Distributional changes in basic services accessibility

Ambra Poggi
University of Milan Bicocca, Italy
and LABORatorio R. Revelli, Collegio Carlo Alberto, Italy

Very preliminary version : April 8, 2015

Abstract

We explore distributional changes in basic services accessibility (public transport, postal services, banking services, grocery services and primary health service) across EU regions over the period 2007-2012. We investigate distinct facets of distributional changes. On one hand, we study σ-convergence as the change in the Gini coefficient over time expressing this change as the net effect of β-convergence when offset by re-ranking among EU regions. On the other hand, we analyze the surfacing of two-peaks in the 2012 distribution using stochastic kernels since the emerging of a two-peaks distribution has a natural interpretation in terms of polarization. We find evidence of divergence since re-ranking offsets the substantial reduction in inequality induced by β-convergence between 2007 and 2012. We also find evidence of a polarization process: regions are polarizing into low-levels services and high-levels services groups. Regions catch up with one another but only within particular subgroups.

Keywords: services accessibility, inequality, re-ranking, polarization, kernel density

JEL: I30, D3, O47
1. Introduction

Access to basic services (public transport, postal services, banking services, grocery services and primary health services) is important as it enhances individuals’ opportunities and living conditions. For example, access to public transport enhances individuals’ access to jobs, which is a key to exit from poverty. Even if differences in access to basic services exist across EU regions (Lelkes and Gasior, 2012; Poggi et al, forthcoming), we expect these differences reduce over time as consequence of regional policies aimed to equalize individuals’ opportunities and living conditions across the EU.

In this paper, we analyze distributional changes in services accessibility across EU regions over the period 2007-2012. In particular, we investigate the distribution dynamics exploring four distinct facets of distributional changes: σ-convergence, β-convergence, re-ranking and polarization. We investigate σ-convergence as the change in the Gini coefficient over time and use the exact additive decomposition suggested by Jenkins and Van Kerm (2006) to express this change as the net effect of β-convergence when offset by re-ranking among EU regions (O’Neill and Van Kerm, 2008). Then, we investigate the surfacing of two-peaks in the 2012 distribution using stochastic kernels. The emerging of a two-peaks distribution has a natural interpretation in terms of polarization: those portions of the underlying population of regions collecting in the different peaks may be said to be polarized, one group versus another.¹

We contribute to the literature offering fresh empirical evidence about distributional changes in services accessibility across EU region during a period (2007-2012) characterized by a deep economic crises. We find evidence of divergence in services accessibility across EU regions since re-ranking offsets the substantial reduction in inequality induced by β-convergence between 2007 and 2012. We also find evidence of a polarization process: regions are polarizing into low-levels services and high-levels services groups. Regions catch up with one another but only within particular subgroups.

Section 2 describes the empirical methods used in the paper reviewing the main literature on distributional analysis. In section 3, we illustrates the data and our measure of services accessibility. Empirical findings are presented in section 4. Section 5 concludes.

2. Methods for studying distributional changes across regions

Measuring distributional dynamics presents some complexities since distinct facets of distributional changes need to be considered.

We can investigate whether all regions are assumed to converge towards the same steady-state. In other words, we can investigate the concept of β-convergence, that is a process in which disadvantaged regions improve faster than advantages ones and therefore catch up on them (Sala-i-Martin, 1996a, Barro and Sala-i-Martin, 1992; Mankiw et al., 1992). However, several authors (Friedman, 1992; Quah, 1993) have argued that that β-convergence tell us little about whether dispersion across countries has fallen: it is possible to observe disadvantaged regions improving faster than advantaged regions and yet observing divergence. For this to happen it must be the case that initially disadvantaged regions overtake the advantaged regions, so that the rankings of regions changes. In this case, we should investigate the concept of σ-convergence (that is if the dispersion tends to decrease over time) testing whether a consistent diminution of variance among regions is observed (Friedman, 1992). In reality, both concepts of convergence are interesting and should be analysed empirically.

¹ The notion of convergence clubs (that is of areas catching up with one another but only within paricular subgroups is also apposite (Baumol, 1986)
O’Neill and Van Kerm (2008) propose an integrated framework for studying convergence that incorporates traditional measure of β-convergence and σ-convergence based on the fact that these concepts are formally linked by a measure of re-ranking. They allow for non-linearities in the improvement process and explicitly identifies the contribution of faster improvements among disadvantaged regions to reduction in overall inequality. Adopting the Gini index as measure of inequality, they decompose changes in inequality across regions over time (from period 1 to period 2) as follow:

\[
\Delta G = G_t^2 - G_t^1 = (G_t^2 - C_t^2) - (G_t^1 - C_t^1) = R - BC
\]

where \(G_t^i\) is the Gini coefficient of period \(t\) and \(C_t^i\) is the concentration coefficient computed according to the ranking observed in period 1 and assigning the value registered in each region at time 2. The left-hand side provides measure of σ-convergence: \(\Delta G > 0\) corresponds to rising inequality (divergence) while \(\Delta G < 0\) corresponds to falling inequality (convergence). Equation 1 decompose this change over two parts, \(R\) and \(BC\). The second term, \(BC\), captures the extent to which faster improvements among disadvantaged regions reduces inequality and it has a clear interpretation as distributive measure of β-convergence. As argued by Friedman (1992) the effect of β-convergence on inequality is mitigated by re-ranking of regions, which occurs if initially disadvantaged regions overtake advantaged regions. \(R\) measures this effect. Thus, O’Neill and Van Kerm (2008) presenting the change in inequality in this way allows us to identify easily the relative contribution of both re-ranking and progressive improvements (β-convergence) to the overall change in inequality (σ-convergence). In this paper, we apply this decomposition to investigate an eventual process of convergence in services accessibility across EU regions (due to policies/investments aimed to improve access to basic services in disadvantaged areas). Note that we use the normalization of the Gini index proposed by Raffinetti et al. (2014) to allow for negative values in the distribution of services accessibility.

Let \(N\) be the number of considered regions and \(y_i\) the level of services accessibility in region \(i\), the normalized Gini index, \(G_p\), can be written as

\[
G_p = \frac{\mu}{\mu_p} G
\]

where \(G\) is the standard Gini coefficient, \(\mu\) is the average of \(Y\), \(\mu_p\) is normalization term defined as

\[
\mu_p = \frac{([N-1]/N^2)(T^*+T^-)}{(N-1)/N}
\]

with \(T^* = \sum_{i=1}^{N} \max(0, y_i)\) and \(T^- = |\sum_{i=1}^{N} \min(0, y_i)|\). The normalization proposed by Raffinetti et al. (2014) is also applied to the concentration coefficient.

The above framework has the advantage to be a more parsimonious representation of distributional change than full-scale analysis of the joint distributions. However, it fails to identify polarization process as the emerging of a two peaks distribution. In other words, it is no informative on processes where regions catch up with one another but only within particular subgroups. Quah (1997) suggests to study the dynamic of cross-section distribution of regions using stochastic kernels to better shed light on intra-distribution dynamics. The focus is on the distribution across regions to find uncovered empirical regularities (as emerging of two peaks) that are hidden to traditional methods of empirical analysis. Let \(\lambda_t\) be the probability measure (one each year) associated with the cross sectional distribution \(F_t\). A simple way to modelling the dynamics is using the following probability model

\[
\frac{\text{Kalaitzidakis et al. (2000) and Fiaschi and Lavezzi (2003) critize earlier worker on β-convergence for failing to allow for non-linearities in the growth process.}}{\text{Raffinetti et al. (2014) introduces a reformulation of the Gini coefficient with a normalization that overcomes some drawbacks inherent in the normalization suggested by Berrebi and Silber (1985).}}
\]
\[ \lambda_t = T^*(\lambda_{t-1}, u_t) \]  
(4)

where \( T^* \) is an operator mapping \( \lambda_{t-1} \) together with disturbances \( u_t \) into \( \lambda_t \). Equation 4 is analogous to the first-order auto regression model in time series and can be rewritten as

\[ \lambda_t = T_{ut}^* \lambda_{t-1} \]  
(5)

where \( T_{ut}^* \) absorbs the disturbances into the operator itself. The stochastic kernel (and its related contour plot) is a graphical representation of the transition probabilities \( T^* \) and therefore maps \( \lambda_{t-1} \) into \( \lambda_t \), tracking where \( F_t \) points in \( F_{t-1} \) end up, then encoding information on intra-distribution dynamics.

Since stochastic kernels are graphs showing how the cross-sectional distribution at time \( t-1 \) evolves into that at \( t \), they need to be interpreted. If most of the graph are concentrated along the 45-degree diagonal then elements in the distribution remain where they begin. If, by contrast, most of the graph mass in the graph are rotated 90 degrees counter-clockwise from the 45-degree diagonal then substantial overtaking occurs. If most of the mass located parallel to the period \( t-1 \) axis, with projections on period \( t-1 \) values equal to each other, then kernel is one a single iteration takes any initial distribution to the same long-run distribution. If a two peaks sitting on the 45-degree diagonal emerges, these peaks can be called “basins of attraction” (Durlauf and Johnson, 1995).

3. Data and services accessibility

We explore the access to basic services, as assessed by the individuals themselves. We use data from the European Community Statistics on Income and Living conditions (EU-SILC), a survey that provides comparable cross-sectional and longitudinal micro data mainly referred to objective living (including income, poverty and deprivation) and employment conditions. In particular, we use 2007 and 2012 EU-SILC cross-sectional data in order to analyse information contained in a special module on “housing conditions” that was carried out for the first time in 2007 and repeated in 2012 on a different sample of households.

Individuals are asked about accessibility to specific basic services (on scale 1-4): public transport, postal services, banking services, grocery services and primary health services.\(^4\) The accessibility of the services is assessed in terms of physical and technical access, and opening hours (i.e. appropriate timetable), but not in terms of quality, price and similar aspects. Consequently, the access should refer to an objective and physical reality. The accessibility is considered at the level of the household, the difficulty to access is evaluated for the household as a whole. See Table 1 for descriptive statistics.

We take advantage of having household data measuring services accessibility at household level. We use factor analysis as a dimension reducing strategy. Factor analysis is a statistical data reduction technique used to explain variability among observed random variables in terms of fewer unobserved random variables called factors. In general, with the factor analysis, we model the observed variables as linear combinations of the factors, plus “error” terms. Thus, we are able to reduce data relating to different attributes down to just a few important dimensions. This reduction is possible because the attributes are related. The algorithm produces a factor structure matrix representing the correlations between the

\(^4\) Grocery services: services which can provide most of the daily needs. Banking services: withdraw cash, transfer money and pay bills. Postal services: send and receive ordinary and parcel post. Public transport: bus, metro, tram and similar. Primary health care services: general practitioner, primary health centre or similar
variables and the factors and is called the factor loading matrix. The interpretation of each factor is marked by high loadings on a certain sub-sample of attributes that give information on a specific kind of unobservable. As we extract consecutive factors, they account for less and less variability. The decision of when to stop extracting factors depends on when there is only very little “random” variability left. We retain only factors which account for sufficient variance: meaning that unless a factor extracts at least as much as the equivalent of one original variable, we do not consider it (Kaiser criterion). Since factor analysis is based on a correlation matrix, it assumes that the observed variables are measured continuously, are distributed normally, and that the association among indicators is linear. Since our observed variables are ordinal variables, we assume that they are indicators of underlying continuous unobserved variables and we use the polychoric correlation in the factor analysis as appropriate. Table 2 reports the results of the factor analysis: we identify only one factor (labeled services accessibility) which explains the 68% of the total variance. The Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) reports a value of about 0.9 thus confirming that the variables have enough in common to run a factor analysis. The factor score has zero mean and variance of one by construction. Values below zero indicate limitations in services accessibility respect the average.

Although the EU-SILC questions about access to services are not meant to explore subjective feelings, expectations and reference points may play a role at the individual level. Some individuals are perpetual optimists (they regularly overestimated the services accessibility) while others are perpetual pessimists (they regularly underestimated the access to services). However, information on services accessibility obtained by individuals are aggregate at regional level: the level of services accessibility registered in a certain region is the average level assessed by individuals living in the area. The bias in optimists should cancel out the bias in the pessimists since in aggregate individuals expectations are rational. In other words, we assume that informational anchors and bias of individuals are randomly distributed. The geographical breakdown used in the paper is defined according to the NUTS-1 regional classification and the degree of urbanization (rural, intermediate and urban). The choice is mainly related by practical considerations on data availability. Our final sample is composed by 134 regions covering 17 countries.  

4. Results

We now apply the decomposition presented in Equation (1) to study the change in basic services accessibility in the EU during the period 2007-2012. Table 3 reports the normalized Gini coefficient for the distribution of services accessibility at 5-year interval. We see an increase of dispersion over the period 2007-2012: the normalize Gini coefficient rises from 0.659 to 0.685. This result indicates that access to basic services slightly diverge across the EU regions over the period.

Table 3 also report the contributions of progressive improvements (BC) and re-ranking (R) to the change in overall inequality coefficient. We observe substantial progressive improvements (β–convergence) indicating that improvements in services accessibility tends to be higher in the most disadvantaged regions. However, re-ranking offsets the substantial reduction in inequality induced by β-convergence between 2007 and 2012. As result, progressive improvements have no equalizing effect.

The general picture emerging above analyzing the conditional distribution of services accessibility is not so comforting as it describes a process in which all regions are slightly diverging. Looking to the univariate kernel densities for services accessibility in 2007 and 2012 seems to confirm important distributional

5 AT, CY, CZ, DK, EE, ES, FI, FR, GR, HU, IS, IT, LT, LU, LV, PL, SK.
changes over the period (Figure 2). In 2007, the kernel exhibits a somewhat unequal level of services accessibility across EU regions. Five year later, a two-peak structure emerges.

The above picture is not a complete description of what goes on during the period. It fails to uncover the strong degree of polarization that continues to exist among the EU regions. Let us see what the distribution dynamics approach tell us about the divergence process. Figure 2 shows the bivariate stochastic kernel estimated. From the three-dimensional surface plot, we observe the existence of two local maxima and sizeable dip among them. In particular, we observe a dip around the average level of services accessibility (around zero by construction) and two local maxima respectively in the disadvantaged and the advantaged parts of the services accessibility range. This is made even more evident from the counter plot (Figure 3): both the peaks and the dip in the stochastic kernel lay approximately on the 45-degree diagonal. The density mass surrounding the peaks is steeper than the main diagonal while the density mass corresponding to the dip is somewhat flatter, suggesting that the two peaks act as “basins of attraction” for abutting observations. In other words, regions are polarizing into low-levels services and high-levels services groups. Moreover, the low-services peak is slightly above the 45-degree diagonal while the high-services peak is slightly below indicating that the distance between the two peaks exhibited a tendency to increase over the 1996-2002 period.

4 Conclusions

(to add)
References


Fiaschi and Lavezzi (2003)


Kalaitzidakis et al. (2000) and


### Table 1. Descriptive Statistics (household level)

<table>
<thead>
<tr>
<th>Variable</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>of grocery services (1=with great difficulty; 4=very easily)</td>
<td>2.24</td>
</tr>
<tr>
<td>of banking services (1=with great difficulty; 4=very easily)</td>
<td>2.05</td>
</tr>
<tr>
<td>of postal services (1=with great difficulty; 4=very easily)</td>
<td>2.02</td>
</tr>
<tr>
<td>of public transport (1=with great difficulty; 4=very easily)</td>
<td>2.05</td>
</tr>
<tr>
<td>of primary health (1=with great difficulty; 4=very easily)</td>
<td>2.02</td>
</tr>
<tr>
<td>age (household respondent)</td>
<td>53</td>
</tr>
<tr>
<td>No. Households</td>
<td>224277</td>
</tr>
</tbody>
</table>

### Table 2. Factor analysis

<table>
<thead>
<tr>
<th>Accessibility</th>
<th>Services</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>of grocery services (1=with great difficulty; 4=very easily)</td>
<td>0.8396</td>
<td>0.2951</td>
</tr>
<tr>
<td>of banking services (1=with great difficulty; 4=very easily)</td>
<td>0.8565</td>
<td>0.2664</td>
</tr>
<tr>
<td>of postal services (1=with great difficulty; 4=very easily)</td>
<td>0.8219</td>
<td>0.3245</td>
</tr>
<tr>
<td>of public transport (1=with great difficulty; 4=very easily)</td>
<td>0.7637</td>
<td>0.4167</td>
</tr>
<tr>
<td>of primary health (1=with great difficulty; 4=very easily)</td>
<td>0.8304</td>
<td>0.3104</td>
</tr>
<tr>
<td>proportion explained</td>
<td>0.6774</td>
<td></td>
</tr>
<tr>
<td>Kaiser-Meyer-Olkin measure of sampling adequacy</td>
<td>0.87</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3. Convergence over the period 2007-2012

<table>
<thead>
<tr>
<th></th>
<th>Services accessibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial normilized Gini</td>
<td>0.659</td>
</tr>
<tr>
<td>Final normilized Gini</td>
<td>0.685</td>
</tr>
<tr>
<td>Change</td>
<td>0.026</td>
</tr>
<tr>
<td>R-component</td>
<td>0.182</td>
</tr>
<tr>
<td>BC-component</td>
<td>0.156</td>
</tr>
<tr>
<td>No. Regions</td>
<td>134</td>
</tr>
</tbody>
</table>
Figure 1. Services accessibility: unidimensional distributions

Figure 3. Stochastic kernel - services accessibility
Figure 4. Contour plot - services accessibility