

Dipartimento di Scienze Statistiche Sezione di Statistica Economica ed Econometria

M. Grazia Pittau, Shlomo Yitzhaki, Roberto Zelli

The make-up of a regression coefficient: An application to gender

DSS Empirical Economics and Econometrics Working Papers Series

DSS-E3 WP 2011/3

DSS Empirical Economics and Econometrics Working Papers Series

- 2011/1 Massimo Franchi, Paolo Paruolo "Normal forms of regular matrix polynomials via local rank factorization"
- 2011/2 Francesca Di Iorio, Stefano Fachin "A Sieve Bootstrap range test for poolability in dependent cointegrated panels"

Dipartimento di Scienze Statistiche Sezione di Statistica Economica ed Econometria "Sapienza" Università di Roma P.le A. Moro 5 – 00185 Roma - Italia

http://www.dss.uniroma1.it

The make-up of a regression coefficient: An application to gender

M. Grazia Pittau^{1*}, Shlomo Yitzhaki², and Roberto Zelli¹

¹Department of Statistical Science, Sapienza University of Rome ²The Hebrew University of Jerusalem and Central Bureau of Statistics

September 14, 2011

Abstract

In this paper we illustrate the potential use of an old/new methodology which combines the use of concentration curves in order to investigate the components that make up a regression coefficient. The illustration is based on examining gender differences in the effect of age on labor market participation in Italy. Women participation rate is substantially lower than men, but their age profile is similar. The most striking difference is in terms of hours of work: while Italian men increase their work effort until the age of 35, Italian women reduce it until the age of 39. These results do not differ substantially when we split the working population into employed and self-employed. Earnings increase with age for both men and women, however the local regression coefficient is negative for Italian women in the age of 38–42. This evidence is accentuated when we focus on the employees.

Keywords: Gini, OLS, Concentration curves, Regression decomposition, Italian labor market.

JEL Classification: C30, J16, J21.

^{*}Corresponding author. Email: grazia.pittau@uniroma1.it.

Acknowledgement: we are very grateful to Yolanda Golan for helpful comments.

1 Introduction

The aim of this paper is to validate an old/new methodology which uses concentration curves within the Gini regression framework, to detect whether the relationship between two variables is monotonic, that is if the local slope does not change along the range of the explanatory variable.

Our starting point is based on Yitzhaki (1996) which interprets the OLS and Gini regression coefficients as weighted averages of slopes defined between adjacent observations of the explanatory variable. The method used, whether OLS or Gini regression, determines the weighting scheme. This distinction enables one to differentiate between the contribution of the basic data, i.e., the slopes that are used by both methods, and the contribution of the method used, which determines the weighting scheme. Both weighting schemes are based on the variability of the explanatory variable. The difference is in the way variability is defined. The OLS weighting scheme is based on the variance of the explanatory variable while in Gini regression it is based on the Gini's Mean Difference (GMD) of the same variable. This leads to the Lorenz curve which describes the way the GMD is composed of. Yitzhaki (1998) presents the transformation of the Lorenz curve that is behind the variance, which is substituting the horizontal axis in the Lorenz curve by the variate itself. An additional development that is relevant for the method used in this paper is based on Yitzhaki and Schechtman (2004) who presented the weighting scheme of the Gini regression coefficient as derived from the Concentration Curve of the dependent variable with respect to the explanatory variable. In this paper, we use variations of the Lorenz curve and the Concentration Curve in order to decompose the regression coefficient into the contribution of each segment of the explanatory variable to the regression coefficient. In particular, we are interested in finding out the contributions of positive and negative segments of the regression coefficient so that one can see how the regression coefficient is composed of. The methodology is also applicable to the OLS, but some of the properties that can be revealed by using the GMD are lost when using the OLS. Therefore, we concentrate on Gini regression and we will point out the properties that are relevant to the OLS too. This means that the appropriate interpretation of the linear regression model is actually as a linear approximation to an unknown regression curve (Schechtman, Yitzhaki, and Artzev, 2008). We are interested in exploring the monotonicity of the regression curve. Monotonicity of a general function is investigated by several authors, and the research that is closely related to the approach presented in this paper is Davidov and Zitikis (2005). However, while they use Lorenz curve and Gini index, we are using concentration curves and we decompose the areas enclosed by the curves to decompose the regression coefficient into the contribution of different sections of the explanatory variable. Also, while most of the research in this area is theoretical, we provide the validity of this methodology in examining gender differences in the effect of age on labor market participation in Italy. The structure of the presentation is the following: Section 2 presents a brief review of the theory used in the paper. Readers who are interested only in the practical application can read only section 2.3, the practitioner's guide. Section 3 describes the data and the labor market variables we used in the analysis. Section 4 reports the main empirical results on gender differences in the Italian labor market in terms of labor participation, hours of work and earnings. Section 5 concludes and gives suggestions for further research.

2 A brief review of the theory

2.1 The LMA and NLMA curves: definition and properties¹

The LMA (Line of independence Minus the Absolute concentration curve) is based on an extension of the Lorenz curve. The Absolute Lorenz Curve (ALC), which is the Lorenz curve multiplied by the expected value of the variable defines the GMD or simply Gini, which is similar to the variance. The GMD can be extended to define the equivalents of the Pearson correlation coefficient and the Ordinary Least Squares regression coefficient. The Gini has decomposition properties that nest the decomposition of the variance as a special case². The NLMA curve is a Normalized LMA curve intended to relate the curve to a different parameter such as a regression coefficient or correlation. The use of curves to describe variability and correlation enables a better understanding of the contribution of different sections to the overall parameter. Hence, in what follows we will be skipping from curves to parameters and vice versa.

Let us start with the definition of LOI:

$$\mathrm{LOI}(\mathbf{F}_X^*) = \mu_Y \mathbf{F}_X^* \tag{1}$$

The LOI, which is an abbreviation for Line Of Independence (represented by OCB in Figure 1), presents the cumulative value of the F_X^* smallest values of the explanatory variable, would the dependent and the explanatory variables be statistically independent³. The second curve is:

$$ACC(\mathbf{F}_X^*) = \int_0^{\mathbf{F}_X^*} \mathbf{E}(Y|\mathbf{F}_X) d\mathbf{F}_X$$
(2)

as the Absolute Concentration Curve (ACC) of the dependent variable with respect to the explanatory variable. The ACC (represented by OAB on Figure 1) describes the expected cumulative values of the dependent variable, Y, that are associated with the F(x)observations of the explanatory variable⁴. The LMA curve is defined as the vertical difference between the curves:

$$LMA(F_X^*) = LOI(F_X^*) - ACC(F_X^*)$$
(3)

The LMA presents for each level of the explanatory variable the cumulative difference between the value of the dependent variable would the dependent and the explanatory variable be statistically independent and the actual expected value of the dependent variable. The properties of the LMA curve that are used in this paper are the following:

a) The curve starts at (0,0) and ends up at (1,0). It can take any shape depending on properties of E(Y|X = x).

¹This section is actually an adaptation of Yitzhaki, Itzhaki and Pudalov (2011) to the application presented in this paper.

 $^{^{2}}$ For a review of the properties of the Gini and a comparison of its properties to the properties of the variance, see Yitzhaki (2003). Frick *et. al.* (2006) describe ANOGI, the decomposition of the Gini which is equivalent to ANOVA.

 $^{^{3}}$ The use of the cumulative relationship creates a smoother relationship than using the original observations because random deviations cancel each other

⁴For definitions and properties of Absolute Concentration Curves, see Yitzhaki and Olkin (1991).

Figure 1: The LOI and ACC curves.



- b) The derivative (slope) of the curve is equal to $\mu_Y E(Y|X = x_F)$, where x_F is the value of X at F_X^* . That is the curve has a positive slope if the conditional expected value of the dependent variable is smaller than its expected value, it is horizontal if the conditional expected value of the dependent variable is equal to its expected value, and it declines if $E(Y|X = x_F) > \mu_Y$.
- c) The curve is concave (convex) if $E(Y|X = x_F)$ is monotonically increasing (decreasing) with X.

The properties listed above enable one to follow the monotonicity of the regression curve from the LMA curve. In the next section we will "translate" properties of the LMA curve into the properties of the regression curve.

The LMA curve has several additional properties that enable us to connect it to the familiar concepts of regression coefficients and correlations. Before we proceed we have to define and present the regression and correlation coefficients.

2.2 The Ordinary Least Squares and Gini regression coefficients

The OLS regression coefficient does not need an introduction. However, the LMA curve is directly related to a different concept of a regression coefficient - the Gini regression coefficient (Olkin and Yitzhaki, 1992). The properties listed below hold only with respect to the Gini regression while others hold for both regression coefficients.

The simple OLS regression coefficient is defined as:

$$\beta^{\text{OLS}} = \frac{\text{cov}(Y, X)}{\text{cov}(X, X)} \tag{4}$$

The Gini simple regression coefficient is defined as (Olkin and Yitzhaki, 1992):

$$\beta^{\mathbf{G}} = \frac{\operatorname{cov}\left(Y, \mathbf{F}(X)\right)}{\operatorname{cov}\left(X, \mathbf{F}(X)\right)},\tag{5}$$

where F(X) is the cumulative distribution of X. The Pearson correlation coefficient is defined as:

$$\rho = \frac{\operatorname{cov}\left(Y, X\right)}{\sqrt{\operatorname{cov}\left(X, X\right)\operatorname{cov}\left(Y, Y\right)}},\tag{6}$$

and the relevant Gini correlation coefficient (Schechtman and Yitzhaki, 1987) is:

$$\Gamma = \frac{\operatorname{cov}\left(Y, \mathbf{F}(X)\right)}{\operatorname{cov}\left(Y, \mathbf{F}(Y)\right)}.$$
(7)

Note that the difference between the regression coefficient and the correlation coefficient in each method is in the denominator, which is always positive. That is, in the normalizing factor. The sign of the regression coefficient, in both cases, is determined by the numerator. The normalizing factor is only relevant when we want to distinguish between the correlation and the regression coefficients. At this stage we ignore it.

The next set of properties enables connecting the LMA curve to regression methodology.

- d) The area enclosed between the LMA curve and the horizontal axis is equal to cov(Y, F(X)), which is the numerator in both the Gini regression coefficient and the Gini correlation (Yitzhaki, 2003).
- e) If we redraw the LMA curve with X instead of F_X^* plotted on the horizontal axis, then the area enclosed between the curve and the horizontal axis is equal to cov (Y, X), that is the numerator of the OLS and Pearson coefficients (Yitzhaki, 1998).
- f) If the curve is always above (below) the horizontal axis, then the sign of Gini and Ordinary Least Squares (OLS) regression coefficients of Y on X will be positive (negative) for all monotonic non-decreasing transformations of X^5 .
- g) If the curve intersects the horizontal axis then there exists a monotonic non-decreasing transformation of X that can change the sign of the OLS regression coefficients.
- h) To check whether a monotonic transformation of Y can change the sign of the OLS regression coefficient, one has to plot the LMA(X|Y) and apply the rules in (f) and (g).
- i) If in a given section the curve is convex (concave) then the sign of Gini regression coefficient in that section is negative (positive).

Property (d) enables us to draw two additional curves that will improve the content of the curve. By dividing the vertical axis by cov(X, F(X)) the area enclosed between the curve and the horizontal axis becomes equal to the Gini regression coefficient. This curve is referred to as the Normalized LMA (NLMA) curve. If, on the other hand, one divides the vertical axis by cov(Y, F(Y)) then the area enclosed between the curve and the horizontal axis is equal to the Gini correlation.

Property (e) is listed in order to relate the curve to the OLS regression coefficient. However, the use of X instead of F(X) makes the LOI (eq. 1) non-linear so that all the

⁵ See Olkin and Yitzhaki (1992) for additional properties of the simple Gini regression coefficient, and Schechtman, Yitzhaki and Pudalov (2011) for its properties in a multiple regression framework.

properties that are based on the convexity/concavity of the LMA curve are lost. What remains are the properties that are based on intersection of curve with the horizontal axis.

Figure 2 presents an a-typical curve intended to demonstrate the properties of the NLMA curve. From O to A and from C to D the curve is increasing which means that values of Y are lower than the average value of Y. From A to C the curve declines which means that the values of Y are greater than the average. Between O to B the curve is concave which means that the regression curve increases with X. The curve reaches a peak at A so that it reaches a turning point at this point, and it has a trough at C which means that it has a "negative" turn, from above average to below average.

If the total area enclosed between the curve and the horizontal axis is positive (negative) then the Gini regression coefficient is positive (negative). However, if the curve intersects the horizontal axis it means that by applying a monotonic transformation to X one can change the sign of OLS regression coefficients. For example, shrinking X at the range BD and increasing the value of X in the range OB will increase the range with positive contribution to the regression coefficient, making the coefficient more positive than the one we have.

Figure 2: The NLMA curve.



Having described graphically the Gini regression coefficient, we can follow Yitzhaki and Schechtman (2010) in order to see the contribution of different sections of X to the regression coefficient on Y.

Assume that the observations are partitioned into M disjoint groups, according to different levels of X, denoted by m = 1, ..., M and let $p_m = n_m/n$ be the relative size of group m. Then, the Gini and OLS regression coefficients can be decomposed as follows:

$$\beta^i = \sum_{m=1}^M w^i_m \beta^i_m + w^i_B \beta^i_B, \tag{8}$$

where i=OLS, Gini. For i=OLS we have:

 $w_m = p_m \frac{\operatorname{cov}_m(X,X)}{\operatorname{cov}(X,X)}, \quad \beta_m = \frac{\operatorname{cov}_m(Y,X)}{\operatorname{cov}_m(X,X)}, \quad w_B = \frac{\operatorname{cov}_B(\bar{Y}_m,\bar{X}_m)}{\operatorname{cov}(X,X)}$ $\beta_B = \frac{\operatorname{cov}_B(\bar{Y}_m,\bar{X}_m)}{\operatorname{cov}(\bar{X}_m,\bar{X}_m)} \text{ and } \bar{Y}_m \text{ and } \bar{X}_m \text{ denote the vectors of group means.}$

The structure of the decomposition for i=Gini is identical to the one presented in eq.(8) except that the terms are the equivalent terms of the Gini regression coefficients while the weights are the shares of each section of X in the overall Gini of X⁶. Equation (8) allows us to investigate the contribution of each section of X to the regression coefficient, with one major difference. For the Gini we can present the contribution in an easy way to follow an LMA curve. The decomposition (8) is based on 4 types of components: the group's weight (w_m) , the group's regression coefficient (β_m) , the between-group weight (w_B) and the between-group regression coefficient (β_B) . We will use this decomposition in the application section.

To see the connection to OLS regression it is worth to point out that a variation of the LMA (and NLMA) curves can be used to describe the contribution of each section to the Ordinary Least Squares coefficient. Would we change the horizontal axis to be X instead of F_X then the area enclosed between the new LMA curve and the horizontal axis will be equal to the numerator of OLS regression coefficient. By dividing the new curve by cov(X, X), i.e., the variance of X, then the area enclosed between the new curve and the horizontal axis will be equal to the OLS regression coefficient. However, the equivalent of the LOI ceases to be linear and the convexity/concavity properties of the curve may change. Therefore, to get as much as we can from the curve, it is worth to use the Gini regression rather than the OLS.

2.3 A practitioner's guide

In this section we present the relationship between some theoretical patterns of the slope of the regression curve, i.e., $\partial E(Y|X = x) / \partial x$, and the representations of those patterns in LMA. There are several properties of the LMA curve that are needed for the analysis:

- (a) If the curve increases (decreases, horizontal) then E(Y|X) is below (above, equal) average Y.
- (b) If the LMA curve is concave (linear, convex) then the slope of the regression curve is positive (flat, negative) locally.
- (c) If the LMA curve is always above (below) the horizontal axis then the sign of the regression coefficient is positive (negative) for all possible monotonic non-decreasing transformations of X, i.e., for all possible assumptions on the distribution of X.
- (d) If the LMA curve intersects the horizontal axis then there are sections of X with negative and positive correlation with Y. This implies that one can find monotonic non-decreasing transformation of X that can change the sign of the regression coefficient.

Following are some examples for the relationship between the regression curve and LMA curve.

Figure 3 presents the curvature of a regression curve that is discontinuous and "jumps" from zero to one about the median of X. Then the LMA curve would create a symmetric triangle with a slope of 0.5 at the left-hand side of the peak and then the curve symmetrically

⁶This formula is applicable to non-overlapping groups. For an application to overlapping groups, see Yitzhaki, Itzhaki and Pudalov (2010).

declines at the second half. Changing the values of Y from zero to another constant changes the slope of the LMA curve, but the shape will remain the same.



Figure 3: The discrete regression curve

Figure 4 presents a linear regression curve. That is Y is a linear function of X. In this case the LMA curve looks like a bell curve with the peak being at the median. At the left-hand side of the figure the curve is increasing, an indication that the dependent variable is below the average, while at the right hand side of the figure the curve is declining which indicates that Y is above its average⁷. The curious reader may ask why we propose the use of the LMA curve over alternative presentation of the regression like a scatter diagram or a diagram with X on the horizontal axis and E(Y|X) on the vertical axis, when actually they rely on the same data and carry the same information. The answer to this question is based on two properties of the LMA curve: the first is that the LMA curve is a cumulative curve and as such it tends to reduce the effect of random perturbations and the second is that the LMA curve is informative about the overall regression coefficient. It is the only curve that can tell whether a monotonic transformation of the variable can change the sign of the regression coefficient.

 $^{^7}$ A linear regression curve with a negative slope will result in a mirror image LMA curve that will be below the horizontal axis.



2.4 Estimation and testing

Estimation of the parameters involved that are related to the Gini methodology (Gini regression and Gini correlation coefficients) were developed in Schechtman and Yitzhaki (1987), Schechtman, Yitzhaki and Artzev (2008), and Schechtman, Yitzhaki and Pudalov (2011). The estimators are the sample's representation of the population parameters with the empirical cumulative distribution substituting for the cumulative distribution. It is shown in the above mentioned papers that the estimators are composed of functions of U-statistics, so that for large samples the distributions of the estimators converge to the normal. The estimation of the LMA curve is also based on the sample's version of the population curve. However, we do not have yet tests that can be used for testing concavity or convexity of the curve, but we can use the Gini parameters that are based on truncated segments of the curve to test for concavity/convexity. Concerning the intersection of the LMA curve with the horizontal axis, Schechtman et. al. (2008a) suggest a test for intersections of Absolute Concentration Curves which can be used to test for whether the LMA curve intersects the horizontal axis. However, in evaluating the power of the test one has to assume the alternative hypothesis. Therefore, our recommendation at this stage is to rely on large samples and pay attention only to large sections in order to be sure that the findings are significant.

3 The data

The analysis on the gender gap in the labor market is based on data drawn from the European Union Survey on Income and Living Conditions (EU-SILC). The EU-SILC project is carried out under European Union legislation (council regulation No. 1177/2003) and it was formally launched in 2004 for the EU15. Soon after its launch, EU-SILC became the principal source for comparative statistics on labor market conditions, income and related phenomena in the member states of the European Union. The Survey is annual and it has a longitudinal and a cross-sectional component. These national surveys are based on a standard questionnaire and provide *ex-ante* harmonized micro-data on European countries.

In this paper we focus on Italy, but similar analyses can be carried out also for other European countries.

 Table 1: Population aged 18–65 years by activity status and gender (percentage values). Italy 2007

	MEN	WOMEN	ТОТ
Population	49.9	50.1	100.0
Employed	71.9	46.8	59.2
Unemployed	6.2	7.1	6.7
Retired	9.0	6.6	7.8
In other inactivities	12.3	38.8	25.6
Undefined	0.6	0.7	0.7
Total	100	100	100
Participation rate	78.1	53.9	65.9

Source: authors' calculation on weighted data from EU-SILC 2008. Cross-sectional weights. Note: Activity status defined according to the PX050 derived variable. Participation rate defined as the percentage of employed and unemployed over the total population aged 18–65

A comprehensive age profile of economic variables is currently possible only using crosssectional data. Adoption of a cohort analysis would have required a much larger number of waves than those currently available. Therefore, our analysis focuses on the latest wave released for the year 2008, which collects information on household activities related to the year $2007.^{8}$

For each respondent we have selected age, personal cross-sectional sampling weight, activity status, average weekly hours of work in main and secondary activities, employee income and self-employed income. Note that we are able to identify the age in terms of years and quarters. We selected individuals aged between 18 and 65 years old, that corresponds to the end of the high school and the most frequent year of retirement in Italy, respectively.

In the following we give a description of the variables (see for more details the reference manual: Eurostat, 2009). Each household respondent (of any age) is labelled by her activity status (PX050), which is a derived variable of the survey. The activity status is calculated on the basis of questions concerning the number of months spent at full-time work, at part-time work, in unemployment, in retirement, studying, and in inactivity. The respondent is identified as employed (full-time and part-time), unemployed, retired, or in other inactivities if the respondent has spent more than six months in that status⁹.

Table 1 reports the structure of the population aged between 18 and 65 by activity status and gender.

The hours of work corresponds to the number of hours the respondent normally works in a week in his/her job. This covers all hours including extra hours, either paid or unpaid, which the person normally works, but excludes the travel time to work as well as the main meal breaks. If multiple jobs are held, it is asked to indicate the number of working hours in the the main job and in the subsidiary jobs. The main job is the one with the greatest number of hours usually worked. Some respondents, particularly the own-employed and family workers, may not have usual hours, in the sense that their hours vary considerably from week to week or month to month. When the respondent is unable to provide a figure for usual hours for this reason, the average of the hours actually worked per week over the past four weeks is used as a measure of usual hours (Eurostat, 2009).

Employee income is defined as the total annual remuneration, in cash or in kind, payable by an employer to an employee for work done by the latter during the income reference

⁸Cross-sectional UDB SILC 2008 Rev.2, October 2010.

 $^{^{9}}$ If the respondent has not spent at least six months in none of those status, the variable is coded as missing/undefined.

Table 2: Employees and self-employed over total employment aged 18–65 years by gender (percentage values). Italy 2007

	MEN	WOMEN	TOT
Employees	73.6	83.7	77.6
Self-employed	26.4	16.3	22.4
Total employment	100.0	100.0	100.0

Source: authors' calculation on weighted data from EU-SILC 2008. Cross-sectional weights. Note: Employees and self-employment defined according to labour income prevalence criterion.

period. Employee income is composed by gross employee cash or near cash income, gross non-cash employee income, and employer's social insurance contribution. Self-employment income includes gross cash benefits or losses from self-employment including royalties, and value of goods produced for own consumption.

An individual can earn his/her labor income from multiple activities. Therefore an individual can get simultaneously wages and self-employment income. In Italy this phenomenon is particularly prominent. Based on EU-SILC data, the estimated percentage of employees who only work exclusively for a public or private employer represent 58.3% of the total population in work, 19.2% is the percentage of self-employed persons who exclusively work in their own business, whereas people who receive simultaneously employee and selfemployment income represent 22.5% of the employment. Unfortunately, only few of them (12.2%) declare the number of hours usually worked in a second job, thus we are not able to disentangle the hours of work worked in an employment activity and the hours worked in self-employment activities. To allocate the respondent as employed or self-employed, we decided to apply an income predominance criterion. According to this criterion, we allocated 82% of the multiple income respondents to the employment income group and the remaining 18% to the self-employment one. This criterion is in good accordance with the self-declared status in employment of the respondent, that is the prevailing income criterion matches (on average) the individual own perception of his/her current main activity. Those allocated to the employee group have an average employee income of 32,517 euros and an average selfemployment income of 2,194 euros. Those allocated to the self-employment group, instead, have an average employee income of 8,676 euros and an average self-employment income of 42,406 euros.

Table 2 reports the percentage of employees and self-employed in relation to the total employment. For men workers for themselves represent over one quarter of the total. Self-employed women, instead, represent about 16% of women in work.

Table 3 reports average weekly hours of work and annual earnings by gender in Italy in 2007. On average, women tend to work less than men by around 17%, corresponding to less than seven hours a week, and earn annually about 25% less, corresponding to about 9,200 euros. Note, however, that gender differences in participation rate are much larger than gender differences in hours worked by participants.

4 Analysis of age profiles

Our empirical strategy is to find whether there is a gender difference in the age profiles of key variables in the labor market, under the plausible assumption that the working age population is between 18 and 65 years old. We have selected the regression with the age as the explanatory variable because it is well documented that in many areas age does not

Week	ly hours o	f work	
	MEN	WOMEN	TOT
Employees	41h 14'	34h 34'	38h 25'
Self-Employed	46h 52'	40h 7'	44h 56'
Total Employment	42h 43'	35h 31'	39h 54'
Annu	ual earning	gs (€)	
	MEN	WOMEN	TOT
Employees	37,494	28,638	33,745
Self-Employed	$38,\!887$	28,266	$35,\!856$
Total Employment	37,708	28,520	34,107

Table 3: Weekly hours of work and annual earnings in euros by type of employment and gender(averages). Italy 2007

Source: authors' calculation on weighted data from EU-SILC 2008. Cross-sectional weights. Note: Employees and self-employment defined according to labour income prevalence criterion.

create monotonic relationship with other variables (Deshpande and Suresh, 1990). The first variable is participation. Next we take working hours per week. Finally we take employee and self-employment earnings. For each variable, we look at the NLMA curve of men and women to decompose the overall Gini and OLS regression coefficient into the contributions of each section of X (age groups). The uncertainty of the estimated regression coefficients is assessed by a Jackknife resampling methods, a procedure that recalculates the coefficients dropping one observation at a time and adjusting the weights accordingly (Yitzhaki, 1991).

4.1 Participation

Individual participation is a binary variable, assuming the value of one if the respondent's activity status is employed or unemployed and zero otherwise, according to the activity status defined above. The participation rate (at each age) is defined as the percentage of employed and unemployed relative to the corresponding total population. Figure 5 reports the NLMA curve for participation and age by gender. The overall Gini regression coefficient of participation on age is negative. This can be seen from the figure because the area below the horizontal axis is greater than the area above it. Also can be seen from the figure that the Gini regression coefficient for women is more negative than the regression coefficient for men because the area above the axis for women is smaller than that for men and the area below the horizontal axis is greater for man than for women. This can be verified from Table $4.^{10}$ Both NLMA curves intersect the horizontal axis. This means that the Gini covariance (the numerator of the Gini regression coefficient) is partly positive (for low age section) and partly negative (for high age section). A monotonic non-decreasing transformation applied to X expands (shrinks) the regions with a derivative greater (smaller) than one. But since the numerator of the Gini regression coefficient uses only the cumulative distribution of X, its magnitude will not be affected. However, the denominator is affected. But since the denominator is always positive, a monotonic transformation of X can change the magnitude of the Gini regression coefficient but it cannot change its sign. On the other hand, the OLS regression coefficient relies on the same NLMA curve with one difference: instead of having F_X on the horizontal axis we have X. Clearly, if there is intersection with the horizontal axis for NLMA for Gini there is also one for OLS. Applying a transformation on X that

 $^{^{10}}$ All the computing was done with R (R Development Core Team, 2010).

shrinks the section we do not like and increases the section we favor can change the sign of the OLS regression coefficient. For example, if we sufficiently expand the variable age for young ages and shrink it for high ages we will be able to change the sign of the OLS regression coefficient from negative to positive. For example, the log transformation applied to the age variable may be able to change the sign of the OLS regression coefficient. If it is not sufficient to do the job, we may try the double log. This is another advantage of the Gini regression coefficient over the OLS, the greater robustness with respect to transformations of variables.¹¹

From the slope of the NLMA curve we can identify ranges with above and below the average participation rate. If the curve increases (decreases) then the participation rate is below (above) the average. As expected, participation rate of men and women is below the (corresponding) averages at the extremes of the range, that is under 27 (corresponding to the 14th percentile) and above 57 years (corresponding to the 84th percentile) for men, and under 25 (corresponding to the 12th percentile) and above 55 years (corresponding to the 80th percentile) for women. Until the age of 27 for men and 25 for women the participation rate is rapidly increasing, while past the age of 57 for men and 55 for women it rapidly decreases. In the middle, the participation rate is quite stable and above average; the NLMA curve is almost linear for men and slightly convex for women, indicating that the changes in the local coefficients are close to zero for men and slightly negative for women. Table 4 presents the contributions of different sections of the variable age to the Gini and OLS regression coefficients, according to this age grouping. The linear segments connecting the extreme points of each group enables us to learn about the behavior of the intra-group regression coefficients. To see this, note that by truncating the overall distribution from above and below and normalizing the axis we get another NLMA figure that portrays the behavior of the intra-group regression coefficient. As can be seen from Figure 5, both for men and women, it is impossible to find a monotonic non-decreasing transformation of X that can change the sign of the regression coefficient of the first and third age groups, while in the second group, the curve for both women and men intersect the linear segment. It is interesting that in this middle group, we could have an additional intra-group decomposition that would have created positive (in the first section) and negative (in the second section) regression coefficients. In conclusion, although average participation rates are different, it seems that the pattern of the regression coefficients for women and men over the life time are similar with the changes in women behavior precede, in terms of age, the changes in men.

The NLMA curve is above the horizontal axis until the 40th percentile for men, corresponding to the age of 38, and until the 27th percentile for women, corresponding to the age of 33. If we define the groups according to the intersections with the horizontal axis, then the between-group component in both regressions would be equal to zero, the regression coefficients for young ages will be positive, the regression coefficients for old ages negative. Obviously, the overall regression coefficients do not change. Table 5 reports the decomposition of the regression coefficients according to this age grouping.

¹¹To search for a monotonic transformation that can change the sign of the regression coefficient we suggest the following procedure. Separate the sections of the explanatory variable into positive and negative regression coefficient. To study the properties of a function that can do the job, start with a piece-wise continuous linear function. Shrink unfavorable sections and expand favorable one. This way one can identify the simplest transformation that is needed. The closer this function is to linearity the more likely that the sign of the regression coefficient is sensitive to manipulations. It should be also noted that in order to check whether monotonic transformation of the dependent variable can change the sign of the OLS regression coefficient, one should reverse the role of the variables.

		M	EN		WOMEN				
Section	Gi	ni	OLS		Gi	ni	OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.012	6.52	0.007	6.51	0.006	6.32	0.002	6.34	
		(0.43)		(0.43)		(0.75)		(0.74)	
II	0.175	-0.09	0.307	-0.10	0.184	-0.31	0.316	-0.29	
		(0.04)		(0.05)		(0.07)		(0.07)	
III	0.009	-5.82	0.004	-5.87	0.015	-5.01	0.008	-5.05	
		(0.53)		(0.52)		(0.34)		(0.33)	
Between	0.803	0.30	0.682	0.31	0.795	-0.69	0.670	-0.69	
		(0.00)		(0.00)		(0.05)		(0.05)	
Overall	1.000	-0.23	1.000	-0.24	1.000	-0.59	1.000	-0.58	
		(0.00)		(0.00)		(0.00)		(0.03)	

Table 4: Gini and OLS regression coefficients of participation on age-age grouping according to the slope of NLMA, Italy 2007

Men. Section I: under 27. Section II: between 27 and 57. Section III: above 57. Women. Section I: under 25. Section II: between 25 and 55. Section III: above 55. In parentheses Jackknife standard errors of the regression coefficients.

		M	EN		WOMEN				
Section	Gi	ni	OI	OLS		ni	OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.111	3.15	0.093	3.26	0.021	3.68	0.036	3.78	
		(0.10)		(0.10)		(0.20)		(0.19)	
II	0.191	-2.56	0.207	-2.65	0.100	-2.01	0.352	-2.06	
		(0.07)		(0.07)		(0.06)		(0.06)	
Between	0.699	0.00	0.700	0.00	0.879	0.00	0.612	0.00	
		(0.04)		(0.04)		(0.05)		(0.05)	
Overall	1.000	-0.23	1.000	-0.24	1.000	-0.59	1.000	-0.58	
		(0.04)		(0.02)		(0.04)		(0.03)	

Table 5: Gini and OLS regression coefficients of participation on age-age grouping according to the NLMA intersection with the horizontal axis, Italy 2007

Men. Section I: under 38. Section II: above 38.

Women. Section I: under 33. Section III: above 33.

In parentheses Jackknife standard errors of the regression coefficients.

Figure 5: NLMA curve for participation of men and women aged 18-65 in Italy (2007).



4.2 Hours of work

In Italy the behavior of men and women in terms of working hours is very different. The OLS and Gini regressions estimates result in an overall positive effect of age on working hours for men, whereas for women the overall association is negative (but/eventhough in both cases they are not statistically significant). Figure 6 reports the NLMA curve for weekly hours of work and age by gender. Table 6 reports the estimates of the OLS and Gini coefficients and their corresponding decompositions according to the age sections identified by the NLMA curve. The NLMA curve for men is increasing and concave until the 26th percentile, corresponding to 35 years old, indicating that the weekly hours of work are below the average for men and increasing: the β coefficient is positive (β OLS =0.113 and β Gini=0.098). Since age is measured in quarters, this roughly means that men increase their weekly effort at work by less than half an hour, each year, until the age of 35. When they are passed 35 years, they stabilize their effort above the men average. The NLMA curve is declining and slightly convex until the age of 52, and the β coefficients are negative and close to zero. Having past the age of 52, the NLMA curve is increasing and convex, indicating that working hours are below average (for men) and declining. The β coefficient is negative $(\beta \text{ OLS} = -0.083 \text{ and } \beta \text{ Gini} = -0.063)$, indicating that the time weekly spent for work fade at a pace of around 20 minutes each year older. The women age profile of working hours is almost specular to men. From the NLMA curve we can identify three age groups: up to 39 years old; between 39 and 51; and over 51. In the first section the NLMA curve declines and it is slightly convex. This means that working hours are above average (for women) and declining. The estimated negative coefficient (β OLS =-0.132 and β Gini=-0.130) implies that adding one year in age corresponds to an expected weekly reduction of more than half an hour at work until the age of 39. Between 39 and 51 years old the effort oscillates around the average and then slightly increase. Past 51 years, women drastically reduce their weekly effort at work at an expected annual rate of 1 hour and 40 minutes.

The NLMA curve for men intersects the horizontal axis. Therefore, one can split the sample into two groups: the young and the elderly, with the young having a positive regres-

Figure 6: NLMA curve for weekly hours of work and age by gender in Italy–Total employment, Italy 2007



sion coefficient while the elderly a negative one. Because the curve crosses the horizontal axis one can find a monotonic transformation of age that can change the sign of the regression coefficient. On the other hand the NLMA for women does not cross the horizontal axis and therefore, although it is possible to find ranges with positive regression coefficient, it is impossible to change the sign of the regression coefficient by a monotonic increasing transformation of X.

When we split the working population in employed and self-employed, the substantive results do not change, but each subpopulation has its peculiarities. Let us start with the employees, which represent the 77.6% of the total population employed. The corresponding NLMA curve is presented in Figure 7 and Table 7 reports the overall estimated OLS and Gini coefficients and their decomposition according to the NLMA curve. Hours of work of employed men are increasing until the age of 27 (the β coefficients are positive), then they are stable around values above the average until the age of 48 and then they start decreasing. For employed women until the age of 39 working hours are declining then they slightly increase between 39 and 50 years old and then start declining again.

The intersections with the x-axis of the NLMA curve of working hours of the selfemployed by gender suggests to partition men in two age groups, with a divide at the age of 50 (see Figure 8). Up to 50 years old, the estimated OLS and Gini coefficients are positive (see Table 8), indicating that self-employed men annually increase their weekly effort at work by more than half an hour. Having turned 50, self-employed men are expected to annually reduce their working hours by two hours a week. Behavior of self-employed women is different. Until the age of 36 their effort at work substantially decreases every year (by 70 minutes per week). Between 36 and 53 years old working hours increase and after 53 they start declining. Monotonic transformations of the (explanatory/dependent) variable can change the overall sign of the regression coefficients.

		М	EN		WOMEN				
Section	Gi	ni	OLS		Gi	ni	OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.050	0.098	0.045	0.113	0.118	-0.130	0.119	-0.132	
		(0.048)		(0.047)		(0.044)		(0.042)	
II	0.130	-0.038	0.130	-0.037	0.066	0.059	0.045	0.057	
		(0.022)		(0.0227)		(0.067)		(0.067)	
III	0.022	-0.063	0.015	-0.083	0.026	-0.444	0.016	-0.405	
		(0.084)		(0.084)		(0.112)		(0.106)	
Between	0.807	0.010	0.810	0.008	0.790	-0.053	0.820	-0.050	
		(0.001)		(0.001)		(0.016)		(0.016)	
Overall	1.000	0.005	1.000	0.005	1.000	-0.062	1.000	-0.062	
		(0.010)		(0.008)		(0.011)		(0.014)	

Table 6: Gini and OLS regression coefficients of weekly hours of work on age by gender - Totalemployment - Italy 2007

Men. Section I: under 35. Section II: between 35 and 52. Section III: above 52. Women. Section I: under 39. Section II: between 39 and 51. Section III: above 51.

In parentheses Jackknife standard errors of the regression coefficients.

Table	7:	$Gini \ and$	OLS	regression	coefficien	ts oj	f weekly	hours	of	work	on	age	by	gender	 Employee	s
- Italy	200	07														

		M	EN		WOMEN				
Section	Gi	ini	OLS		Gini		OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.012	0.301	0.005	0.341	0.112	-0.160	0.113	-0.161	
		(0.133)		(0.132)		(0.045)		(0.044)	
II	0.207	-0.026	0.210	-0.025	0.057	0.098	0.037	0.100	
		(0.022)		(0.022)		(0.072)		(0.072)	
III	0.005	-0.189	0.039	-0.132	0.031	-0.214	0.021	-0.238	
		(0.052)		(0.054)		(0.097)		(0.091)	
Between	0.726	-0.058	0.746	-0.056	0.800	-0.097	0.829	-0.094	
		(0.011)		(0.011)		(0.016)		(0.016)	
Overall	1.000	-0.053	1.000	-0.053	1.000	-0.094	1.000	-0.096	
		(0.010)		(0.008)		(0.010)		(0.011)	

Men. Section I: under 27. Section II: between 27 and 48. Section III: above 48. Women. Section I: under 39. Section II: between 39 and 50. Section III: above 50. In parentheses Jackknife standard errors of the regression coefficients.

Figure 7: NLMA curve for weekly hours of work and age by gender in Italy (2007)–Employees.



Figure 8: NLMA curve for weekly hours of work and age by gender in Italy (2007)-Self-Employed.



4.3 Earnings

We first analyse the relationship between age and earnings of total employment. The overall estimated Gini regression coefficient is 586.3 for men and 345.4 for women (similar values for the OLS estimates, see Table 9). From Figure 9 we can see that both the NLMA curves do not cross the horizontal axis and therefore no monotonic transformation can change the sign of the overall coefficients. The men's curve is positive and concave until the age of 47 and becomes convex between the range of 47–52, indicating that the covariance is positive for men aged less than 47 and therefore the sign of the local regression coefficients are also

		M	EN		WOMEN				
Section	Gi	ini	OLS		Gi	ini	OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.077	0.135	0.343	0.148	0.061	-0.295	0.039	-0.280	
		(0.043)		(0.044)		(0.228)		(0.235)	
II	0.018	-0.517	0.046	-0.528	0.205	0.125	0.179	0.127	
		(0.108)		(0.107)		(0.103)		(0.105)	
III					0.023	-0.211	0.011	-0.283	
						(0.366)		(0.350)	
Between	0.904	0.009	0.611	0.009	0.710	-0.003	0.770	-0.003	
		(0.031)		(0.031)		(0.046)		(0.046)	
Overall	1.000	0.035	1.000	0.032	1.000	0.021	1.000	0.013	
		(0.026)		(0.019)		(0.043)		(0.037)	

Table 8: Gini and OLS regression coefficients of weekly hours of work on age by gender - Self-Employed - Italy 2007

Men. Section I: under 50. Section II: above 50.

Women. Section I: under 36. Section II: between 36 and 53. Section III: above 53.

In parentheses Jackknife standard errors of the regression coefficients.

positive. For men between 47 and 52, instead, the covariance is negative. As shown in Table 9, in that range mens' earnings are expected to decline at a rate of approximately 2,500 euros per year. After that age they increase.

The NLMA shape for women changes from a concave to a convex curve in the range of 39 percentile and 56 percentile of the cumulative distribution of age, that corresponds to a range of about five years, between 38 and 42. In this range, that accounts 12% of the women employed, the covariance between income and age is negative and therefore the sign of the regression coefficient is negative. Table 9 shows the contribution of each age group to the overall regression coefficient. Until the age of 38, annual income is expected to have an increase of about 2,200 euros per year. Then for women between 38 and 42 income is expected to decline by a rate of about 800 euros per year. Having passed the age of 42, earnings start increasing again.

This evidence is accentuated when we focus on the employees (see Figure 10 and Table 10). Income of employee women under the age of 37 is expected to increase at a rate of about 2,400 euros per year, then to drastically decline by about 5,500 euros per year and, turned 42, to increase by about 1,200 euros per year. Such a pattern can not be easily detected for self-employed women.

In all cases the regression coefficients obtained by the OLS and the Gini method are similar, making easier the interpretation of the results. However, only using the concentration curves based on the Gini regression we are able to group the data into sections that contribute differently to the overall regression coefficient.

5 Conclusion and suggestion for further research

In this paper we illustrated the potentiality of using concentration curves within the Gini regression framework in depicting gender differences in the age profiles in the labor market in Italy. Building up the Line of independence Minus the Absolute concentration curve (NLMA) for each key variable in the labor market, we detected sections in the curve that

		M	EN		WOMEN				
Section	Gi	ni	OLS		Gi	ni	OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.263	765.4	0.292	780.8	0.103	556.3	0.094	581.3	
		(78.1)		(72.0)		(71.6)		(66.1)	
II	0.014	-685.2	0.003	-690.6	0.006	-200.4	0.001	-227.1	
		(128.4)		(124.9)		(66.3)		(66.1)	
III	0.034	399.7	0.019	171.6	0.132	228.2	0.137	209.2	
		(78.1)		(79.4)		(84.7)		(82.3)	
Between	0.689	520.1	0.686	515.7	0.759	342.3	0.768	342.6	
		(30.5)		(30.5)		(27.6)		(27.6)	
Overall	1.000	586.3	1.000	582.6	1.000	345.4	1.000	346.2	
		(29.3)		(22.5)		(24.8)		(19.8)	

Table 9: Gini and OLS regression coefficients of earnings on age by gender - Total employment -Italy 2007

Men. Section I: under 47. Section II: between 47 and 52. Section III: above 52. Women. Section I: under 38. Section II: between 38 and 42. Section III: above 42. In parentheses Jackknife standard errors of the regression coefficients.

Table 10: Gini and OLS regression coefficients of earnings on age by gender - Employees - Italy2007

		M	EN		WOMEN				
Section	Gi	ini	OLS		Gini		OLS		
	weight	β	weight	β	weight	β	weight	β	
Ι	0.075	941.9	0.063	974.3	0.093	608.5	0.082	638.2	
		(78.1)		(72.0)		(71.6)		(66.1)	
II	0.005	-119.3	0.001	-169.3	0.010	-1403.4	0.003	-1408.5	
		(128.4)		(124.9)		(66.3)		(66.2)	
III	0.161	412.4	0.193	395.5	0.123	315.2	0.131	298.6	
		(78.1)		(79.4)		(84.7)		(82.3)	
Between	0.759	628.2	0.743	628.6	0.771	372.9	0.784	372.9	
		(30.5)		(30.5)		(27.6)		(27.6)	
Overall	1.000	606.0	1.000	605.0	1.000	377.1	1.000	380.3	
		(29.3)		(22.7)		(24.8)		(19.8)	

Men. Section I: under 36. Section II: between 36 and 39. Section III: above 39.

Women. Section I: under 37. Section II: between 37 and 42. Section III: above 42.

Figure 9: NLMA curve for earnings and age by gender in Italy (2007)-Total employment.



are convex and sections that are concave. The Gini regression enabled to quantify the contribution of each section to the overall regression coefficient. This connection between the Gini regression coefficients and the concentration curves allowed us to verify monotonicity of regression curves between age and each variable. The search for a monotonic relationship is important because conclusions that are based on non-monotonic regression curve may be sensitive to the selection of the range of the variables in the model and to monotonic transformation of the variables so that two researchers, using the same data can reach contradicting conclusions concerning the effect of one variable on the other.

Based on the European Statistics on Income and Living Conditions (EU-SILC), we analyzed the Italian labor market focusing on participation, hours of work and earnings for the total population employed and, separately, for employees and self-employed. Although Italian women participation rate is markedly lower than men, their age profiles are quite similar. The behavior of men and women in terms of hours of work is, instead, almost specular. The estimated effect of age on hours of work is positive for men aged less than 35 years, whereas is negative for women until the age of 39. Past those ages, the work effort is slightly decreasing for men and increasing for women until the early 50s. Over 50, the reduction of working hours is much more pronounced for women than men. These results do not differ substantially when we split the working population into employees and self-employed. The covariance of earnings with age is always positive, for men and women, indicating a positive overall relationship between earnings and age. However, for women, it was possible to identify a range between 38 and 42 years, so that the sign of the covariance becomes negative. This evidence is accentuated when we focus on the employees.

Further research is required to fully exploit the advantages of the Gini regression. The conclusion is that between the two variables that represent participation in the labor market, weekly hours of work is the major player in differentiating between men and women, and as such it may play an important role in explaining the wage gap between men and women. Further research is needed in order to see whether this result holds also in other European countries. Comparison with Israeli results indicates that while in Israel the difference between women and men is both in participation and in hours worked (Yitzhaki and





Schechtman, 2010), in Italy the difference is essentially in terms of hours of work. Looking at working hours, while Israeli men are similar to Italian men, the behavior of Italian women is totally different. However, the comparison with Israeli data may be sensitive to different fertility rates and labor market legislation. Therefore, it seems that the way forward is to compare Italy to other European countries.

Further theoretical research is needed in order to move from simple Gini regression to multiple Gini regression. It can be shown that the regression coefficients in a multiple regression framework are actually composed of a (complicated) function of the simple regression coefficients of all the variables that are participating in the multiple regression. Hence, it seems that analyzing the sensitivity of regression coefficients in a multiple framework is actually based on the properties of the simple regression coefficients.

References

DeGroot, M. H. (1975). *Probability and Statistics*, London: Addison-Wesley Publishing Company.

Deshpande, J. V. and R. P. Suresh, (1990). Non-monotonic Ageing. Scandinavian Journal of Statistics, 17, 257–262.

Davidov, Y. and R. Zitikis (2005). An index of monotonicity and its estimation: a step beyond econometric applications of the Gini index, *Metron*, LXIII, 3, 351–372.

Eurostat (2009). Eurostat Community Statistics on Income and Living Condition, Study Documentation, Luxembourg.

Frick, R. J., J. Goebbel, E. Schechtman, G. G. Wagner, and S. Yitzhaki (2006). Using Analysis of Gini (ANOGI) for Detecting Whether Two Sub-Samples Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience, *Sociological Meth*ods and Research, 34(4), 427–68.

Olkin, I. and S. Yitzhaki (1992). Gini Regression Analysis. *International Statistical Review*, 60, (2), 185–196.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Schechtman, E.; A. Shelef; S. Yitzhaki; R. Zitikis (2008). Testing hypothesis about absolute concentration curve and marginal conditional stochastic dominancem, *Econometric Theory*, 24 (4), 1044–1062.

Schechtman, E. and S. Yitzhaki (1987). A Measure of Association Based on Gini's Mean Difference, *Communications in Statistics: Theory and Methods*, A16 (1), 207–231.

Schechtman, E., S. Yitzhaki and Y. Artzev (2008). Who Does Not Respond in the Household Expenditure Survey: An Exercise in Extended Gini Regressions, *Journal of Business & Economic Statistics*, 26 (3), 329–344.

Schechtman, E.; A. Shelef; S. Yitzhaki; R. Zitikis (2008a). Testing hypothesis about absolute concentration curve and marginal conditional stochastic dominance, *Econometric Theory*, 24 (4), 1044–1062.

Schechtman, E., S. Yitzhaki and T. Pudalov (2011). Gini's multiple regressions: two approaches and their interaction, *Metron*, LXIX, n. 1, 66–97.

Yitzhaki, S. (1991). Calculating Jackknife Variance Estimators for Parameters of the Gini Method, Journal of Business & Economics Statistics, 9, 2, 235–239.

Yitzhaki, S. (1996). On Using Linear Regressions in Welfare Economics, *Journal of Business* & *Economic Statistics*, 14(4), 478–86.

Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini, *Research on Economic Inequality*, 8, 13–30.

Yitzhaki, S. (2003). Gini's mean difference: A superior measure of variability for non-normal distributions, *Metron*, LXI, 2, 285–316.

Yitzhaki, S. and I. Olkin (1991). Concentration Indices and Concentration Curves, in Karl Mosler and Marco Scarsini (eds.) *Stochastic Orders and Decisions under Risk*, Institute of Mathematical Statistics: Lecture-Notes Monograph Series, 19, 380–392.

Yitzhaki, S. and E. Schechtman (2004). The Gini Instrumental Variable, or the "double instrumental variable" estimator, *Metron*, LXII, 3, 287–313.

Yitzhaki, S. and E. Schechtman (2010). Identifying Monotonic and Non-monotonic Relationships. Available at http://ssrn.com.

Yitzhaki, S., Itzhaki, R. and Pudalov, T. (2011). A Nonparametric Item Response Theory. Available at SSRN: http://ssrn.com/abstract=1785197.