal indexes. The upper panel computes indexes on the full sample of sectors. The middle panel focuses on traded sectors (Agriculture, Forestry and Fishing, Mining and Quarrying, Manufacturing, Transport, Storage and Communication, Banking and Insurance). The lower panel focuses on non-traded sectors (Construction, Trade, Hotels and Restaurants, Real Estate, Business Services, Public Administration and Other Services). Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels.

Table 4: China

	Specialization (All)	Agglomeration (All)	Dissimilarity (All)
Per capita GDP	-68.30^{**} (-2.82)	1.83^{***} (8.85)	-1.08 (-1.01)
Obs.	7	7	7
	Specialization (T)	Agglomeration (T)	Dissimilarity (T)
Per capita GDP	-100.10^{*} (-2.48)	3.07^{***} (9.33)	7.65^{***} (4.54)
Obs.	7	7	7
	Specialization (NT)	Agglomeration (NT)	Dissimilarity (NT)
Per capita GDP	15.20^{***} (3.98)	1.13^{***} (6.36)	-8.97^{**} (-2.89)
Obs.	7	7	7

Notes: The Table reports coefficients estimates in regressions of various indexes on Chinese per capita GDP. Specialization and agglomeration are measured using Herfindhal indexes. The upper panel computes indexes on the full sample of sectors (minus Farming, Forestry, Animal Husbandry and Fishery). The middle panel focuses on traded sectors (Mining and Quarrying, Manufacturing, Transport, Storage and Communication Services, Finance and Insurance). The lower panel focuses on non-traded sectors (Electricity, Gas, and Water, Construction, Wholesale and Retail Trade, Catering Services, Real Estate, Scientific Research and Polytechnic Services, Geological Prospecting and Water Conservancy, Education, Culture and Arts, Radio, Film and Television, Health Care, Sports and Social Welfare, Government Agencies, Party Agencies and Social Organizations, Social Services and Others). Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels.

Table 5: USA

	Specialization (All)	Agglomeration (All)	Dissimilarity (All)
Per capita GDP	0.407^{***} (8.68)	-0.333^{***} (9.35)	-0.083^{***} (−12.83)
Obs.	21	21	21

Notes: The Table reports coefficients estimates in regressions of various indexes on US per capita GDP. Specialization and agglomeration are measured using Herfindhal indexes. Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels.

Table 6: IPUMS - International Data

	Specialization (All)	Agglomeration (All)	Dissimilarity (All)
Low	-39.80^{***} (−11.71)	4.18^{***} (2.78)	0.100 (0.16)
High	−0.648 (−1.50)	−0.340 (−0.86)	-0.834^{***} (−10.23)
Obs.	51	51	51
	Specialization (T)	Agglomeration (T)	Dissimilarity (T)
Low	-49.80^{***} (−12.38)	6.73^{***} (3.95)	7.19^{***} (6.64)
High	-3.27^{***} (−3.96)	0.006 (0.01)	-1.54^{***} (−9.02)
Obs.	51	51	51
	Specialization (NT)	Agglomeration (NT)	Dissimilarity (NT)
Low	2.73 (1.20)	-5.34^{***} (−3.50)	-6.93^{***} (−4.16)
High	1.73^{***} (3.26)	-0.067^{*} (−1.74)	-0.712^{***} (−4.18)
Obs.	51	51	51

Notes: The Table reports coefficient estimates in a regression of various indexes on individual country's real per capita GDP estimated for "low" and "high" income sub-samples. The mid-point corresponds to median per capita GDP. The number of observations refers to each sub-sample. Specialization and agglomeration are measured using Herfindahl indexes. The upper panel computes indexes on the full sample of sectors. The middle panel focuses on traded sectors (Agriculture, Hunting and Fishing, Mining, Manufacturing, Transport,Telecommunication, Finance and Business Services). The lower panel focuses on non-traded sectors (Construction, Public Administration, Public Services, Retail, Hotels, and Other Services). Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels. All estimations include country-specific fixed effects.

Table 7: Real Value Added Data - National Sources

	Specialization (All)	Agglomeration (All)	Dissimilarity (All)
Low	-15.70^{***} (−3.48)	1.91^{*} (1.87)	-0.760^{*} (−1.80)
Medium	-0.601^{***} (−6.43)	-0.229 (−1.49)	-0.443^{***} (−10.20)
High	0.138 (0.98)	0.396^{*} (1.95)	-0.088^{*} (−1.86)
	Specialization (T)	**Agglomeration (T)**	**Dissimilarity (T)**
Low	-26.70^{***} (−2.86)	5.29^{***} (3.57)	5.40^{***} (3.90)
Medium	7.29^{***} (14.98)	-0.381^{**} (−2.23)	-1.95^{***} (−10.68)
High	-1.34^{*} (−1.69)	3.01^{***} (3.61)	1.11^{***} (3.80)
	Specialization (NT)	**Agglomeration (NT)**	**Dissimilarity (NT)**
Low	-4.70^{***} (−2.97)	-0.058 (−0.06)	-3.59^{***} (−2.85)
Medium	-0.698^{***} (−6.47)	-0.166 (−0.96)	-0.661^{***} (−8.49)
High	0.491^{***} (2.64)	-0.501^{***} (−4.39)	0.084 (0.78)

Notes: The Table reports coefficient estimates in a regression of various indexes on individual country's real per capita GDP estimated for "low", "medium" and "high" levels of per capita GDP. The thresholds are chosen to split observations on per capita GDP in three samples of equal size, with bounds \$6,500 and \$27,000 at 2005 PPP exchange rates. Specialization and agglomeration are measured using Herfindahl indexes. The upper panel computes indexes on the full sample of sectors. The middle panel focuses on traded sectors (Agriculture, Hunting and Fishing, Mining, Manufacturing, Finance and Business Services). The lower panel focuses on non-traded sectors (Construction, Public Administration, Public Services, Retail, Hotels, and Other Services). Coefficients are multiplied by 10^6. Student's t-statistics are reported between parentheses. *** $^{(**,*)}$ denote significance at 1% (5%, 10%) significance levels. All estimations include country-specific fixed effects.

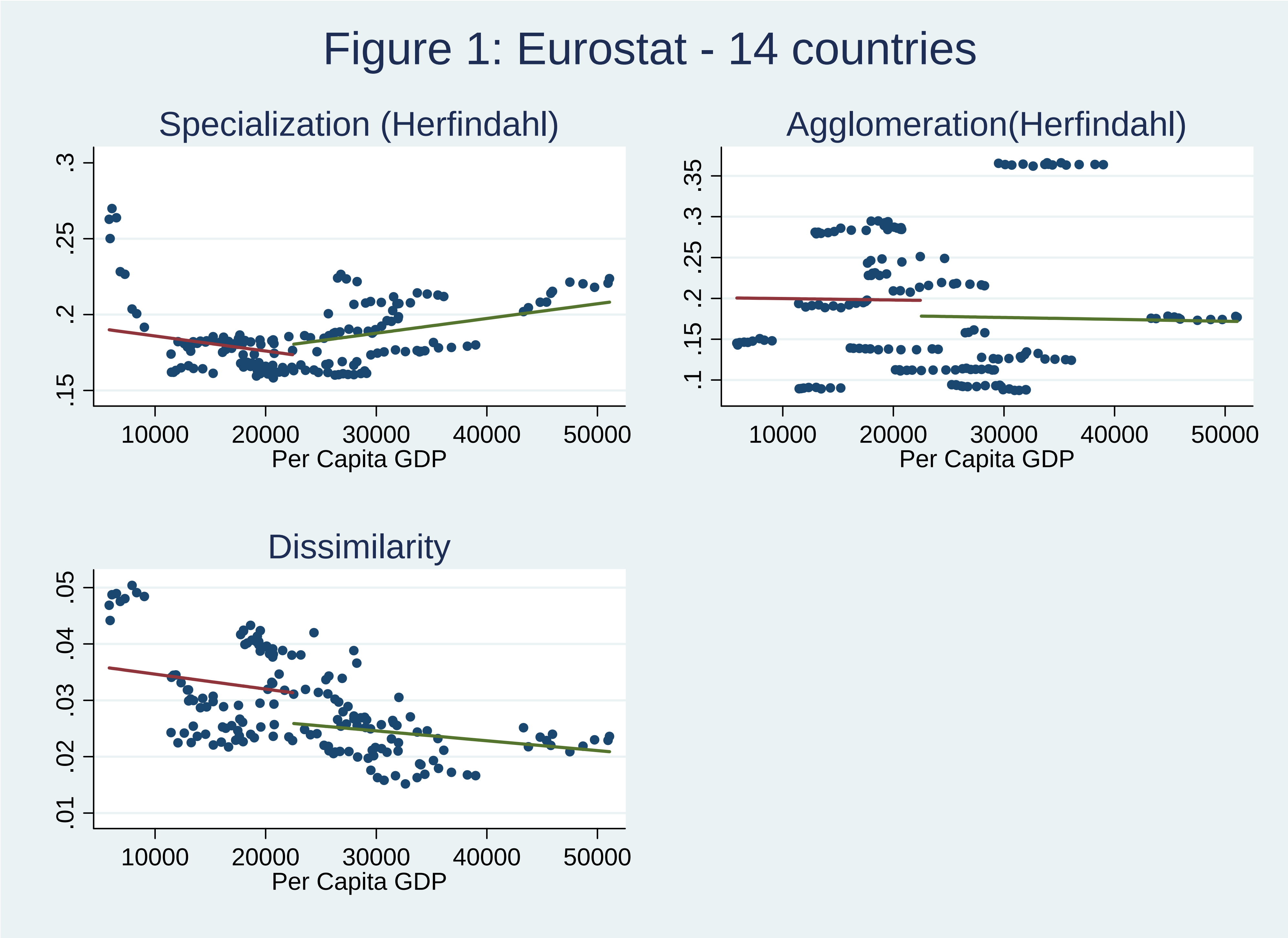
Figure 1: Eurostat - 14 countries
Specialization (Herfindahl)
.15
.2
.25
.3
10000
20000
30000
40000
50000
Per Capita GDP
Agglomeration(Herfindahl)
.1
.15
.2
.25
.3
.35
10000
20000
30000
40000
50000
Per Capita GDP
Dissimilarity
.01
.02
.03
.04
.05
10000
20000
30000
40000
50000
Per Capita GDP

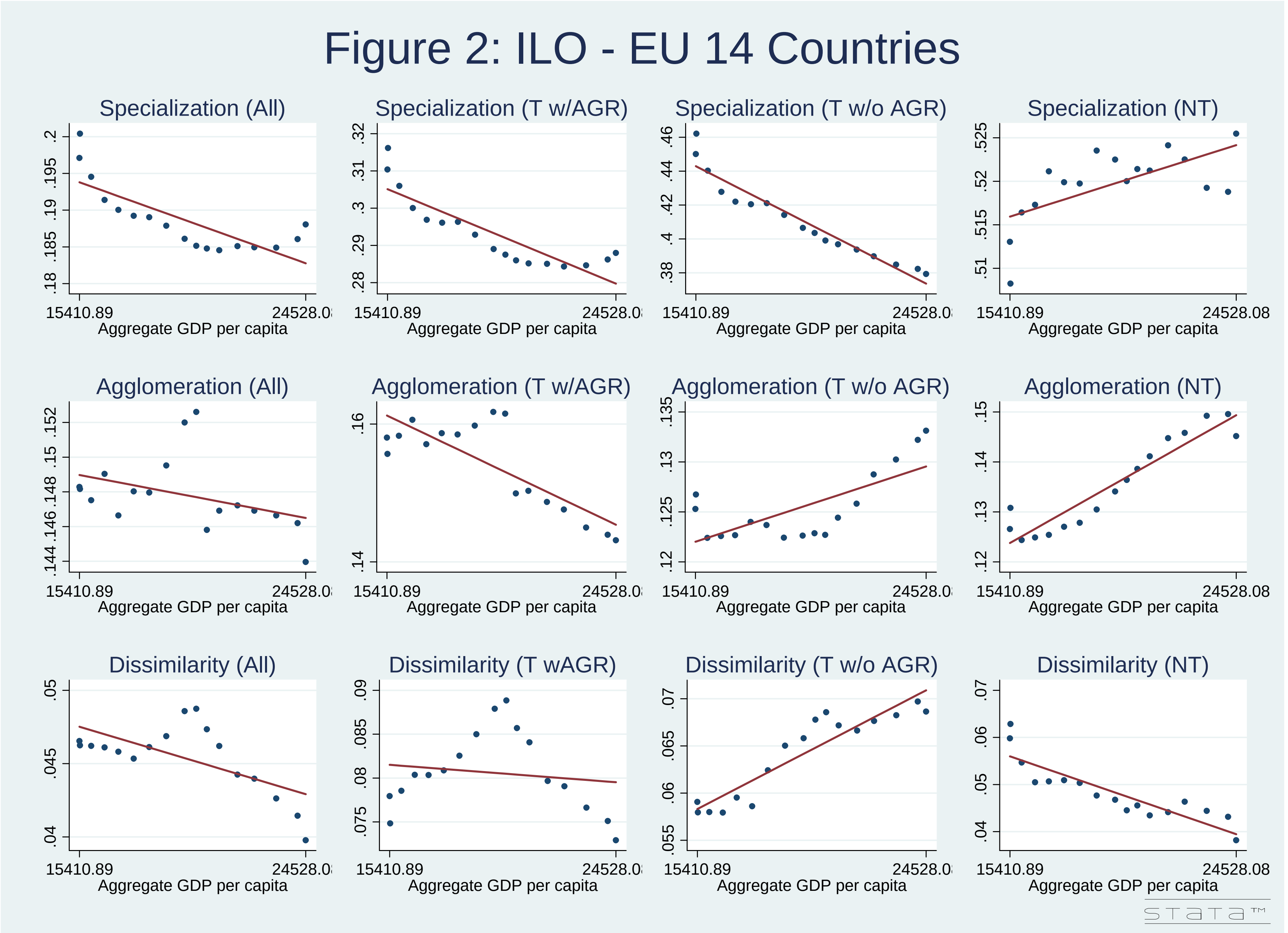

Figure 2: ILO - EU 14 Countries
Specialization (All)
Specialization (T w/AGR)
Specialization (T w/o AGR)
Specialization (NT)
Agglomeration (All)
Agglomeration (T w/AGR)
Agglomeration (T w/o AGR)
Agglomeration (NT)
Dissimilarity (All)
Dissimilarity (T wAGR)
Dissimilarity (T w/o AGR)
Dissimilarity (NT)
15410.89
24528.08
Aggregate GDP per capita

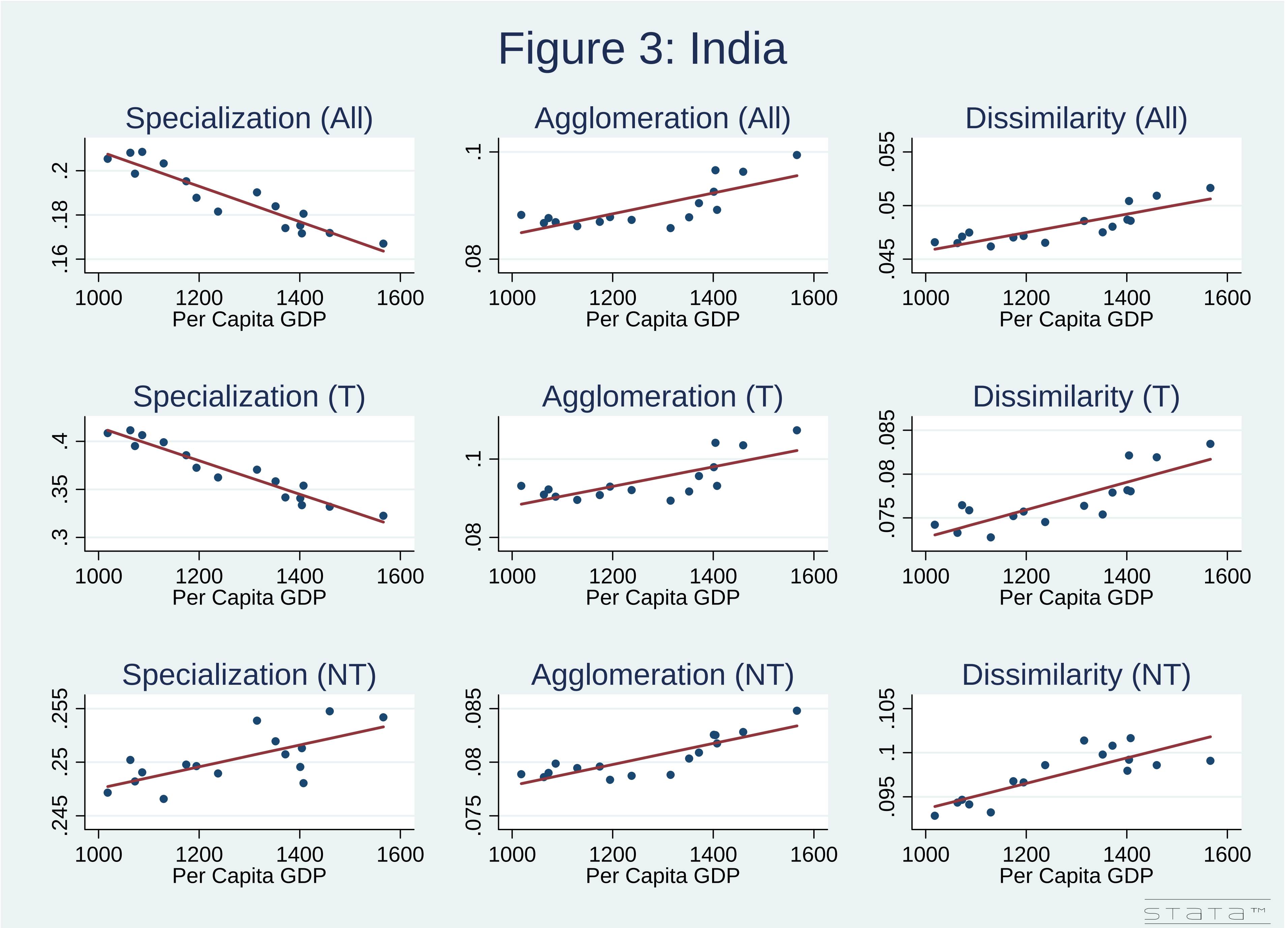

Figure 3: India
Specialization (All)
Agglomeration (All)
Dissimilarity (All)
Specialization (T)
Agglomeration (T)
Dissimilarity (T)
Specialization (NT)
Agglomeration (NT)
Dissimilarity (NT)
Per Capita GDP
1000
1200
1400
1600
.16
.18
.2
.08
.1
.045
.05
.055
.3
.35
.4
.075
.08
.085
.245
.25
.255
.095
.105
stata™

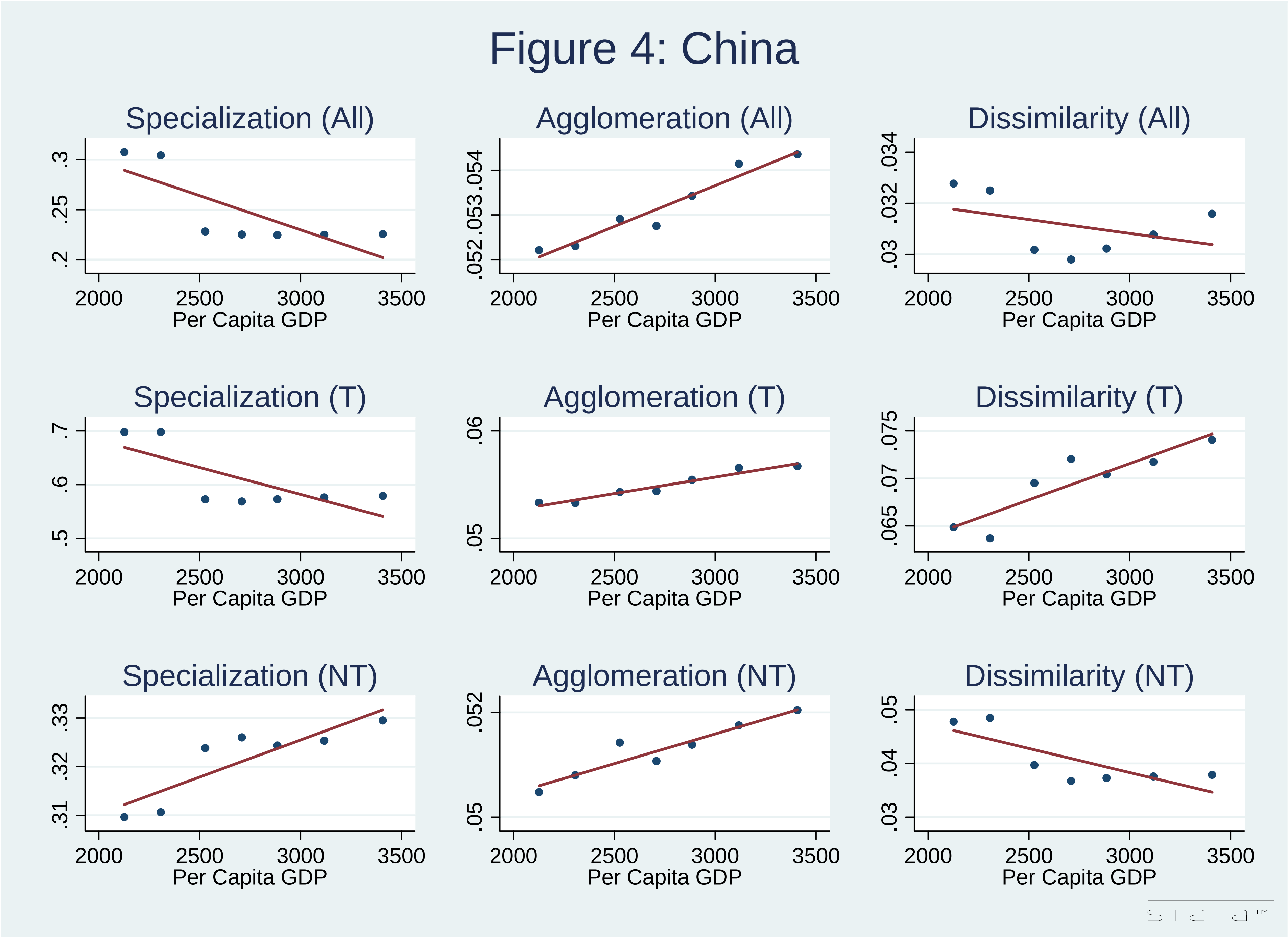
Figure 4: China
Specialization (All)
Agglomeration (All)
Dissimilarity (All)
Specialization (T)
Agglomeration (T)
Dissimilarity (T)
Specialization (NT)
Agglomeration (NT)
Dissimilarity (NT)
Per Capita GDP
2000 2500 3000 3500
.2 .25 .3
.052 .053 .054
.03 .032 .034
.5 .6 .7
.05 .06
.065 .07 .075
.31 .32 .33
.05 .052
.03 .04 .05
STATA™

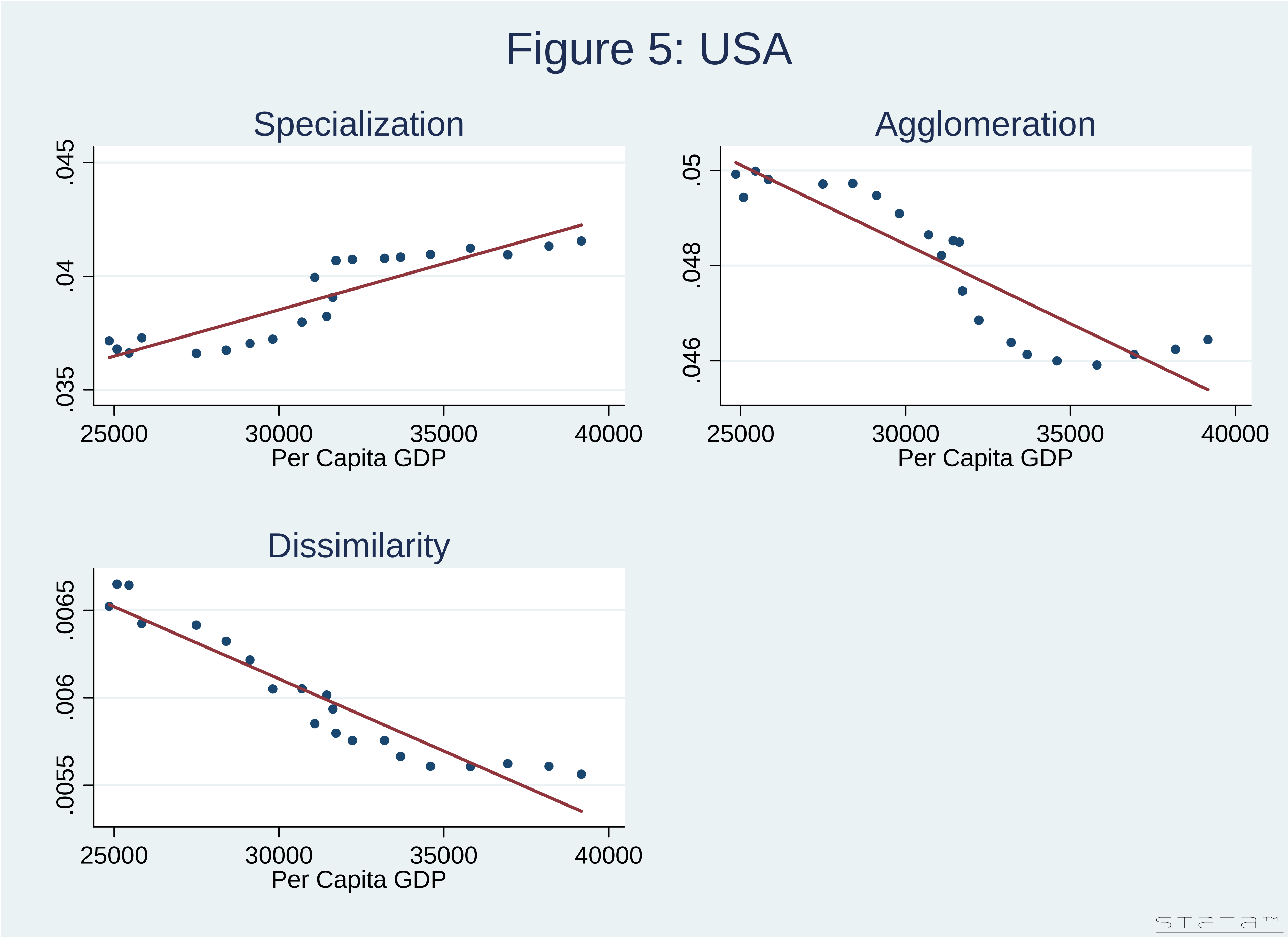

Figure 5: USA
Specialization
.045
.04
.035
25000
30000
35000
40000
Per Capita GDP
Agglomeration
.05
.048
.046
25000
30000
35000
40000
Per Capita GDP
Dissimilarity
.0065
.006
.0055
25000
30000
35000
40000
Per Capita GDP

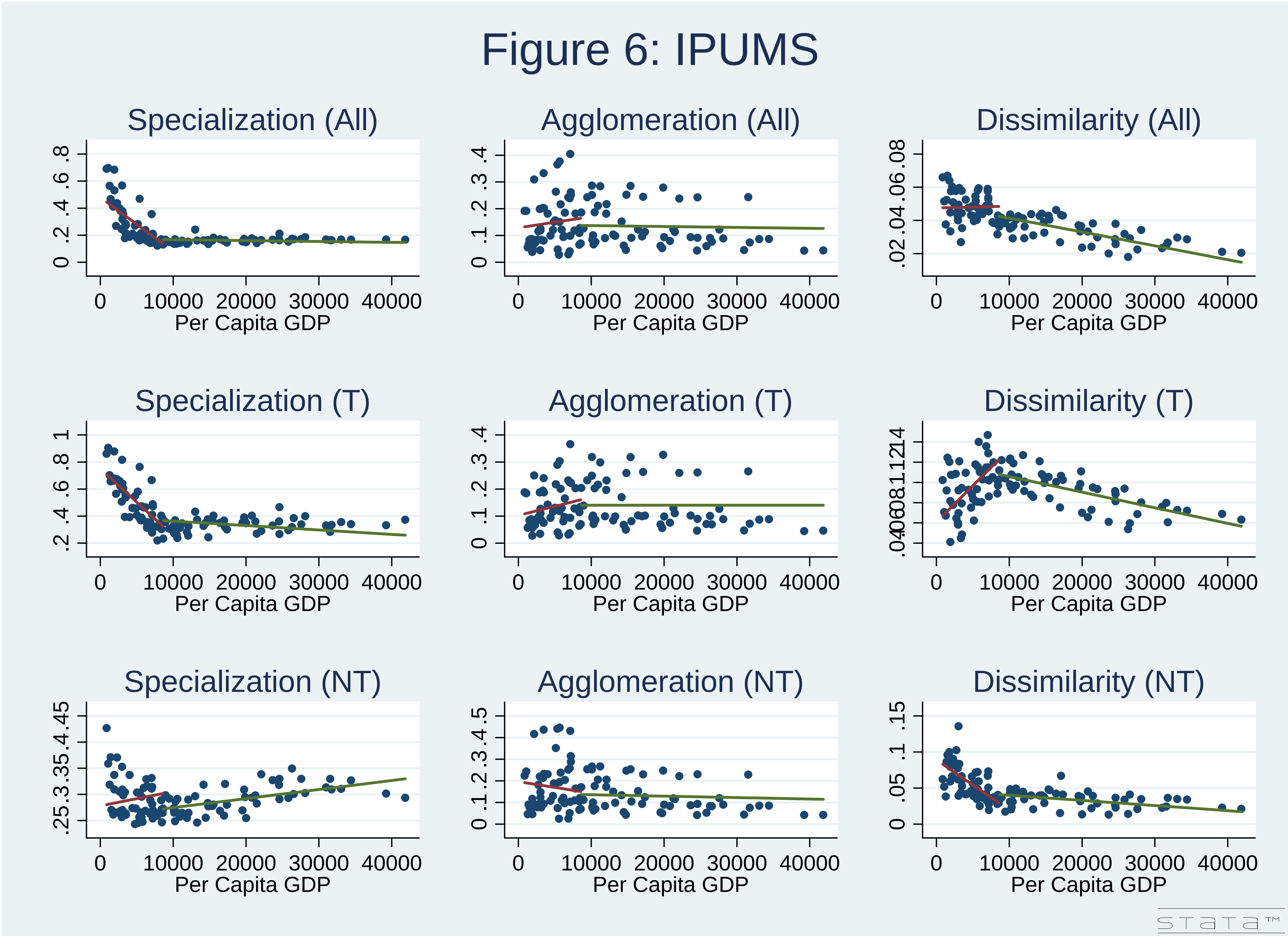

Figure 6: IPUMS
Specialization (All)
Agglomeration (All)
Dissimilarity (All)
Specialization (T)
Agglomeration (T)
Dissimilarity (T)
Specialization (NT)
Agglomeration (NT)
Dissimilarity (NT)
0 10000 20000 30000 40000
Per Capita GDP

Figure 7: Examples with Output Data

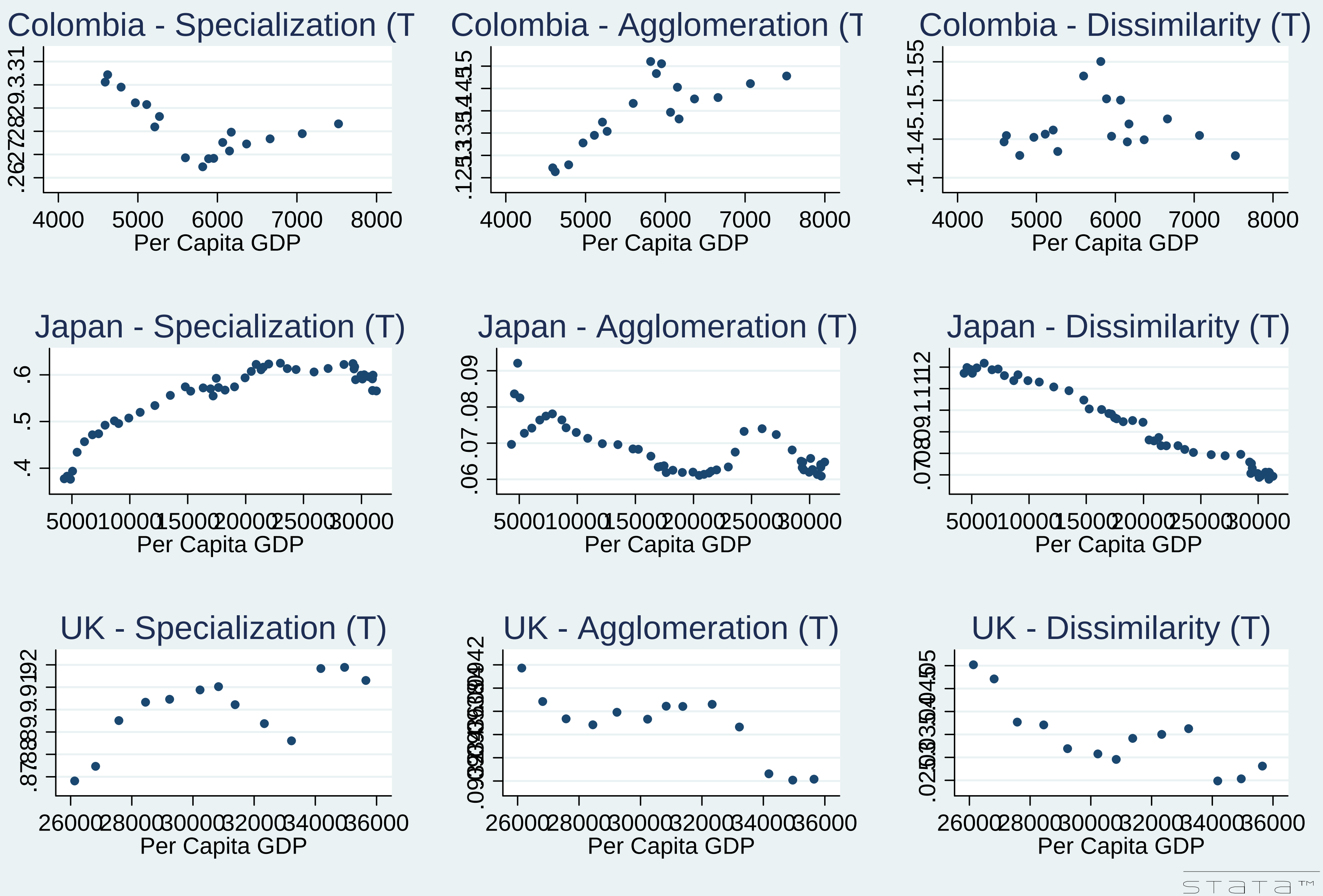

Figure 8: Output Data