Scalable shift-share analysis: Novel framework and application to France

Jean-François Ruault, Yves Schaeffer

Scalable Shift-Share Analysis: Novel Framework and Application to France.

Ruault[*] J.-F., Schaeffer[*] Y. [‡]

[*] Univ. Grenoble Alpes, INRAE, LESSEM, 2 rue de la Papeterie-BP 76, F-38402 St-Martin-d'Hères, France

[‡]Authorship is equally shared among the two co-authors.

jean-francois.ruault@inrae.fr (the addressee) and yves.schaeffer@inrae.fr

Abstract

The shift-share analysis (SSA) of regional employment growth disparities aims at disentangling the effects of regional differences in industry mix and industrial competitiveness. Yet, the spatial concentration of industry is a blindspot of this approach. We generalize the SSA to encompass this salient feature of the economic geography. Besides, industry emergence and catastrophic growth events - booming or collapsing industries - are integrated in our framework. This novel method is applied to study regional disparities in manufacturing employment growth in France in a dynamic way over a 22-year period at a fine geographical level.

Keywords: Dynamic shift-share analysis**,** Regional inequality**,** Regional employment growth**,** Manufacturing**,** France

JEL classifications: R11, R12, C65

1. Introduction

A well-known characteristic of Krugman's (1991) pioneering two-regions two-sector microeconomic model - with transportation costs and increasing returns to scale - is that it features a process of agglomeration of manufacturing firms when the decrease of transportation costs reaches a critical threshold level. The outcome is a stable locational equilibrium, named core-periphery, where all manufacturing firms are concentrated in one region and only agricultural activities remain in the other. The main lesson of this simplified model is that strong regional economic inequalities can arise from regions that are initially perfectly identical. Later on, Krugman and other scholars investigated more realistic settings, involving multiple regions, multiple industries, endogenous growth factors, technological spillovers and the role of natural geography and history in determining the economic geography (Behrens and Robert-Nicoud, 2009). All these modeling efforts, known as the New Economic Geography (NEG) research field, were motivated by a salient empirical fact: the very high geographical concentration of industry and economic growth from the Industrial Revolution to the present. Beyond the possibility of core-periphery configurations, a key insight with respect to the spatial distribution of industries is that "*distance matters in such a way that agglomerations of each industry (roughly) take place at a certain industry-specific distance, where the area between agglomerations is in the agglomeration shadow of the industry, and thus not a profitable location*" (Fujita and Mori, 2005, p. 392). In Krugman's world, industries should not be expected to spread out everywhere and, indeed, they do not.

However, such important and illustrative findings for regional economics, both theoretical and empirical, are currently disregarded by the shift-share analysis (SSA). The SSA, usually credited to Dunn (1959; 1960), is a more rudimentary but still popular approach for the study of interregional differences of growth in a country. Its essential aim, well summarized by

Esteban (2000, p. 355), is to empirically analyze: "*the extent to which the difference in growth between each region and the national average is due to the region performing uniformly better than average on all industries or to the fact that the region happens to be specialized in fast-growing sectors*." The focus is on decomposing the regional growth, usually of employment, between (at least) two components: a competitive effect independent of the regional economic structure and an industry-mix effect which depends on it. Three fundamental decomposition formulae have been proposed by Dunn (1959), Esteban-Marquillas (1972) and Artige and van Neuss (2014) and several extended versions are available, either to refine the interpretation of the components (Arcelus 1984), make it dynamic (Barff and Knight, 1988), or incorporate spatial relationships (Nazara and Hewings 2004; Espa et al. 2014; Mussini 2019). However, none of these methods acknowledges empirical evidences or NEG lessons properly: they all assume, at least implicitly, that all national industries exist in all regions, whatever the geographical scale of these latter and the refinement of the breakdown of employment into industrial sectors.

Actually, the difference between the sets of industrial sectors at the regional versus the national level is a blindspot of the SSA. Its original focus was on the shift between the observed growth of a region and the counterfactual growth that could have been expected in this region considering the "*forces operating at the national level*" (Dunn 1960:97). Formally, Dunn's growth decomposition formulae was strictly related to the set of industrial sectors observed in a region of interest and the comparison of its local and national growth dynamic. And "*for simplicity's sake*," the statistical technique was applied on an example of "*region[s] having the same four industrial sectors*" (Dunn 1960:110). As far as we know, the following developments have kept this prerequisite of a common set of sectors in all regions. This requirement appears much too stringent. For instance, let us take the case of France, which is illustrative of advanced economies, and where a time series of employment data at different spatial scales is available. The European statistical classification of economic activities lists 24 divisions for manufacturing activity (see Appendix 1). In the French Hexagon in 2015, the number of manufacturing industries varies between 22 and 24 at the NUTS 2 level, between 18 and 24 at the NUTS 3 level, and between 1 and 24 at the LAU 1 level.[1] This example shows that the SSA should be qualified to be useful in the common context where many regions have missing sectors (see also Appendix 2 for striking maps). In addition, sectors initially missing may emerge over time: in the French Hexagon, industry emergence between 2014 and 2015 occurred, respectively, in 8%, 6% and 26% of NUT 2, NUTS 3 and LAU 1 regions. Over the 1993-2015 period, it is 83%, 85%, and 100% of these regional units that are concerned. These phenomenons are not taken into account in standard methods, making SSA not spatially scalable.

The main objective of this article is thus to make the SSA spatially (and sectorally) scalable by incorporating these industry concentration and industry emergence issues, as well as a third important issue also hitherto ignored: the effect of extraordinary growth or decline events. At very aggregated geographical and industry-structure levels, sectoral employment growth rates are expected to be moderate, but as soon as more disaggregated levels are used, some industries may present outstanding positive or negative growth rates. The competitive and industry-mix effects are likely to be affected by these outlier values, and the growth decomposition may no longer make sense. Again, the French example shows that this phenomenon is empirically relevant: for instance, between 2014 and 2015, 75% of NUTS 2, 91% of NUTS 3, and 86% of

[1] The Nomenclature of territorial units for statistics (NUTS) is a hierarchical system for dividing up the economic territory of the EU for the purpose of European regional statistics; Eurostat also maintains a system of Local Administrative Units (LAUs) compatible with NUTS (see https://ec.europa.eu/eurostat). In France, the LAU 1 region is a demographic and electoral grouping of one or more entire municipalities: a populated municipality is usually a single LAU1 region while less populated municipalities are grouped in LAU1 regions.

LAU 1 regions have hosted some manufacturing sectors with tremendous annual growth rates (as defined below). Overall, all regions without exception have hosted such sectors between 1994 and 2015.

To address these issues, we suggest a generalized decomposition of regional employment growth. A key idea taken from Artige and van Neuss (2014) is that a pro-growth local industrial mix is one that deviates from a uniform distribution in such a way that the largest sectors are those which display *locally* the higher growth (whereas "*fast-growing sectors*" in the industry-mix effect of Dunn and Esteban are defined with respect to national dynamics). Hence, our method does not indicate whether or not a regional industry-mix is beneficial given national sectoral growth trends, and whether or not a region is growing faster or slower than expected in view of its industrial mix. Regional growth effects are now defined independently of any national reference. They reflect two facets of regional performance, linked either to the advantageous allocation of specific local sectoral performances in the local economic structure, or to growth performances that are widely shared across the local economic structure. Contrarily to Artige and van Neuss (2014), however, we do not require each region to have the same number of sectors as the nation. Also, we recognize that some sectors can disappear or emerge over time at the regional scale, and that emergent sectors - the emergence effect - matter to regional growth. Last but not least, we treat "outlier" industries separately from "normal" ones and define an additional "outliers effect" in the growth decomposition framework.

The value of our generalized decomposition approach is illustrated by an empirical study. We investigate regional growth inequality using French data on the year-to-year dynamics of manufacturing employment over the period 1994-2015, at three distinct spatial scales (22 NUTS 2 regions, 96 NUTS 3 regions, and about 2000 LAU 1 regions), using a breakdown into

24 industries. We answer the question of whether regional inequality in manufacturing employment growth rates is attributable mainly to competitive or industry-mix effects. To do so, we contrast the variance decomposition approach proposed by Esteban (2000) to an original approach based on the median absolute deviation around the median. Finally, spatial autocorrelation issues are considered.

The article is organized as follows. Section 2 presents the most usual shift-share decompositions and our alternative proposal. Section 3 applies our approach to answer our empirical question. Section 4 concludes.

2. The shift-share analysis of regional inequalities

The SSA attempts to explain the differences in growth rates between two economies by making the contributions of regional specialization and regional business performance independent. However, the three main SSA techniques have drawbacks: they fail either to design independent components or to make SSA spatially and sectorally scalable, or both (2.1). We therefore introduce a novel SSA technique adressing these failures (2.2).

2.1. Principles and drawbacks of shift-share methods

In their recent paper, Artige and van Neuss (2014) provide a neat presentation of the canonical shift-share methods proposed by Dunn (1959, 1960) and Esteban-Marquillas, (1972), and a critical assessment justifying their own proposal.

The original idea of Dunn is to compare the regional employment growth with the growth that the region would have experienced were its growth rate equal to the national one. If the region performs actually better than in this hypothetical scenario, the question raised is whether it is because it has more advantageous industrial specialization than the nation or because its industry performs better. The objective of the SSA is thus to decompose this growth differential into two components: an *industry-mix effect*, related to the specific structure of the regional industry but independent of the local sectoral growth rates; and a *competitive effect*, stemming from the growth performances of the local sectors (supposedly) independently of the first component:

$$g^j_{t+1} - r_{t+1} = \sum_{i=1}^{I} (\omega^j_{i,t} - \theta_{i,t}) r_{i,t+1} + \sum_{i=1}^{I} (g^j_{i,t+1} - r_{i,t+1}) \omega^j_{i,t} \tag{1}$$

where g^j_{t+1} is the growth rate of region j between time t and $t+1$, r_{t+1} is the national growth rate between time t and $t+1$, $\omega^j_{i,t}$ is the share of employment of region j in the sector i at time t, $\theta_{i,t}$ is the national share in the sector i at time t, $g^j_{i,t+1}$ is the growth rate of region j in the sector i between time t and $t+1$, and $r_{i,t+1}$ is the national growth rate in the sector i between time t and $t+1$.

The main criticism of this approach has been immediately raised by Rosenfeld (1959), who noted that two regions with identical sectoral growth rate distributions but different underlying economic structures (i.e. growth rate values are not attached to the same specific sectors) would have different competitive effects. This shows that the competitive effect is not genuinely independent of the local industry mix.

A solution to this problem has been proposed by Esteban-Marquillas (1972), with the following formula:

$$g_{t+1}^{j} - r_{t+1} = \sum_{i=1}^{I} (\omega_{i,t}^{j} - \theta_{i,t}) r_{i,t+1} + \sum_{i=1}^{I} (g_{i,t+1}^{j} - r_{i,t+1}) \theta_{i,t} + \sum_{i=1}^{I} (\omega_{i,t}^{j} - \theta_{i,t})(g_{i,t+1}^{j} - r_{i,t+1}) \qquad (2)$$

The industry-mix effect is left unchanged. But Dunn's competitive effect is decomposed into two terms: a renewed *competitive effect* and an *allocation effect*. The competitive effect is computed assuming that the region's industry mix coincides to the national one. In Esteban-Marquillas' words (1972, p. 254): "*we consider the consequences of the specialization of its employment as being independent of the sectorial dynamism (industry-mix effect), and the consequences of the sectorial dynamism independently of the specialization of its employment (competitive effect)*". If the competitive effect is made independent of the local economic structure, then the comparability between regions is (supposed to be) ensured, which responds to Rosenfeld's criticism.

Artige and van Neuss (2014) contend that Esteban's approach does not really solve the inconsistency identified by Rosenfeld. To prove it, they present a simple two-regions two-sectors shift-share test. The two regions have equal employment growth rates and perfectly similar joint distributions of sectoral growth rates and structural weights, and only the names of the principal and secondary sectors differ (A is the larger and more dynamic sector in region 1 but the smaller and less dynamic sector in region 2; the reverse is true for B). According to Artige and van Neuss, the conclusion of the test should be obvious: both regions have equivalent growth performances, and their specialization and competitiveness are equally good. However, Esteban's method provides different industry-mix and competitive effects for the two regions. Both effects are null in region 1, whose economy is perfectly similar to the one of the nation. But in region 2, the industry-mix effect is negative: the regional growth performance would be lower had the national sectoral dynamics applied to its structure. And the competitive effect is also negative: region 2 has lower growth rates in the sectors that are prominent at the national scale and is thus considered less competitive. In other words, the competitive effect depends on which local sectors – specifically – are growing, and not merely on the shape of the local distribution of sectoral growth rates. As a consequence, Rosenfeld criticism still holds: inter-regional comparisons of effects are problematic due to the use of national references in the definition of the industry-mix effect and the competitive effect.

Artige and van Neuss thus propose a major departure from previous approaches: to define both effects independently of any national reference. Growth rates are decomposed in the first place at the geographical unit level:

$$g_{t+1}^{j} = \sum_{i=1}^{I} \left(\omega_{i,t}^{j} - \frac{1}{I} \right) g_{i,t+1}^{j} + \sum_{i=1}^{I} \ldots_{1} \frac{1}{I} \quad (3)$$

$$r_{t+1} = \sum_{i=1}^{I} \left(\theta_{i,t} - \frac{1}{I} \right) r_{i,t+1} + \sum_{i=1}^{I} r_{i,t+1} \frac{1}{I} \quad (4)$$

where I is the number of sectors in the national economy (and $\frac{1}{I}$ the share of employment in any sector assuming a uniform distribution).

The key idea is that a pro-growth industry-mix is one that deviates from an uniform distribution in such a way that the largest sectors are those which display the higher growth. Thus the regional industry-mix effect equals zero when the economic structure is perfectly uniform or when overrepresented and underrepresented industries have equal average growth rates; it is positive when the overrepresented industries grow faster than the under-represented ones and, conversely, negative when the latter grow faster than the former. Second, the average sectoral growth rate (that would be obtained if each sector had the same weight) can be considered as a performance indicator truly independent of the industrial mix.

The decomposition of the growth rates differential between the region and the nation into an industry-mix effect and a competitive effect is then obtained merely by grouping and rearranging the corresponding terms of equations (3) and (4):

$$g_{t+1}^{j} - r_{t+1} = \underbrace{\left[\sum_{i=1}^{I}\left(\omega_{i,t}^{j} - \frac{1}{I}\right)g_{i,t+1}^{j} - \sum_{i=1}^{I}\left(\theta_{i,t} - \frac{1}{I}\right)r_{i,t+1}\right]}_{Industry-mix\ effect} + \underbrace{\sum_{i=1}^{I}\left(g_{i,t+1}^{j} - r_{i,t+1}\right)\frac{1}{I}}_{Competitive\ effect} \quad (5)$$

2.2. Reassessment and proposals

We endorse the substantial shift made by Artige and van Neuss (2014), which consists in defining regional growth effects independently of a national reference. Hence, the industry-mix effect does not reflect "*national forces*" any more, and the competitive effect is no more a regional residual. Rather, these effects reflect two facets of the differences in regional growth, either based on differences in overall sectoral competitiveness (i.e. some regions outperforming others in all their sectors), or based on differences in the distribution of local sectoral growth rates in local economic structures (i.e. some regions being favored by the dynamism of their main industries). This region-centered approach remains quite in line with Esteban's (2000, p. 355) statement that the SSA should help identify: "*the extent to which the difference in growth between each region and the national average is due to the region performing uniformly better than average on all industries or to the fact that the region happens to be specialized in fast-growing sectors*," although "*fast-growing sectors"* are now defined at the local rather than at the national level. That being said, we believe that three qualifications must be made to Artige and van Neuss's proposal in order to make it fully exploitable and scalable in practice.

First, like previous scholars in the field, Artige and van Neuss (2014) did not consider the extent of agglomeration shadows (Fujita and Mori 2005; Burger et al. 2015) seriously enough (cf. Appendix 2 and introduction section). Their method still requires that all regions whose growth performances are to be compared have the same number of industries – I in equation 5 – regardless of their geographical scale and the refinement of the breakdown into sectors. Otherwise, $g_{i,t+1}^{j}$ is undefined for missing sectors and neither the industry-mix effect nor the competitive effect can be computed. Therefore, the decomposition formula must recognize that the number of sectors can be different between a region and the nation it belongs to, and more broadly between any two regions:

$$g_{t+1}^{j} - g_{t+1}^{k} = \underbrace{\left[\sum_{i\in S^j}\left(\omega_{i,t}^{j} - \frac{1}{|S^j|}\right) g_{i,\,t+1}^{j} - \sum_{i\in S^k}\left(\omega_{i,t}^{k} - \frac{1}{|S^k|}\right) g_{i,\,t+1}^{k}\right]}_{Industry-mix\ effect} + \underbrace{\left[\sum_{i\in S^j} g_{i,\,t+1}^{j}\frac{1}{|S^j|} - \sum_{i\in S^k} g_{i,\,t+1}^{k}\frac{1}{|S^k|}\right]}_{Competitive\ effect} \qquad (6)$$

where S^j and S^k are sets of industrial sectors in region j and k, respectively, and $|S^j|$ and $|S^k|$ indicate the number of sectors in each set.

Equation 6 is very similar to equation 5, but no longer involves any reference economy, that is why it is now formulated as a two-region difference (i.e. $g_{t+1}^{j} - g_{t+1}^{k}$ in equation 6) and not specifically as a region-nation difference (i.e. $g_{t+1}^{j} - r_{t+1}$, as in equation 1, 2 and 5). For each region, a pro-growth economic structure (i.e. with a positive industry-mix effect) is one in which the main industries grow faster. We do not assume that it is an advantage to be specialized in sectors that grow fast elsewhere, because there may be good reasons why they develop elsewhere and not here: the condition for growth may not be met locally. Specialization is also not considered better or worse depending on the number and quality of missing sectors, since we do not know what the performances of these sectors would be in this region, and it is possible that some sectors just cannot exist in this place (e.g., think about oil extraction industry if there is no oil). A competitive industry - abstracting from structural advantages - is one in which existing sectors perform well on average (i.e. they are growing, which means the competitive effect is positive), and again we do not make hypotheses on the potential behaviors of absent sectors.

The second qualification is also connected to absent sectors, but relates to the time dimension of the SSA. This dimension imposes to add a third component to our decomposition. Indeed, sectors that are missing in a region at time t can emerge and be present at time $t+1$. These sectors obviously contribute to the overall growth performance of the region but it is not meaningful to incorporate them into the industry-mix or competitive effects. Indeed, we cannot evaluate the growth performance of an emergent sector since its growth rate is by definition infinite. Thus "emergence effect" must be considered as an additional independent growth component. This novel component is introduced in equation 7: for each region, it is merely the percent increase in jobs between t and $t+1$ that stems from emergent sectors, and the effect (ε) on the two-region growth gap is the difference between these regional increases.

The third qualification addresses an important outlier issue. At geographically or structurally disaggregated levels, some industries may present tremendous growth rates. A local industry that represents only a small number of jobs at the time *t* may easily experience a double or triple-digit growth rate, which is virtually unattainable for large industries (for example, the manufacture of electrical equipment had 5 jobs in the Brumath LAU 1 region in 2014 and 106 jobs one year later, and thus experienced a soaring growth rate of +2000%). Mixing these booming sectors with the others makes the interpretation of the competitive effect problematic. The competitive effect should indeed reflect the central tendency of sectoral growth rates in a region (i.e. "*the extent to which […] the region [is] performing uniformly better than average on all industries*" to cite again Esteban, 2000). But including the explosive growth rate of a booming sector in the computation of the average sectoral growth rate is likely to mask the general tendency followed by most industries in the region.[2] A similar problem might exist for the decline: it is more likely for small sectors to have strong negative growth rates (although there is a floor value of -100%). Put differently, the competitive effect, which is based on simple means, is very sensitive to extreme values.[3] Due to this sensitivity, the decomposition between competitive and industry-mix effects is meaningful only for "normal" industries, which experience "normal" growth rates.

Our proposal is thus to distinguish between normal and extreme growth rates based on Tukey's interquartile range (IQR) approach[4]: booming and collapsing sectors are those which display

2 This issue has previously been mentioned by Lamarche et al. (2003). The (unsatisfactory) solution proposed was to use more sectorally or geographically aggregated nomenclatures to mitigate the problem.
3 Esteban's competitive effect - although not based on simple means - is also highly sensitive to the presence of booming or collapsing sectors: it is computed assuming that the region's industry mix coincides to the national one, so that large weights can be applied to actually very small sectors experiencing tremendous growth (or decline) dynamics. Dunn's competitive effect is less sensitive, as local sectoral growth rates are weighted by the regional economic structure, but it is not immune to this problem.
4 Tukey fences have the advantage not to take a symmetric view on the dispersion and to depend on a robust scale estimator (i.e. the interquartile range). In the empirical application that follows, we identify these extreme values in the overall distribution of sectoral growth rates for all regions included in the analysis.

extreme growth rates ranging respectively above or below Tukey's fences (resp. the 3rd Quartile + 1.5*IQR or the 1st Quartile - 1.5*IQR). The contribution of these sectors to the regional growth differential is then introduced as an independent "outliers effect" in equation 7. As for sectoral emergence, it computed for each region as the percent increase (or decrease) in jobs between *t* and *t+1* that stems from those specific sectors, and the effect (ω) on the two-region growth gap is the difference between these regional figures.

Our final decomposition formula is thus:

$$g_{t+1}^{j} - g_{t+1}^{k} = \underbrace{\left[\frac{{L_{N,t}}^{j}}{L_{t}^{j}} \sum_{i \in N^{j}} \left(\omega_{i,t}^{j} - \frac{1}{|N^{j}|}\right) g_{i,t+1}^{j} - \frac{{L_{N,t}}^{k}}{L_{t}^{k}} \sum_{i \in N^{k}} \left(\omega_{i,t} - \frac{1}{|N^{k}|}\right) g_{i,t+1}^{k}\right]}_{Industry-mix\ effect\ (\mu)} \tag{7}$$

$$+ \underbrace{\left[\frac{{L_{N,t}}^{j}}{L_{t}^{j}} \sum_{i \in N^{j}} g_{i,t+1}^{j} \frac{1}{|N^{j}|} - \frac{{L_{N,t}}^{k}}{L_{t}^{k}} \sum_{i \in N^{k}} g_{i,t+1}^{k} \frac{1}{|N^{k}|}\right]}_{Competitive\ effect\ (\pi)}$$

$$+ \underbrace{\left[\sum_{i \in E^{j}} \frac{L_{i,\,t+1}^{j}}{L_{t}^{j}} - \sum_{i \in E^{k}} \frac{L_{i,\,t+1}^{k}}{L_{t}^{k}}\right]}_{Emergence\ effect\ (\varepsilon)} + \underbrace{\left[\sum_{i \in O^{j}} \frac{\left(L_{i,t+1}^{j} - L_{i,t}^{j}\right)}{L_{t}^{j}} - \sum_{i \in O^{k}} \frac{\left(L_{i,t+1}^{k} - L_{i,t}^{k}\right)}{L_{t}^{k}}\right]}_{Outlier\ effect\ (\omega)}$$

where E^{j} is the set of emergent sectors in region j over the period ($E^{j} \notin S^{j}$). $L_{i,t+1}^{j}$ is the number of jobs of sector i in region j and L_{t}^{j} is the total number of jobs in region j. O^{j} is the set of sectors with outlier growth rates (i.e. ranging above or below Tukey's fences) in region *j* and N^{j} is the set of normal sectors in region *j* (such as $N^{j} = S^{j} - O^{j}$). $|N^{j}|$ is the number of sectors i in the set N^{j}. Notations are the same for the region k. Note that each component is now named by a Greek letter, useful in section 3.2.

As we can note, the competitive (π) and industry-mix (μ) effects are similar to those presented in equation 6 (and explained at that point), except that each regional figures within them are now weighted by their corresponding regional percent jobs in normal sectors, as regional growth rates can now also be affected by emergent and outliers sectors.

To conclude this section, we should stress one last time that our innovative SSA equation continues the ideas of Artige and van Neuss's (2014) by really decomposing regional growth rates independently in each region: it finally makes the SSA both spatially and sectorally scalable.

3. A case study on regional growth inequality

The French case study is here illustrative of the SSA disaggregation issue in advanced economies and chosen considering the availability of detailed data at different spatial scales. A key objective of the SSA is to shed light on the roots of regional growth disparities in a country and, more specifically, on the extent to which they are mainly due either to some regions performing better than others on all industries, or to some regions being favored by suitable specializations in high-growth industries. Our first aim is thus to provide an answer to this question based on French data on manufacturing industries dynamically[5] analyzed over a 22-year period. We also consider whether our new SSA framework was necessary to address this analysis and its added value. Eventually, the issue of the spatial dimension of growth is introduced and examined empirically.

3.1. Data and context

Two data sources were used to monitor the French manufacturing employment year after year over a twenty-year period. For years 2008 to 2015, we used the CLAP database of the National

[5] As formerly suggested by Barff and Knight III (1988), a dynamic SSA provides a more accurate allocation of job changes among shift-share components.

Institute of Statistics and Economic Studies (Insee),[6] which reports the number of salaried job positions (except for Defense and household employers) on December 31 of each year at the municipal level, distinguishing among the 732 different activities of the French statistical classification of economic activities (NAF rev. 2). For years 1993 to 2008, we used the database of the former French agency for unemployment insurance system (Assédic), which provides a similar information, but excluding state employees and agricultural sectors.

Based on these datasets, we computed gross and sectoral annual growth rates at three spatial scales (LAU 1, NUT 3, NUT 2) for the years 1994 to 2015. These rates were systematically calculated with annual data coming from the same dataset: the annual growth rate of 2008 based on the number of Assédic jobs in 2007 and 2008, and the annual growth rate in 2009 based on the number of CLAP jobs in 2008 and 2009.

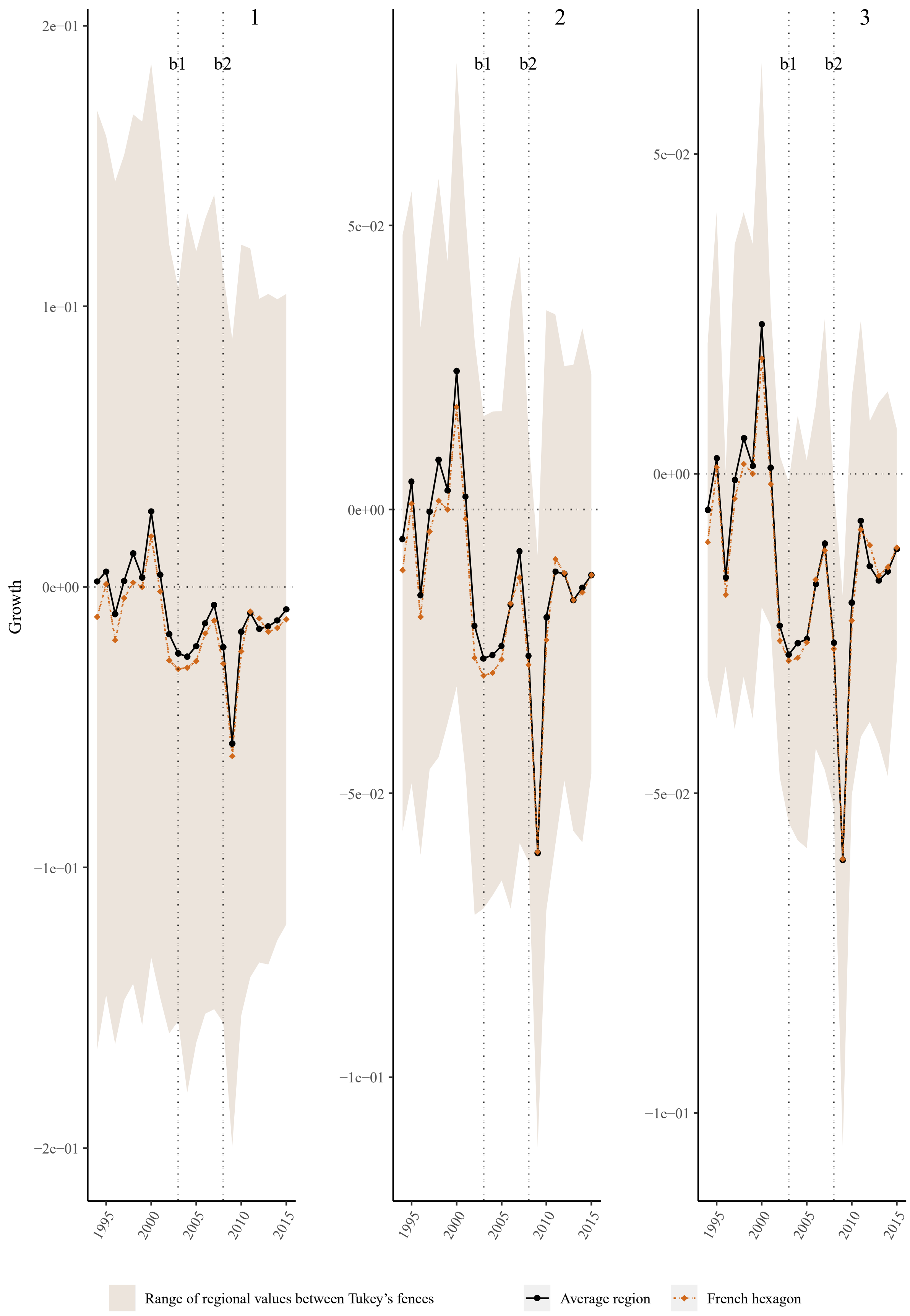
Despite this precaution, the source change unmistakably introduce a first statistical break in 2008. In addition, the NAF has been officially revised twice during the period, in 2003 and 2008, introducing another statistical break in 2003. These statistical breaks are not innocuous: the overall trend with respect to the growth of manufacturing activity is not significantly affected by the source change of 2008, but the distribution of jobs between business sectors is; and the distribution of jobs between sectors is obviously affected by the NAF revisions.

To alleviate this problem, we decided to conduct analysis on a breakdown of the manufacturing activity into 24 divisions (equivalent to the divisions of the NACE rev. 2, see Appendix 1), rather than on more detailed breakdowns. We also focused on the general patterns observed over the whole period, rather than on particular years. At the end, we did not see any significant impacts from revisions to the NAF or change in sources that could call our results into question.

[6] Connaissance locale de l'appareil productif (CLAP) : Tabulation sur mesure, INSEE [producteur], ADISP [diffuseur].

Figure 1 presents the French Hexagon manufacturing employment growth rate at the LAU 1 (Figure 1.1), the NUTS 3 (Figure 1.2) and the NUTS 2 (Figure 1.3) spatial scales, as well average value and Tukey fences for regional growth rates. The year-to-year dynamics of the French manufacturing activity were varied; however the trend is bleak during this period. At the Hexagon level, we observe three years of near-zero growth (1995, 1998 and 1999) and only one year of significantly positive growth (2000). Manufacturing employment is declining all other years, most remarkably the year following the world financial crisis (2009).[7] From the beginning to the end of the period, disregarding breaks in statistical series, 954,209 jobs were lost in this activity (-26%). Finally, this figure shows that the average regional growth rate is very close to the hexagonal rate at any spatial scale, and that the smaller the spatial scale, the larger the Tukey fences.

[7] The sharp decline in growth observed in 2009 exists independently in each of the two datasets and is attributable to the consequences of the global financial crash.



Note: b1 and b2 are statistical breaks

Source: CLAP, INSEE, ADISP | Assédic

Figure 1. Hexagonal manufacturing employment growth rate and average regional shift over the period 1994-2015.

3.2. Methods

Our generalized SSA method help decompose properly the growth rates differential between two regions, but we also need a method to evaluate the contribution of each component to the overall regional inequality in regional growth rates.

Esteban (2000) suggests two such methods in his study of regional inequality in aggregate productivity per worker in Europe. In this study, the productivity shifts between the regions and a benchmark European region are decomposed using Esteban-Marquillas' SSA formulae (1972). Then, on the one hand, the author computes the relative weight of the variance of each SSA component in the overall observed variance of the productivity shifts. On the other hand, he compares several regression models, each model correlating one single component of the SSA to the productivity shifts. In a more recent study on regional inequality in labor productivity, (Mussini 2019) advocates the use of the Gini index, standard for inequality measurement, in place of variance. Le Gallo and Kamarianakis (2011) reject the variance for not being a typical measure of inequality, and extend Esteban's regression approach, applying space-time econometric models to study the evolution of regional productivity in Europe.

In our view, variance is indeed an imperfect measure of inequality, but a regression is not better in this respect. The Gini index is clearly more appropriate to the study of inequality for variables such as labor productivity. However, regional growth rates admit negative values, hence the use of the Gini is not a suitable option for our study.[8] Among the main limits of the variance to

[8] The Gini index is sometimes used for variables which can admit some negative values (usually few), such as negative wealth or negative income taxes. The most common and simple practices are either to eliminate the observations with negative values or convert them into zero, but these methods have

reflect regional growth inequality is the fact that it gives utmost importance to the few observations that differ most from the mean (due to its quadratic form). An alternative approach could be to use a robust scale estimator, which has the property to be less sensitive to extreme values and can be meaningfully computed on a distribution admitting negative values. The median absolute deviation about the median (MAD) is a member of this class of estimators. It is fairly insensitive to extreme values as it is based on absolute deviations rather than quadratic deviations.[9] Like the Gini and unlike the variance, it reflects the behavior of the data in the middle of the distribution rather than in its tails. Considering a batch x of N numbers $\{x_1, \dots, x_N\}$, the formula of the MAD is given by:

$$MAD(x) = b\ med_m|x_m - \quad (8) \quad x)|$$

where $med(x)$ is the median of x, $med_m|x_m - med(x)|$ is the median of the absolute deviations from $med(x)$, and the constant $b = 1.4826$ makes the MAD comparable to the standard deviation when x is drawn from the normal distribution and N is large (Rousseeuw and Croux 1993).

The formula of the empirical variance is:

$$var(x) = \frac{1}{N-1}\sum_{m=1}^{N}(x_m \quad (9) \quad ^2$$

important drawbacks (De Battisti, Porro, and Vernizzi 2019) and, in any case, obviously cannot be applied here. More appropriate methods have been proposed, notably by Raffinetti et al. (2015), but even in this case, "*it is important to remark that when dealing with negative values the standard Gini coefficient G is no longer a concentration measure: it can still be computed for comparing different distributions, but it can be interpreted just as a relative measure of variability with respect to the mean value*" (Battisti et al., 2019, p. 106).

[9] The Sn and Qn robust scale estimators proposed by Rousseeuw and Croux (1993) have the additional advantage not to be slanted toward symmetric distribution. We have contrasted the results obtained with the MAD with the ones obtained with these alternative estimators. These results are not presented due to lack of space and as no substantial differences appeared.

where $\bar{x}$ is the mean of x.

As proposed by Esteban (2000), the variance of the regional shift (between a region and a benchmark) can be decomposed into the variances of each of the SSA components and a total covariance term:

$$var\left(g_{t+1}^{j} - g_{t+1}^{b}\right) = var(\mu) + var(\pi) + var(\varepsilon) + var(\omega) + COVAR$$

$$COVAR = 2 \times [cov(\mu,\pi) + cov(\mu,\varepsilon) + cov(\mu,\omega) + cov(\pi,\varepsilon) + cov(\pi,\omega) + cov(\omega,\varepsilon)] \quad (10)$$

where g_{t+1}^{b} is the benchmark region's manufacturing employment growth rate, μ is the industry-mix component, π the competitive component, ω the outlier component and ε the emergence component of equation 7.

For ease of interpretation, we computed in addition two specific covariance terms:

$$COVAR1 = 2 \times cov(\mu,\pi)$$

$$COVAR2 = 2 \times cov(\omega \quad (11)$$

Another way of handling the variance decomposition approach in the SSA literature is by following Shorrock's rule, which commands to distribute each covariance term into its associated component term (Shorrocks 1982; Duro, Alcántara, and Padilla 2010):

$$var\left(g_{t+1}^{j} - g_{t+1}^{b}\right) = var(\mathrm{P}) + var(\mathrm{M}) + va \quad (12) \quad var(\mathrm{E})$$

$$var(\mathrm{M}) = var(\mu) + cov(\mu,\pi) + cov(\mu,\varepsilon) + cov(\mu,\omega)$$

$$var(\mathrm{P}) = var(\pi) + cov(\mu,\pi) + cov(\pi,\varepsilon) + cov(\pi,\omega)$$

$$var(\mathrm{E}) = var(\varepsilon) + cov(\pi,\varepsilon) + cov(\mu,\varepsilon) + cov(\omega,\varepsilon)$$

$$var(\Omega) = var(\omega) + cov(\pi,\omega) + cov(\mu,\omega) + cov(\omega,\varepsilon)$$

We thus propose, on a yearly basis over the period 1994-2015: (i) first to compute and decompose the variance of the regional growth based on our four SSA components (following

Shorrock's rule or not); and (ii) second to measure the MAD of the regional growth and the MAD of each one of the four components.[10]

Both analyses were performed after that regions with extreme values for manufacturing employment growth were identified each year (based on Tukey's fences, see Figure 1) and deleted (i.e. 9% of regions on average with at least 7% and at most 11% at the LAU 1 level). This clean-up was necessary for the variance analysis to be interpretable, but uncleaned results are also provided in appendix and are qualitatively unchanged for the MAD. This deletion approach could not reasonably be applied to the issue of sectoral outlier values, which affects all regions with many sectors candidate for deletion (on average at the LAU 1 level, a region would have lost 17% of its industrial sectors and more than 25% of them for the regions in the most affected quartile).

3.3. Results

3.3.1 Main result: Industry-mix vs Competitive effects

The main question addressed by the SSA is whether regional inequality in manufacturing employment growth rates is attributable mainly to competitive or industry-mix effects. Figure 2 provides the answer to this question (see Appendix 3 for a similar figure providing results uncorrected for regional outliers). The first two rows present the variance decomposition approach for each of three spatial scales, either displaying all variance and covariance terms

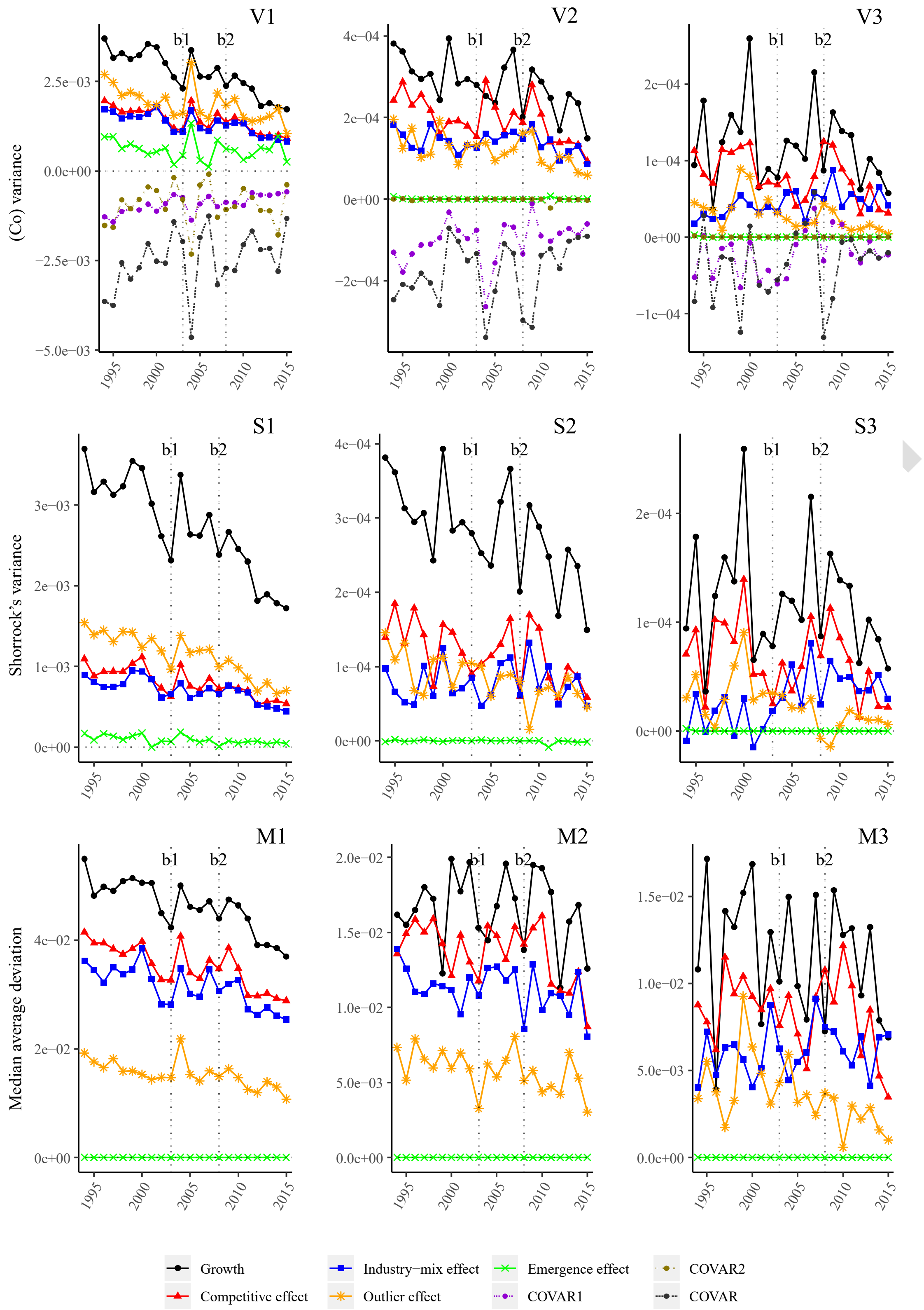
[10] A benchmark region is introduced in equations 10 and 12 to follow the usual practice in the SSA literature of decomposing growth differentials. However, $var\left(g_{t+1}^{j} - g_{t+1}^{b}\right) = var\left(g_{t+1}^{j}\right)$ as g_{t+1}^{b} Is a constant, and all the variance and covariance terms of equations 10 and 12 are unaffected by the benchmark. In the same way, $MAD\left(g_{t+1}^{j} - g_{t+1}^{b}\right) = MAD\left(g_{t+1}^{j}\right)$, and $MAD(\mu)$, $MAD(\pi)$, $MAD(\varepsilon)$ and $MAD(\omega)$ are unaffected by the benchmark.

(panels V1 to V3), or applying Shorrock's rule (panels S1 to S3). The third row presents the median average deviation (MAD) approach (panels M1 to M3).

For the whole period, whatever the spatial scales, the response is qualitatively unchanged: *it is the competitive effect that contributes most to regional growth inqualities, although differences between dispersion indicators for competitive and industry-mix effects are not large*, especially on a fine spatial scale (see V1, S1 and M1). This result prevails for almost all years with few exceptions (see the most recent years of V3, S3 and M3), and whether the measure of inequality is sensitive to extreme values (variance) or not (MAD). Besides, we can also observe in all panels that dispersion indicators for competitive and industry-mix effects show comparable decreasing trends, which result in a similar trend affecting regional growth inequalities.

To say more, the fact that differences between dispersion indicators for competitive and industry-mix effects tend to be smaller on a fine spatial scale makes the proof that spatial disaggregation bring additional information: it reveals local differences in industrial mix that are hidden at coarser spatial scales. Our novel SSA technique makes it possible to perform multi-scale analyses and thus observe more of the local economic heterogeneity. In addition, the interannual variability of the dispersion indicators appears to be much smaller at a fine spatial scale. This may stem from statistical efficiency: scale estimators are expected to converge to their underlying "true" values as sample sizes grow; fine spatial scales provide more observations, such that estimations are less contaminated by measurement errors and by chance (i.e. an erroneous or odd observation does not take the estimator value too far away from its correct value). Our scalable SSA allows exploiting this efficiency advantage associated with fine spatial scales. Of course, there are also more extreme sectoral growth events at fine spatial scales, but this outlier issue is well accounted for in our scalable SSA framework (see below).

The panels V also display the curve of $COVAR1$. Although the competitive effect and the industry-mix effect are mathematically independent components of the final equation (7), our empirical results show a negative correlation between the two over the period (-0.33 on average at the LAU 1 level, for example). The growth of some regions where business is doing well on average is in fact driven by small fast-growing sectors and not (or less) by larger ones, inducing a positive competive effect while the industry-mix effect is negative. Conversely, some regions with widespread business difficulties have large sectors that tend to fare better than the smaller ones, supporting once again a negative correlation between the two SSA components.



Note: b1 and b2 are statistical breaks, see section 3.1 for more details on data sources
Source: CLAP, INSEE, ADISP | Assédic

Figure 2. Dispersion indicators for regional growth and SSA components.

3.3.2 On ‘extra’ effects and the value added of our SSA framework

We have got our answer regarding the comparison between competitive and industry-mix effects. But was our novel decomposition framework necessary to make this analysis, and to what extent does it bring additional added value compared to former methods?

First and foremost, it was indeed necessary to use a method that acknowledges and admits sectoral absences. Identifying a shared set of manufacturing sectors among all LAU 1 regions was not a reasonable option and neither was it reasonable to delete all regions featuring missing sectors. In 2015, for instance, 525 LAU 1 regions among 1988 (26%) had missing sectors. Over the period 1994-2015, no region is left untouched. It is then simply impossible to use former methods without excessively reducing the regional and sectoral scope of the analysis to meet a common economic structure across regions. This sectoral absence issue, although more critical at the LAU 1 level, also exists at NUT 2 and 3 levels, as already stressed in the introduction of the paper.

Secondly, a value-added of our decomposition framework is that it allowed to account for the growth effect of sectoral emergence at the regional level. As we can see on panels V and S of the Figure 2, sectoral emergence contributes significantly to regional growth inequality as measured by variance. In 2014, its variance is even larger than the one of the industry-mix effect. However, the panels M of the Figure 2 shows that its contribution to regional inequality can be ignored when the analysis focuses on the middle of the distribution rather than on extremes.

Last but not least, applying our method was important for the competitive and industry-mix effects not to be strongly biased by sectoral growth outliers, especially in the variance analysis (see Appendix 4 for a demonstration of this point and some descriptive statistics on sectoral outliers). In our framework, the effect of tremendous growth events affecting regional industries is isolated in a specific component. Panels V and S of the Figure 2 show that the variance term of this sectoral outlier effect is non-negligible and even always larger than the ones of the competitive and industry-mix effects. Panels M of the Figure 2 shows that the MAD of the sectoral outlier effect is also non-negligible although always smaller than the MADs of the competitive and industry-mix effects. Logically, tremendous sectoral growth events contribute significantly to regional growth inequality, and it is even more obvious when attention is focused on distribution tails (i.e. variance).[11]

3.3.3 On the spatial dimension of growth

"*Everything is related to everything else, but near things are more related than distant things*" (Tobler 1979) is commonly defined as the first law of geography. It has inspired a great deal of research in regional science, particularly in the study of the spatial autocorrelation (Cliff and Ord 1969), which is characterized as a lack of statistical independence between the observed value of a variable in a regional unit and its values in the surrounding regions. Consequently, it is not surprising that recent SSA enhancements have also focused on the spatial issue. It has been addressed, for instance, by isolating the neighborhood influence on regional growth from region-specific explanatory components within spatially structured employment growth

[11] Panels V of the Figure 2 also shows that $COVAR2$, i.e. the covariance between this outlier effect and the emergence effect is negative and comparable in magnitude to $COVAR1$ (the average correlation coefficient over the period is -0.42 at the LAU1 level, for example): the regions where new sectors are emerging are generally not those that are experiencing exceptional growth events. The addition of $COVAR1$ and $COVAR2$ almost amounts to the total covariance term.

decomposition frameworks (Nazara and Hewings 2004; Mayor and López 2009), or by breaking down shift-share component contributions to regional inequality of productivity between neighboring and non-neighboring regions (Mussini, 2019), or - in an econometric approach to the shift-share decomposition of productivity - by correcting for spatial autocorrelation biases (Le Gallo and Kamarianakis 2011).

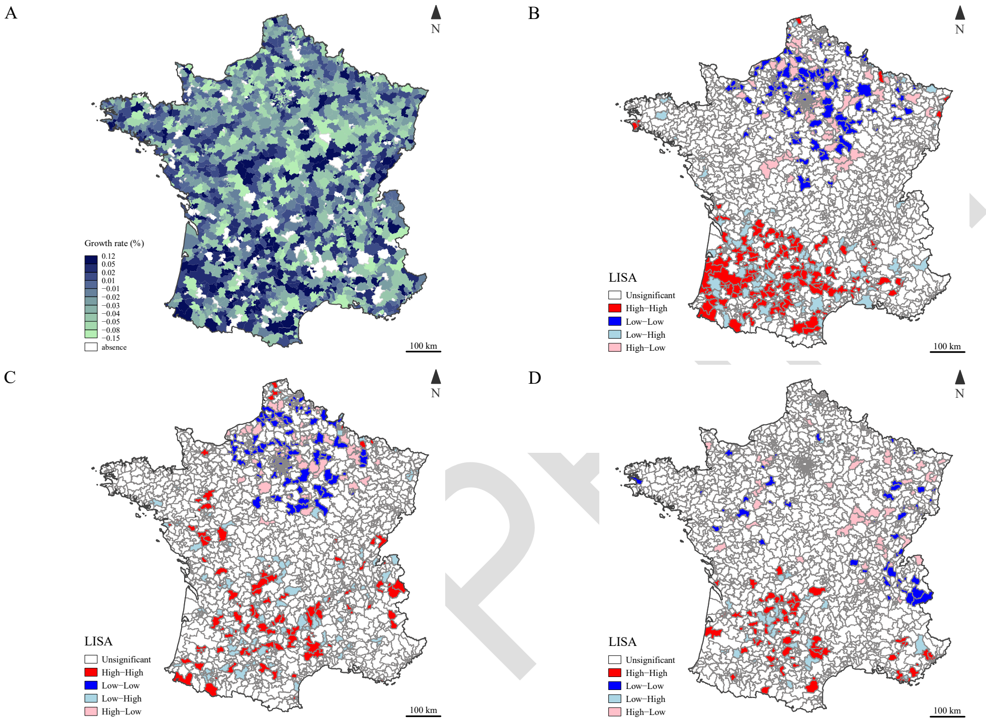
Taking into account the spatial dimension raises new methodological issues: "*a topological concept of spatial proximity must be arbitrarily introduced by researchers*" (Espa et al., 2014, p. 117)**,** and additional data are sometimes required. In practice, spatial weight matrices must be defined and can be so either based on simple contiguity notions, or considering the length of shared common boundaries, or based on between-centroïd Euclidian or gravity-based distances (Nazara and Hewings 2004; Mitchell and Carlson 2005; Mayor and López 2009; Espa et al. 2014), or even based on socio-economic distance variables such as migration or trade flows (Patuelli et al. 2006; Zaccomer 2006; Zaccomer and Mason 2011).

The current paper does not aim to introduce novelty in the treatment of the spatial issue. Nevertheless, an explanatory analysis of global and local spatial autocorrelations was carried out on regional growth rates and on the components of our SSA decomposition. Our expectations were twofold. First, in accordance with Tobler's law, we expected to find a global spatial autocorrelation of regional growth and, more importantly, a positive one. Second, we expected a positive and strong global spatial autocorrelation of competitive effects. Indeed, the competitive effect reflects the (dis)advantages that affect all industries in a region, and it is therefore likely to be based on (dis)advantageous place-based characteristics (e.g. market size, land costs, human and social capitals, institutions, infrastructures, bundles of natural amenities, etc., see for instance Ketterer and Rodríguez-Pose, 2018) that do not have to be strictly limited to a fine geographical scale such as LAU 1. In addition, positive spillovers can be expected between regions endowed with such economic (dis)advantages and neighboring regions. Specific regional industries may also benefit from location-specific characteristics available at coarse geographical scales (e.g. a large forestry massif for the wood industry), potentially creating spatial autocorrelations in industry-mix effects. But the effects affecting different industries could cancel each other out, so we should more likely find local (rather than global) spatial autocorrelations in this case.

To explore these hypotheses, we modeled spatial interactions based on Euclidian distances and critical cut-off boundaries at the smallest spatial scale available (i.e. the LAU 1 level): for each region, the neighbors are the regions within a radius of - alternately - 50, 100, 150 and 200 kilometers. We then first computed the Moran's *I* statistics (Cliff and Ord 1969), which is the standard measure of global autocorrelation in regional science, for each year of our 22-year study period. Table 1 presents summarized results regarding the number of years with a significant global autocorrelation and the values of the significant indices.[12] Second, Anselin's *LISA* (Local Indicator of Spatial Association) is commonly used as a local equivalent of Moran's index (Anselin 1995). For each observation and a given regional variable, the *LISA* indicates the extent of the association between the value observed in this observation and the values observed in its neighborhood. If the *LISA* is significant, it can either be a positive association (neighborhood values are of the same kind, high-high or low-low) or a negative association (neighborhood values are negatively correlated, high-low or low-high). Table 2 presents summarized results regarding the number of regions with significant local

[12] Detailed results are not shown here for space reason but are available upon request to the authors.

autocorrelation over the whole period, and Figure 3 displays maps of the clusters of significant local spatial associations, as well as a map of regional growth, for the year 2010.[13]



Source: CLAP, INSEE, ADISP | Assédic | IGN

Figure 3. Maps of the regional growth and local indicators of spatial association (LISA) in 2010.

[13] Detailed results are available upon request to the authors. A cut-off distance of 150 km is retained for the local analysis as the global analysis shows that the spatial relationships are stronger than for other distances (significant *I* are larger). The year 2010 was chosen as an illustration for the map, as it is the post-crisis year with the highest level of local spatial autocorrelation (i.e. larger *LISA*) for employment growth.

Table 1. Global spatial autocorrelation (Moran's *I* statistics) over the 22-year period

Regional variable	Number of years with significant *I* (average value of significant *I*) Critical cut-off distances:			
	50 km	100 km	150 km	200 km
Growth rate	19 (5.12E^{-3})	19 (1.43E^{-2})	20 (1.57E^{-2})	21 (1.06E^{-2})
Competitive effect	14 (3.08E^{-3})	19 (1.26E^{-2})	17 (1.65E^{-2})	18 (8.09E^{-3})
Industry-mix effect	8 (1.07E^{-3})	3 (4.96E^{-3})	3 (7.60E^{-3})	4 (3.10E^{-3})
Emergence effect	5 (2.79E^{-3})	4 (4.18E^{-3})	4 (7.72E^{-3})	4 (4.05E^{-3})
Outlier effect	4 ($9.20\ \text{E}^{-4}$)	2 (5.27E^{-3})	2 (8.21E^{-3})	4 (2.66E^{-3})

Note: All significant Moran's *I* have positive values; p-values ≤ 0.05 are considered statistically significant; p-values (two-tailed) are calculated for the randomization null hypothesis' test; R package lctools has been used.

Table 2. Local indicator of spatial association (LISA) between LAU 1 regions over the 22-year period

Regional variable	Average number of regions with significant *LISA* (average proportion among all regions)		
	Positive	Negative	Total
Growth rate	164.7	117.7	282.4 (15.6%)
Competitive effect	151.5	107.4	258.9 (14.3%)
Industry-mix effect	59.4	53.1	112.5 (6.2%)
Emergence effect	25.0	21.6	46.6 (2.6%)
Outlier effect	50.8	45.0	95.9 (5.3%)

Note: A cut-off distance of 150 km is retained and p-values ≤ 0.05 are considered statistically significant; p-values (two-tailed) calculated for the randomization null hypothesis' test; R package lctools has been used.

The results are striking and consistent with our expectations. With regard to global spatial autocorrelations (Table 1), regional growth was indeed positively autocorrelated in most years, as was the competitive effect. In contrast, global spatial relationships were much less often significant for the other effects. Spatial relationships appear always stronger (i.e. larger values of significant *I*) with the 150 km cut-off distance. At the local level (Table 2), we found both positive and negative significant relationships, although the positive ones still dominate. On average, we found again that regional growth and the competitive effect were much more spatially autocorrelated than the industry-mix effect, let alone the emergence and outlier effects. The illustrative map for the year 2010 (Figure 3) clearly shows regional spatial interactions and place-based dependencies. On panel B, a large dynamic southwest region featuring numerous high-high spatial relationships for growth stands out from a northern region with low-low spatial autocorrelations, and the negative high-low and low-high spatial relationships correspond to regions located close to the latter but with opposed profiles. As for competitive effects (panel C), the high-high relationships are less numerous and more scattered in the South, and the low-low clusters are more concentrated in the North. The last map (panel D) shows that local spatial associations also exist for the industry-mix effect, but are less common (in line with *I* being not significant for this effect in 2010). Overall, this analysis shows that regional growth inequality is a multi-level issue, which cannot be considered only at a fine spatial scale, but is also structured at larger regional scales.

4. Discussion and concluding comments

Every year, many French regions are concerned by sectoral absences, industry emergence and tremendous growth events. There are very few regions, if any, that do not experience them over decades. As in Krugman's theoretical world, industries do not spread everywhere, but tend to cluster due to the combination of cost incentives, location advantages and path dependencies. Besides, in line with some Schumpeterian insights, the appearance/destruction of regional economic activities matter.

The SSA framework, popular for decomposing regional growth, fails to adequately address such important issues for regional economics. The challenge is so serious that it is currently only possible to implement the SSA on highly aggregated sectoral and spatial data.

Our novel SSA differ from original frameworks in that it becomes fully scalable spatially and sectorally by including the possible absence, emergence or extreme growth events affecting some business sectors. It follows the paradigm shift suggested by Artige and van Neuss (2014): rather than answering the question of whether a regional industry-mix is beneficial in view of national trends in sectoral growth, and whether a region is growing faster than expected in view of its industrial mix - thus disentangling presumed “national forces” and “local deviations” from them - it now focuses exclusively on effective regional forces. The renewed SSA industry-mix and competive growth components reflect two regional economic levers, namely the structural strenghts and the common local assets: a well-specialized region is one where the major regional industries are particularly dynamic and boost the local growth, and a region is considered competitive *per se* when all industries across the economic structure show favorable growth performances. Put differently, the industry-mix effect reveals an economic specialization that is really adapted to local capitals, whether physical, social, environmental or cultural, and the competitive effects shows the extent to which these capitals and the local business ecosystem in general benefit to all local industrial players.

This change converges with a similar paradigm shift in regional development policy making. There is no longer an attempt to shape regional economies in order to host presumed global or national industrial forces, but regions’ specific assets and capabilities are now presumed to be the driving force behind innovation, growth and successful economic specialization. For example, the 2014-2020 European Regional Innovation Strategy based on Smart Specialization Strategy (RIS3) supports this view by aiming for a concentration of investments in local initiatives (e.g. excellence poles, business incubators…) and the sectoral business assets of regional economies, promoting the capitalization on regional competitive advantages, fostering a regional economy of distinction and excellence, encouraging innovation and economic transformation by regions. This policy is part of a place-based regional development strategy, in which economic activities are recognized to be linked to place identity. The role of the regional development policy is thus to help identify specific strengths, stimulate cooperative processes and combine the "hard" location factors for industrial production (existing industrial core, access to resources and transport infrastructure, links to markets, etc.) with "soft" factors (skilled labor force and educational institutions, administrative support for business, etc.) to promote sustainable growth (Alessandrini, Celotti, and Dallhammer 2019).

From this region-centered perspective, the results obtained with regard to inequalities in French manufacturing employment growth over the period 1994-2015 highlight that *competitive effect contributed the most to regional growth inqualities, although competitive* and industry-mix effects were fairly close to each other at the smallest spatial scale (*i.e.* LAU 1 regions) Both SSA effects appear negatively correlated: regions that are doing well overall generally have small sectors that are doing particularly well (e.g. start-ups or catching-up sectors), and regions with widespread difficulties have large sectors that tend to fare better (i.e. a kind of economic shock absorbers). Industry emergence also contributes significantly to regional growth inequality when importance is given to distribution tails (i.e. the gap between the most leading and lagging regions). As one would expect, sectoral outliers (booming or collapsing industries) contribute even more strongly to inequality. Besides, isolating this last component was necessary to make the competitive and industry-mix effects properly interpretable. Regional growth rates and competitive effects are also shown to be positively spatially autocorrelated: dynamic/competitive regions tend to be located nearby dynamic/competitive regions, and non-dynamic/non-competitive regions nearby non-dynamic/non-competitive regions. Locally, however, we can see that both positive and negative relationships exist, and that the significant spatial relationships occur in specific parts of the national territory.

Due to the paradigmatic shift mentioned above, it is difficult to directly compare our work with previous one. The literature on French manufacturing employment growth that applies SSA decomposition frameworks (or econometric analogs to these methods) is sparse and based on the national-structural vs geographic divide. Besides, their focus is not on analyzing the regional growth inequality as in our case, but to compare average national-structural effects with geographic effects, either for all French administrative regions or employment areas (Carré et al., 2019; Kubrak, 2018, Levratto and Carré, 2013), or for urban vs rural areas (Gaigné et al., 2005; Abildtrup et al., 2018). In these papers, the geographic component always dominates the structural component. The industrial decline appears stronger in urban and former industrial regions. The sprawl of production out of urban centers lead to a negative geographic effect in urban areas, unlike in rural and periurban areas where it is generally positive.

Beyond the French case, the literature based on (deterministic or econometric) national-structural vs geographic decompositions recurringly finds that the geographic context is the main explanatory component, e.g. Blien and Wolf (2002) for regional employment growth, Dinc and Haynes (2005) for regional productivity, Esteban (2000), Benito and Ezcurra (2005) or Le Gallo and Kamarianakis (2011) for regional productivity inequalities. The two applications of Artige and Van Neuss's technique on regional employment growth find on average larger competitive effects than industry-mix effects (Artige and van Neuss 2014; Martin et al. 2016). However, they do not make an analysis of regional inequalities. As such, our approach sheds new light on this issue: with the lens of a novel region-centered method allowing to work at various geographical scales dynamically, it reveals a limited gap between the contribution to inequality of the competitive effect and the industry-mix effect, although the former tends to dominate the latter, especially at coarse spatial scales (NUTS 2 and 3). The competitive effect is certainly greater at these scales because spatial aggregation hides the existence of significant differences in industrial mix at a more local scale.

We can draw some lessons for the design of French regional policies. Our findings may provide general insights, or even region-specific ones, about the growth drivers to be possibly prioritized for support. A classic concern of place-based regional policies is to adequately target investments, either on a few key sectors, or on the business ecosystem as a whole. Considering that regional competitiveness and industry mix could have a fairly equal importance in regional growth inequality, an equal attention should be paid to the two policy options. Also, in small regions, emerging and booming sectors are likely to have a significant impact on the economy and could be targeted. In practice, "*there is no one-size-fits-all approach. Local specificity is both inevitable and desirable to develop the strength of the industry within a region*" (Alessandrini et al., 2019, p. 17). A policy that aims to reduce the spatial inequality, or at least to support 'distributed development' (Iammarino 2019), would benefit from identifying the specific weaknesses of the regions lagging behind. In regions characterized by both an overall lack of competitiveness and a rapid job decline in major industries, improving the business ecosystem is undoubtedly of primary importance, and seeking the emergence of new industries might be a way forward. In those affected by a similar decline in large industries, but where the competitiveness is not that bad, it might be worth investing massively on the small dynamic sectors that have already emerged. Lastly, in regions where large industries are the main factor of dynamism and resilience, it may be highly advisable to support them, while in the same time seeking for fresh industrial opportunities.

To conclude, our SSA approach has been shown here to be both tractable and relevant for the dynamic analysis of employment data broken down according to fine sectoral and territorial classifications. Although our results appeared quite robust to changes in nomenclature,[14] this approach may in general provide varying results according to the sectoral/geographical breakdowns retained, because part of the competitive effect at a given aggregation level may be an industry-mix effect at a more detailed one. Consequently, the choice of economic and geographic structures to compute the SSA is not trivial and should be considered more seriously in future research.

This novel framework opens up many opportunities for research enhancements. Allowing the decomposition of the growth gap between two regions, it can serve as a basis for unprecedented interregional comparisons, such as Paris versus London, for example. Meaningful similarity matrices and clusters could be built from multi-region shift-share comparisons. Our SSA components could also be used in space-time econometric modeling, as the absence of a national reference in the framework makes the control of spatial and temporal autocorrelations even more relevant. Finally, it would be interesting to use it on variables other than employment, with the possibility of renewing, for example, work on labor productivity, by focusing on productivity evolution rather than on static annual productivity data.

[14] Further research was carried out on the same data set, varying either the regional aggregation (NUTS 3) or the sectoral aggregation (95, 230 or 259 sectors). In any cases, the regional competitive effect and the industry-mix effect contribute fairly equally to regional growth inequality, but the importance of the industry-mix effect tends to increase with the degree of disaggregation, up to be slightly greater than the competitive effect.