A NONPARAMETRIC GINI CONCENTRATION TEST FOR LABOUR MARKET ANALYSIS

Chiara Gigliarano, Francesco Maria Chelli

1. Introduction

The Gini index is one of the most important statistical indices employed in social sciences for measuring concentration in the distribution of a positive random variable; it is mainly used in economics as a measure of income or wealth inequality among individuals or households (see, e.g., Gini 1912, 1914). Recently, the Gini coefficient has been used to describe concentration in levels of mortality, or in length of life, among different socio-economic groups, and to evaluate inequality in health and in life expectancy (see, e.g., Hanada 1983; Bonetti et al. 2009).

Aim of this paper is to analyse the differences in survival times of job contracts among subgroups of workers, from the point of view of concentration.

We examine the differences both in the length of the first job contract and in the waiting time between the end of the first contract and the beginning of a new one.

We apply the well known Gini index to measure concentration in survival times within groups of workers, and as a way to compare the distribution of survival times across such groups. We consider a test for differences in the heterogeneity of survival distributions, which may suggest the presence of a differential covariates effect on the job contract survival.

The analysis is based on the Italian Compulsory Communications system data, which record all the activations, transformations, fixed-term extensions and anticipated terminations of employment relationships between any worker and employer in Italy since January 2009 until June 2012. The target population is made up by the young workers, between 18 to 35 years old.

The rest of the paper is structured as follows: in Section 2 we briefly review the Gini test for survival data; in Section 3 we analyse the Italian labour market from the point of view of concentration; in Section 4 we conclude.
2. The Gini index for survival data: a brief review

The Gini index measures concentration in the distribution of a positive random variable. Bonetti et al. (2009) propose to apply the Gini index in survival analysis in order to measure concentration in survival times within groups of subjects. In particular, they apply a restricted version of the Gini index to right-censored survival data in order to detect differences in concentration (heterogeneity) between the survival time distributions of two groups.

A number of nonparametric statistical tests exist in the literature to test the difference in survival distribution functions between groups. Common tests are in the class of weighted linear rank tests, including the log-rank test (LR test), the Wilcoxon test (W test), the Gray and Tsiatis test (GT test); see, e.g., Harrington and Fleming 1982; Gray and Tsiatis 1989. Testing for differences between survival distributions via a concentration measure may prove more powerful than these methods, for example when one is far from the proportional hazard structure.

The Gini coefficient of concentration for a positive random variable $X$ with cumulative distribution function $F$ and survival function $S$ is defined as

$$G = 1 - \frac{\int_0^\infty [1 - F(x)]^2 dx}{\int_0^\infty \Pr(X > x) dx} = 1 - \frac{\int_0^\infty [S(x)]^2 dx}{\int_0^\infty S(x) dx} ;$$

see Hanada, 1983.

In survival analysis subjects have usually a finite follow-up time, so we consider the restricted version of the Gini index:

$$G_t = 1 - \frac{\int_0^t [S(x)]^2 dx}{\int_0^t S(x) dx} ,$$

where $t$ represents the longest follow-up time in the data.

Minimum value of $G_t$ is reached when all subjects have the same survival time, while maximum value is obtained when one individual has the maximum survival time and the rest of the population experiences the event immediately.

Bonetti et al. (2009) and Gigliarano and Bonetti (2013) propose a test based on the restricted Gini index $G_t$ for comparing two survival functions related to two different groups. Their Gini test is aimed to test for differences in two survival distributions from the point of view of concentration. The Gini test statistic is
where $\hat{G}_{j,t}$ is the estimator of the restricted Gini index for censored data referred to the group $j$ and $\hat{V}ar(\hat{G}_{j,t})$ is the estimator of the approximate variance of $\hat{G}_{j,t}$, for group $j, j = 1, 2$.

Bonetti et al. (2009) prove that under the null hypothesis of equality of the two survival distributions, the statistic $T$ has an approximate chi-squared distribution with 1 degree of freedom, while, under any alternative to the null hypothesis, $T$ is distributed as an approximate noncentral chi-squared distribution.

### 3. Data description

The empirical illustration is based on a sample of the Compulsory Communications ("Comunicazioni Obbligatorie") data provided by Italian Ministry of Labour and Social Policies.\footnote{The Compulsory Communication Data are used with the permission of the Ministry of Labour and Social Policies thanks to the agreement between the Department of Economics and Social Sciences of Marche Polytechnic University and General Department for the Innovation Technology of the Ministry of Labour and Social Policies. The authors are grateful to Stefano Staffolani and Matteo Picchio for the data preparation.}

The Compulsory Communications (henceforth, CC) data include all activations, transformations, fixed-term extensions, early anticipated terminations of a working relationship, either public or private. The sample refers to all Italian workers born on 15 January, 15 April, 15 July and 15 October of any year. Our database therefore includes about 1 out of 91 of all workers who have been involved in the CC system over the period between January 2009 and June 2012.

The population of interest are the 18-35 aged workers who activated a contract in 2009. Individuals who entered the CC database for the first time after December 31, 2009 are excluded from the analysis.

The CC data have as unit of observation the contract ("contratto di lavoro"), defined as a working relationship between an employer and an employee and characterized by a starting date. However, in the context of mobility analysis, the key concept is the worker rather than the contract; therefore, the worker’s history needs to be reconstructed starting from the original CC data, so that the observation unit becomes the individual.

For more details on the data preparation and cleaning process we refer to Lilla and Staffolani (2011), while further information on the methodology for joining...
different contracts corresponding to same individual can be found in Picchio and Staffolani (2013).

CC data provides information on the daily occupational status of an individual. Here for simplicity a monthly unit of time is considered, and for each month he prevalent contract is selected (according to type and length of contract).

The variable of interest is the occupational status. Four are the types of occupational status considered, that are ordered as follows: (i) not in employment, (ii) temporary contract, including fixed-term contract ("contratto a tempo determinato"), parasubordinate contract ("contratto di collaborazione coordinata e continuativa"), internship contract ("contratto di stage"), interim contract ("lavoro interinale"), (iii) apprenticeship contract ("contratto di apprendistato"), (iv) permanent contract, that is the open-ended contract ("contratto a tempo indeterminato").

We apply the Gini test discussed above to the measurement of concentration in survival times within groups of workers, and as a way to compare the distribution of survival times across such groups.

Analysis of the differences in survival times of job contracts has been performed among subgroups of workers, based on gender, educational level and geographical area.

In particular, we have analysed differences both (i) in the length of the first job contract and (ii) in the waiting time between the end of the first contract and the beginning of the second one. The results are summarised in Table 1 and illustrated in Figures 1 to 4.

**Figure 1 – Male versus female. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new first job.**

A first analysis is aimed at determining whether there are gender differences in the Italian labour market. Figure 1 and Table 1 reveals that there exists no significant difference between young males and young females in the waiting time between
the end of the first contract and the beginning of a new one, while significant differences emerge in the length of the first job contract, which is longer for males and females.

**Table 1** – *P*-values of Gini, Gray-Tsiatis (GT), Log Rank (LR) and Wilcoxon (W) tests for different groups comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Gini</th>
<th>GT</th>
<th>LR</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENDER</strong> (Male versus female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of the first job</td>
<td>0.0152</td>
<td>0.0051</td>
<td>0.4041</td>
<td>0.4997</td>
</tr>
<tr>
<td>Waiting time for new job</td>
<td>0.8366</td>
<td>0.7629</td>
<td>0.9687</td>
<td>0.9865</td>
</tr>
<tr>
<td><strong>EDUCATION</strong> (Tertiary versus non tertiary)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of the first job</td>
<td>0.0000</td>
<td>0.4646</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Waiting time for new job</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td><strong>EDUCATION IN TERTIARY SECTOR</strong> (Tertiary versus non tertiary)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of the first job</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.5114</td>
<td>0.0463</td>
</tr>
<tr>
<td>Waiting time for new job</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td><strong>GEOGRAPHICAL AREA</strong> (North versus South)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of the first job</td>
<td>0.8834</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Waiting time for new job</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

We also test for the presence of significant impact of the educational level on the Italian labour market: Table 1 and Figure 2 shows that tertiary education helps in finding quickly a new job, while it seems not so relevant for activating permanent contracts. With a particular focus on the tertiary economic sector, if a worker has tertiary education he will find quicker a job at the end of the first contract, but the length of his first contract will be shorter, in comparison to workers in the same economic sector but without tertiary education (see Table 1 and Figure 3).
Figure 2 – Tertiary education versus non-tertiary education. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new job.

Figure 3 – Tertiary education versus non-tertiary education within the tertiary economic sector. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new job.

Finally, we compare the Italian macro areas (North, Center and South): no statistically significant differences emerge between North and Center of Italy (data are not shown), while differences emerge between North (or Center) and South of Italy. Table 1 and Figure 4 reveals that the labour market in the North of Italy is characterized by higher percentage of permanent contracts and by shorter waiting time for the activation of the second contract, if compared to the South of Italy.
Figure 4 – *North versus South of Italy. Left-hand side: Length of the first job. Right-hand side: Waiting time for a new job.*

4. Concluding remarks

In this paper we have examined the Italian labour market dynamics from a novel point of view, based on the concentration analysis.

The empirical analysis revealed that there exists no significant difference between male and female in the waiting time between the end of the first contract and the beginning of a new one. Gender differences emerge, instead, in the length of the first job contract, which appears to be significantly longer for males than for females.

Significant differences emerge also among geographical areas: the North of Italy has the highest percentage of permanent contracts and also the shortest waiting time for the second contract.

Finally, different levels of education have different impact on the Italian labour market: tertiary education helps in finding quickly a new job, while it seems not so relevant for activating permanent contracts.

References


GIGLIARANO C., BONETTI M. (2013), Gini test for survival data in presence of small and unbalanced groups, Epidemiology, Biostatistics and Public Health, Volume 10, Number 2, DOI:10-2427/8762.

GINI C. (1912) Variabilità e mutabilità. Contributo allo studio delle distribuzioni e relazioni statistiche. Studi Economico-Giuridici dell’Università di Cagliari III
SUMMARY

Aim of the paper is to analyse the differences in survival times of job contracts among subgroups of workers, based on age, gender, educational level, region.
We examine the differences both in the length of the first job contract and in the waiting time between the end of the first contract and the beginning of a new one.
We apply the well known Gini index to the measurement of concentration in survival times within groups of workers, and as a way to compare the distribution of survival times across such groups. We consider a test for differences in the heterogeneity of survival distributions, which may suggest the presence of a differential covariates effect on the job contract survival.
The analysis is based on the Italian Compulsory Communications system data for the period between January 2009 and June 2012.

Chiara GIGLIARANO, Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, Ancona, e-mail: c.gigliarano@univpm.it
Francesco Maria CHELLI, Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche, Ancona, e-mail: f.chelli@univpm.it