Conference Title
The thought and work of Corrado Gini

- The Gini Methodology
- By
- Shlomo Yitzhaki

The Gini reveals more (Lambert and Decoster) The Gini Methodology |Springer

The GMD offers a response to Leamer's critique

THE PROBLEM:

- Edward Leamer (1983, p. 37): "Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously."
- Angrist and Pischke (2010) response by listing the improvements in research design, better data collection, better definitions of the research question, and more.

- No improvement in methodology

The target of the talk

- The aim of the book is to develop an old/new methodology that will avoid whimsical assumptions.
- The aim of the paper is to illustrate the methodology empirically.
- Whimsical assumptions: Imposed but are not supported by the data.
- They can change the sign of a regression coefficient.(Either alone or in combination with other whimsical assumptions).

- Typical research in economics: Running hundred of regressions, choosing a few.

Whimsical assumptions That can be Avoided by the GMD

- Symmetry in correlation when the distributions are different
- Linearity in regression
- The free use of Transformations

THE ROLE OF THE VARIANCE:

- The measure of variability used, determines the covariance, the correlation, the simple regression coefficient, the multiple regression coefficients.

- Let (X,Y) be two continuous random variables: cov(Y,X); cov(X,X)

$$\beta = \text{cov}(Y,X)/\text{cov}(X,X)$$

The GMD vs. the variance

Let (X_1, X_2) be two i.i.d. continuous random variables:

Then,

$$\sigma^2_X = 0.5\ E\{(X_1 - X_2)^2\} = cov(X,X)$$

$$\text{GMD:} \quad \Delta_X = E\{|X_1 - X_2|\} = 4\ cov(X,F(X))$$

The GMD has about 14 alternative presentations.

Only two have been used in the book

Covariances and Correlations:

$$\mathrm{Gcov}(X,Y) = \mathrm{cov}(X,G(Y)) \;;$$
$$\mathrm{Gcov}(Y,X) = \mathrm{cov}(Y, F(X)) \;;$$

The correlations which are the normalized co-Ginis are written as:

$$\Gamma_{X,Y} = \frac{\mathrm{cov}(X, G(Y))}{\mathrm{cov}(X, F(X))} \;;\; \Gamma_{Y,X} = \frac{\mathrm{cov}(Y, F(X))}{\mathrm{cov}(Y, G(Y))}\,.$$

A sufficient condition for the two Gini correlations to be equal is exchangeability up to a linear transformation.

The range of Pearson correlation coefficient

Simple Regression coefficients:

There are two regression coefficients:

1. Derived by minimization of the GMD of the error term (Called R-regression)
2. Non-parametric: Imitating Ordinary Least squares:

$$\beta^N = \frac{\mathrm{cov}(Y, F(X))}{\mathrm{cov}(X, F(X))}$$

$$\beta^O = \frac{\mathrm{cov}(Y, X)}{\mathrm{cov}(X, X)}$$

Decompositions:

Proposition 4.2.

Let $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. Then the following identities hold:

(a)

$$\Delta_Y^2 - [\beta_1 D_{1Y} \Delta_1 + \beta_2 D_{2Y} \Delta_2] \Delta_Y = \beta_1^2 \Delta_1^2 + \beta_2^2 \Delta_2^2 + \beta_1 \beta_2 \Delta_1 \Delta_2 (\Gamma_{12} + \Gamma_{21})$$

Where Γ_{ij} is Gini's correlation between X_i and X_j and $D_{iY} = \Gamma_{iY} - \Gamma_{Yi}$, i=1,2.

(b) Provided that $D_{iY} = 0$, for i=1,2, and $\Gamma_{12} = \Gamma_{21} = \Gamma$ decomposition (4.6) can be simplified into

$$\Delta_Y^2 = \beta_1^2 \Delta_1^2 + \beta_2^2 \Delta_2^2 + 2\beta_1 \beta_2 \Delta_1 \Delta_2 \Gamma.$$

Which is identical in structure to the decomposition of the variance.

ANOGI VS. ANOVA

Same case: ANOGI is identical to ANOVA provided that the distributions are stratified.

(developed in DIW) (The late Joachim Frick).

Stratification: each sub-group occupies a given range in the distribution.

Sratification: Inverse of overlapping

The simple Gini regression coefficients

Let e be the residuals. $e = Y - \beta X$.

Two regression coefficients:

(a). By minimizing the GMD of the residuals (Linearity assumption)

$$\mathrm{Cov}\,(X, F(e^G)) = 0. \quad ,$$

(b). By imitating the Ordinary Least Squares (Weighted average of slopes)

$$\mathrm{Cov}(e^N, F(X)) = 0.$$

We concentrate on (b) and use (a) for testing linearity

THE GINI NP regression coefficient

$$\beta^{N} = \frac{\mathrm{cov}(Y, F(X))}{\mathrm{cov}(X, F(X))} \quad \text{vs.} \quad \beta^{O} = \frac{\mathrm{cov}(Y, X)}{\mathrm{cov}(X, X)}$$

both can be interpreted as weighted average of slopes defined between adjacent observations (Yitzhaki, JBES, 1996).

The difference is in the weighting scheme

The Effect of transformations (change the distribution, shrink/expand the variable applied too).

Testing specification

Let $e^N = Y - \beta^N X$

Then cov $(e^N, F(X)) = 0$. by construction. (Orthogonality condition).

Testing the specification by cov $(X, F(e^N)) =? 0$.

Multiple regressions:

The Gini multiple regression coefficients are the solutions of linear equations, with the simple Gini regression coefficient serving as parameters.

One can test the specification for each independent variable.

The free use of transformations

- Assume X=1,…, 100.
- Prove that the OLS regression coefficient of log (x) on x is negatve:
- Run $\text{Log10}(x) = a + b\,x + c\,e^{0.01x} + d\,x^2 + e$.
- $\text{Log10}(x) = -13.903 - 0.089x + 14.323\,e^{0.01x} - 0.0014\,x^2$.
- Impossible under Gini

Extended Gini

- Useful in Welfare economics and Finance
- $\Delta(v) = cov(X, -[1-F(X)]^v)$
- v- a parameter chosen by the researcher $v > 0$.
- Necessary conditions for second – degree dominance.
- $M_1 \geq M_2$ and M_1 - $M_1 - \Delta_1(v) \geq M_2 - \Delta_2(v)$ are necessary conditions for distribution 1 to dominate distribution 2 for all E{U} with U'> 0; and U" < 0. (M is the mean of the distribution.
- Purpose: Either impose our social welfare function or investigate the curvature of the regression curve.

Concentration Curves

- Provide necessary and sufficient conditions for second degree stochastic dominance
- Provide necessary and sufficient conditions on whether a monotonic transformation of a variable can change the sign of correlation with another variable.
- Provides conditions for monotonic relationship between variables
- LMA (Line of Independence Minus Absolute concentration curve enables to characterize the transformation that can change the sign of an OLS regression coefficient (but not Gini regression).

Mixed Regression

- One can combine OLS and Gini regression in the same regression.
- If the sign of Gini and OLS regressions are different, we can find which variable(s) are responsible.

Multiple Regressions: Earned Income

Regression Coefficient	OLS		1		2		3		4	GINI
Age	-1.19 (0.06)	O	-1.19	O	-1.24	G	-1.05	G	-0.96	-0.96 (0.07)
Household size	11.19 (0.52)	O	11.19	O	12.59	G	13.47	G	15.87	15.87 (0.60)
Earned Income	-4.32 (0.42)	O	-4.32	G	-27.13	O	-4.40	G	-26.86	-26.86 (0.87)
Secondary school without matriculation	1.73 (3.10)	G	1.73	O	-6.01	G	1.33	O	-6.25	-6.25 (3.04)
BA degree	-16.13 (3.29)	G	-16.13	O	-0.08	G	-15.44	O	0.94	0.94 (3.25)
MA+ degree	-4.61 (3.67)	G	-4.61	O	21.79	G	-4.60	O	20.89	20.89 (4.04)
Jewish Male	-0.15 (2.12)	G	-0.15	O	16.03	G	-0.17	O	15.83	15.83 (2.12)
α(mean)	612.43		612.43		625.95		598.35		601.49	601.49
α(median)	593.67		593.67		608.54		579.88		584.41	584.41

Extensions

- The STATA procedure enables:
- Mixed OLS/Gini/Extended Gini regressions
- Extended Gini –to impose a SWF on the regression
- LMA curve of residuals to check for quality of fit.

The Future

- 50 years of competition with the variance have been lost.
- "Translating" Pittau, Zelli and Yitzhaki (2014) into a regression.
- Re-estimating Models based on OLS to Gini Methodology.
- Imposing Economic Theory (risk aversion or social welfare) on the regression.

Thank you for your patience

- Questions?