Economics Bulletin

Volume 44, Issue 1

On goodness of fit measures for Gini regression

Amit Shelef
Department of Industrial Management, Sapir Academic College, Israel

Edna Schechtman
Ben Gurion University, Beer Sheva, Israel

Abstract

The semi parametric Gini regression is more robust than ordinary least squares (OLS) regression when the underlying assumptions of the OLS fail and therefore has been used by many researchers. Several measures for goodness of fit of Gini regression were suggested in the literature. However, to the best of our knowledge, these were not compared. We examine the effect of one outlier on several goodness of fit measures in the case of a simple linear regression model via simulation. We base our comparison on the sensitivity curve. As expected, all measures under study are less sensitive to the outlier as the sample size increases. Results indicate that the least sensitive measure to an outlier is Gini correlation between the predictor Y_hat, based on Gini regression, and the observed value Y.

Paper #3 in Special issue "In memory of Pr. Michel Terraza" The authors would like to thank the Editor and the Associate Editor for the insightful comments and suggestions.
Citation: Amit Shelef and Edna Schechtman, (2024) "On goodness of fit measures for Gini regression", *Economics Bulletin*, Volume 44, Issue 1, pages 295-307
Contact: Amit Shelef - amitshe@mail.sapir.ac.il, Edna Schechtman - ednas@bgu.ac.il.
Submitted: July 23, 2021. **Published:** March 30, 2024.

Economics Bulletin

Picture credits: Virginie Terraza

Special issue “In memory of Professor Michel Terraza”

1. Introduction

Gini regression was first introduced in Olkin and Yitzhaki (1992). They proposed two Gini regressions: a parametric regression, based on the minimization of the Gini's Mean Difference (GMD) of the error term, and a semi parametric regression, which mimics the ordinary least squares (OLS). These regressions are based on using GMD as the measure of dispersion. There are more than a dozen ways to spell Gini (Yitzhaki, 1998). The relevant presentation for our paper is

$G(X) = 4COV(X, F(X))$, where $F(X)$ is the cumulative distribution function of X (Lerman & Yitzhaki, 1984).

As mentioned in Olkin and Yitzhaki (1992), Gini regression is more robust to outliers than OLS and allows for relaxing some of the traditional assumptions such as the linearity of the model and the normality of the residuals. Several authors investigated the advantages of Gini regression over OLS, focusing on the parameter estimates, on the effects of outliers and on the case of heteroscedasticity. For example, Charpentier et al. (2019) examine the robustness of the semi-parametric Gini regression to outliers and heteroscedasticity and use Pearson R^2 to examine goodness of fit. Mussard and Souissi-Benrejab (2019) suggest two Gini-PLS regressions that improve the quality of the coefficient estimates in the presence of outliers and excessive correlations between the regressors.

The issue of goodness of fit of the fitted regression model was mentioned in several papers: Olkin and Yitzhaki (1992) suggest a measure denoted by GR, based on the GMD of the error term, e, and of the dependent variable, Y (see Olkin and Yitzhaki, 1992 and Yitzhaki and Schechtman, 2013). Yitzhaki and Schechtman (2013) denote the measure by GR^2 and show that under some restrictive assumptions, this measure is equal to the ratio between the square of GMD of the predicted value $\hat{Y}$ and the square of Gini of Y. Hence it is similar in structure to Pearson's R^2, which is the commonly used measure in OLS (R^2 is the ratio between the sum of squared deviations between $\hat{Y}$ and the mean of Y, SSR, to the total sum of squares, TSS). Yitzhaki and Schechtman (2013) further suggest a version under less restrictive assumptions. This version includes Gini correlations between Y and $\hat{Y}$ (the predicted variable according to Gini semi-parametric regression) and between $\hat{Y}$ and e. In addition, they suggest two measures that are based on the covariance between (a function of) Y and (a function of) $\hat{Y}$. They note that when using OLS, the four measures are equivalent. Mussard and Ndiaye (2018) suggest a measure which is based on the Gini covariance between the error, e, and Y. Charpentier et al. (2019) use the well-known Pearson R^2 where the predicted values are obtained from their Aitken-Gini estimators of the Gini regression.

Although several goodness of fit measures appear in the literature, their properties were not discussed and to the best of our knowledge there is no consensus as to which measure to use. The objective of this note is to study (via simulation) the sensitivities of the above-mentioned measures to different underlying distributions of the independent variable and the error term and to the presence of outliers. In addition, we add our own suggestions. We focus on the semi parametric simple Gini regression (i.e., one independent variable). Because the measures are on different scales (some are in terms of squares and some are not, as will be seen below), a reasonable way to compare them is by looking at the behavior of the sensitivity curve which is a translated and rescaled version of the empirical influence function (Hampel et al., 2011; Tukey, 1970). The

influence function was used by Croux and Dehon (2010) to show that Spearman and Kendal's correlation measures are bounded when adding one outlier.

The paper is organized as follows: In section 2 we present the indices that appear in the literature and propose our own indices. Section 3 is devoted to an extensive simulation study that compares the sensitivity of the above-mentioned measures to a single outlier. Section 4 concludes.

2. Measures of goodness of fit

2.1 The existing measures

We focus on a simple linear model:

$Y = \beta_0 + \beta_1 X + \varepsilon$, where Y is the dependent variable, X is the explanatory variable and ε is the error term.

The semi parametric Gini regression estimator of β_1 is

$\hat{\beta}_1 = \frac{cov(Y,R(X))}{cov(X,R(X))}$, where $R(X)$ is the (relative) rank of X (Olkin & Yitzhaki, 1992). The intercept β_0 is estimated by

$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$ (so that the predicted line will pass through the means of the variables), and the error term is $e = Y - \hat{Y} = Y - \hat{\beta}_0 - \hat{\beta}_1 X$.

The commonly used goodness of fit measure in regression analysis is Pearson's R^2, which is the percent of variability in the dependent variable that is explained by the regression model. In Gini regression, the literature mentions several goodness of fit measures, but to the best of our knowledge their properties were not studied and there is no consensus as to which one to use.

Olkin and Yitzhaki (1992) introduced Gini regression and suggested the following measure of goodness of fit:

$$GR = 1 - \left(\frac{cov(e,R(e))}{cov(Y,R(Y))}\right)^2, \quad (1)$$

where $R(e)$ is the (relative) rank of $e = Y - \hat{Y}$. The intuitive reason for this suggestion lies in the fact that in OLS, the total sum of squares (*TSS*) can be decomposed into the sum of two sums of the squares: Sum of squares due to regression (*SSR*) and error sum of squares (*SSE*), where

$TSS = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$,

$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$,

$SSE = \sum_{i=1}^{n} (e_i)^2$,

$TSS = SSR + SSE$,

and Pearson's R^2

$$R^2 = \frac{SSR}{TSS} = 1 - \frac{SSE}{TSS}. \quad (2)$$

Note that equations (1) and the right-hand side of (2) are similar in structure. However, as will be explained below, the equivalent of the middle term in (2) does not apply to (1) in general, because in the Gini regression, an equivalent (in structure) decomposition does not hold. As mentioned in Yitzhaki and Schechtman (2013), equation 7.41, the decomposition

$$G^2(Y) = G^2(\hat{Y}) + G^2(e) \tag{3}$$

(which is the motivation for the similarity in structure of equations (1) and (2)) holds true **only** under some restrictive assumptions on Gini correlations and on the model. Recall that Gini correlation between X and Y is defined as $\Gamma_{XY} = \frac{COV(X,F(Y))}{COV(X,F(X))}$ (Schechtman & Yitzhaki, 1987) and that the two Gini correlations Γ_{XY} and Γ_{YX} are not necessarily equal. The restrictive assumptions for (3) to hold are:
1. The two Gini correlations between e and $\hat{Y}$ are zero ($\Gamma_{e\hat{Y}} = \Gamma_{\hat{Y}e} = 0$), i.e., the model specification is correct.
2. The two Gini correlations between Y and $\hat{Y}$ are equal ($\Gamma_{Y\hat{Y}} = \Gamma_{\hat{Y}Y}$),
and
3. The two Gini correlations between Y and e are equal ($\Gamma_{Ye} = \Gamma_{eY}$).
(See Yitzhaki and Schechtman (2013) for details).

We note in passing that if the distribution of ($\hat{Y}$, e) is bivariate normal, then their sum, Y, is normally distributed, implying that Y and $\hat{Y}$ are exchangeable up to a linear transformation. In this case, assumptions 2 and 3 above are met.

Under the three assumptions above, Yitzhaki and Schechtman (2013) define an R^2-equivalent by

$$GR^2 = \left(\frac{G(\hat{Y})}{G(Y)}\right)^2 = 1 - \left(\frac{G(e)}{G(Y)}\right)^2$$

(Yitzhaki and Schechtman, 2013, equation 7.42).

Note that GR^2 is the same as GR in Olkin and Yitzhaki (1992), except for the middle equation, which is not mentioned in Olkin and Yitzhaki (1992). Olkin and Yitzhaki (1992) claim that the range of GR is [0,1], but as will be shown later by an example, GR is not bounded from below. Therefore, denoting it as a square is a bit misleading.

If one only assumes that the specification of the model is correct (assumption 1), then additional terms appear in the decomposition of the square of GMD of Y:

$$G^2(Y) = \left(D_{Y\hat{Y}}G(\hat{Y}) + D_{Ye}G(e)\right)G(Y) + G^2(\hat{Y}) + G^2(e)$$

(Yitzhaki and Schechtman, 2013, equation 7.40)

and the general form of GR^2 is

$$GR^2 = \left(\frac{G(\hat{Y})}{G(Y)}\right)^2 = 1 - \left(\frac{G(e)}{G(Y)}\right)^2 - \frac{D_{Y\hat{Y}}G(\hat{Y}) + D_{Ye}G(e)}{G(Y)}, \tag{4}$$

where D_{XY} represents the difference between two Gini correlations of two variables, X and Y, namely: $D_{XY} = \Gamma_{XY} - \Gamma_{YX}$,

(Yitzhaki and Schechtman, 2013, equation 7.43).

Note that the third term on the right-hand side of (4) can be positive or negative. This fact can cause the measure to be negative, as will be shown in the example below. Obviously, there is an advantage to have a measure with a bounded range, but because the criterion for a good model is the closeness of the measure to 1, we are less concerned about negative numbers.

Yitzhaki and Schechtman (2013) suggest two additional measures:

$$\Gamma_{Y\hat{Y}} = \frac{cov(Y,F(\hat{Y}))}{cov(Y,F(Y))},\ \Gamma_{\hat{Y}Y} = \frac{cov(\hat{Y},F(Y))}{cov(\hat{Y},F(\hat{Y}))} \tag{5}$$

(Yitzhaki and Schechtman, 2013, equation 7.44)

We note in passing that in OLS, R^2 (equation (2)) and the squares of the two Pearson correlations $corr(Y,\hat{Y})$ and $corr\ (\hat{Y},Y)$ (the equivalents of (5)) are all equal to the square of the correlation between *X* and *Y*.

Mussard and Ndiaye (2018) deal with the multiple regression case and suggest a similar measure of goodness of fit.

Using $Y = \hat{Y} + e$, they start with

$$1 = \frac{cov(\hat{Y},R(Y))}{cov(Y,R(Y)} + \frac{cov(e,R(Y))}{cov(Y,R(Y))}.$$

Hence, their suggested measure is

$$GR^2_{MN} = 1 - \left(\frac{cov(e,R(Y))}{cov(Y,R(Y))}\right), \tag{6}$$

which is the slope of the regression of $\hat{Y}$ on Y.

Note that this measure is similar to the one in the right-hand side of (5), suggested by Yitzhaki and Schechtman (2013). The difference is that Mussard and Ndiaye (2018) use GMD of *Y* in the denominator, while Yitzhaki and Schechtman (2013) use GMD of $\hat{Y}$. Charpentier et al. (2019) deal with the multiple regression case. They propose a Gini-White test and use Pearson R^2 in their Monte Carlo simulations, where the predicted value is obtained by Gini regression. They make some comparisons between the generalized least squares and the Gini regression and show that a better power is obtained by Gini-White test compared with the usual White test when outlying observations contaminate the data.

2.2 Our suggestion

In this section we describe two intuitive suggestions. Unfortunately, our preliminary simulation study showed that they are not competitive with the existing measures. Moreover, one of them is not bounded from above by 1. Therefore, their performances are not shown in the extensive simulation section below (except for an example where the measure exceeds 1 which appears in Appendix B). However, we think that it is worth mentioning them, as they seem to be natural competitors.

Intuitively, a natural way to evaluate the goodness of fit is to base it on the difference between the predicted value (based on the model and the estimation procedure) and the mean of *Y*. The rational is: without any information from *X*, the natural prediction is the mean of *Y*. The difference reflects the advantage we get from using *X* and fitting the model.

In OLS, the partitioning of the sum of squares of Y is well known:

TSS=$\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2+\sum_{i=1}^{n}(e_i)^2$, where $\hat{Y}_i$ and e_i are the predicted values and residuals obtained by OLS regression, respectively.

That is, the cross term is equal to 0.

Hence, the natural measure of goodness of fit is the ratio

$R^2=\frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$.

The first intuitive measure in the case of Gini regression would be to use R^2, where $\hat{Y}$ is calculated by Gini regression. Unfortunately, this measure is not bounded from above by 1, hence it cannot serve as a good measure. See an example in the simulation section. The next step is to try to imitate the decomposition of TSS. Unfortunately, when using Gini regression, the partitioning is not so straight forward.

$TSS_G=\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2+\sum_{i=1}^{n}(e_i)^2+2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})e_i$,

where $e_i=Y_i-\hat{Y}_i$ and $\hat{Y}$ is calculated by Gini regression.

In other words, it contains the sum of squares of the residuals, plus two terms that add up to the contribution of the regression. Therefore, our second suggested measure is

$$R_g=\frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2+2\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})e_i}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}. \quad (7)$$

It is easy to see that this measure is bounded from above by 1. We note that the measure can be negative (as are the measures in (1) and (4)), but as mentioned above, the criterion for a good model is the closeness (from below) of the measure to 1. Therefore, we are less concerned about negative numbers. Our preliminary simulation study shows that R_g is not a good candidate. Therefore, it is not included in the simulation section below.

3. Simulation study

An extensive simulation study was conducted in order to study the sensitivity of the above-mentioned measures to outliers. The measures that are included are: Pearson's R^2, where $\hat{Y}$ is calculated by OLS, Olkin and Yitzhaki's GR, GR^2_{MN}, $\Gamma_{Y\hat{Y}}$ and $\Gamma_{\hat{Y}Y}$. Pearson's R^2, where $\hat{Y}$ is calculated by Gini regression is not bounded from above by 1, hence it is not included in the study. The data was generated from a simple linear regression model $Y=a+bX+\varepsilon$, where both X and ε have normal distributions. In order to study the effect of an outlier, we investigated 6 cases: An outlier was added at the upper (lower) end of the range of X and around the middle of the range. In each location, the outlier was added above or below the regression line. For each case we repeated the generating process K=1000 times and calculated the above-mentioned measures for each replication. Because the measures are on different scales (some are presented in terms of squares and some are not, some are in a bounded range and some are not), we chose the criterion for comparison to be the sensitivity curve, based on K replications. We used Tukey's sensitivity curve (Hampel et al., 2011; Tukey, 1970) as follows: Suppose we have an estimator $\{T_n; n\geq 1\}$ and a sample $(x_1,\dots,x_{n-1})$ of $n-1$ observations. Then the sensitivity curve (SC), which is a translated and rescaled version of the empirical influence function, is defined as

$$SC_n(x) = n[T_n(x_1, \ldots, x_{n-1}, x) - T_{n-1}(x_1, \ldots, x_{n-1})], \quad (8)$$

where x is the outlier. See Hampel et al. (2011), equation 2.1.21 for details.

We now turn to a detailed description of the simulation. In the first step of the simulation we generated n values (n=20, 50, 100, and 200) of X from a Normal(0,1) distribution and n values of ε from a Normal(0,0.2^2) and Normal(0,0.6^2) distribution. Then, we calculated the dependent variable according to a linear model: $Y = 2 + 3X + \varepsilon$. These n pairs of (X,Y) are used to calculate the second term of the right-hand side of Equation (8). In the second step we added outliers, one at a time, to calculate the first term of the right-hand side of Equation (8). We looked at 6 scenarios: an outlier is added at the left-side of the range of X (two cases: the outlier lies above or below the line), right side (two cases) and near the middle (two cases). More specifically, since X is Normal(0,1), we chose 3 values of X: (-2, 0, and 2) which represent the left, middle and right sides of the range of X, respectively. In order to create a SC for each location, we generated 31 new values, starting at the point on the line (that is, calculate Y by $Y = 2 + 3X$, for $X = -2, 0, 2$) and adding (or subtracting) 0,1, 2 , …, 30 to the result. Notice that the first value of each of the resulting SCs (in which 0 is added or subtracted) represents the case without contamination. For example, for the bottom left (x=-2) when the outlier is added below the regression line, the outliers for the SC were {-4 (on the line, to represent the case without contamination), -5, -6, …, -34}. For each case we calculated the 5 measures: Pearson's R^2, where $\hat{Y}$ is calculated by OLS, Olkin and Yitzhaki's GR, GR^2_{MN}, $\Gamma_{Y\hat{Y}}$ and $\Gamma_{\hat{Y}Y}$. The first and second steps were repeated K=1000 times. Each point on the SC represents an average of 1000 runs.

Results for sigma=0.2 and sigma=0.6 are similar, hence we will focus on sigma=0.2 from now on.

The effect of the location of the outlier. As expected, the bottom (below the line) left and upper (above the line) right locations give similar (mirrored) results due to the symmetry. Similarly, the upper left and bottom right give similar results. Also, bottom and upper middle are similar. Hence we will concentrate only on three cases: (a) bottom left, (b) upper left and (c) bottom middle. Figure 1 illustrates the sensitivity curves for the 5 measures for these cases ((a)-(c)) by sample size (in order to illustrate the symmetry between the cases, all 6 cases are reported in appendix A for n=50). The horizontal axis of each plot depicts the value of the outlier ({-4, -5, …, -34} in the example above) and the vertical axis depicts the sensitivity as defined in (8), averaged over 1000 runs. (We ignore the multiplication by n in (8), since it is irrelevant for our comparison which is based on the same n in each SC).

In general, as can be seen, in all cases Pearson OLS (R^2) increases with the size of the outlier (in absolute value), but the values of the sensitivity curve decrease as n gets larger. (Note the decreasing (with n) scale on the vertical axes). This is not surprising. Devlin et al. (1975) showed that the influence function of the classical Pearson correlation is unbounded, proving the lack of robustness of this measure. Also, as expected, all measures are less sensitive to an outlier as the sample size increases, because the percent of affected observations decreases. One outlier out of 20 is different from one out of 200.

In case (a) - bottom left, outliers below the regression line, Olkin and Yitzhaki's GR is affected by the outliers for n=20, and the effect decreases as n gets larger. The other three measures are (almost) not affected by the size of the outliers, even for n=20.

Cases (b) (upper left, outliers are above the regression line) **and (c)** (bottom middle, outliers below the regression line) are similar, but different from case (a). At n=20 all measures except $\Gamma_{\hat{Y}Y}$ are sensitive to outliers. As n gets larger, we see two pairs. While GR and $\Gamma_{\hat{Y}Y}$ stabilize as the size of the outlier increases, the other two, GR^2_{MN} and $\Gamma_{Y\hat{Y}}$ are more affected.

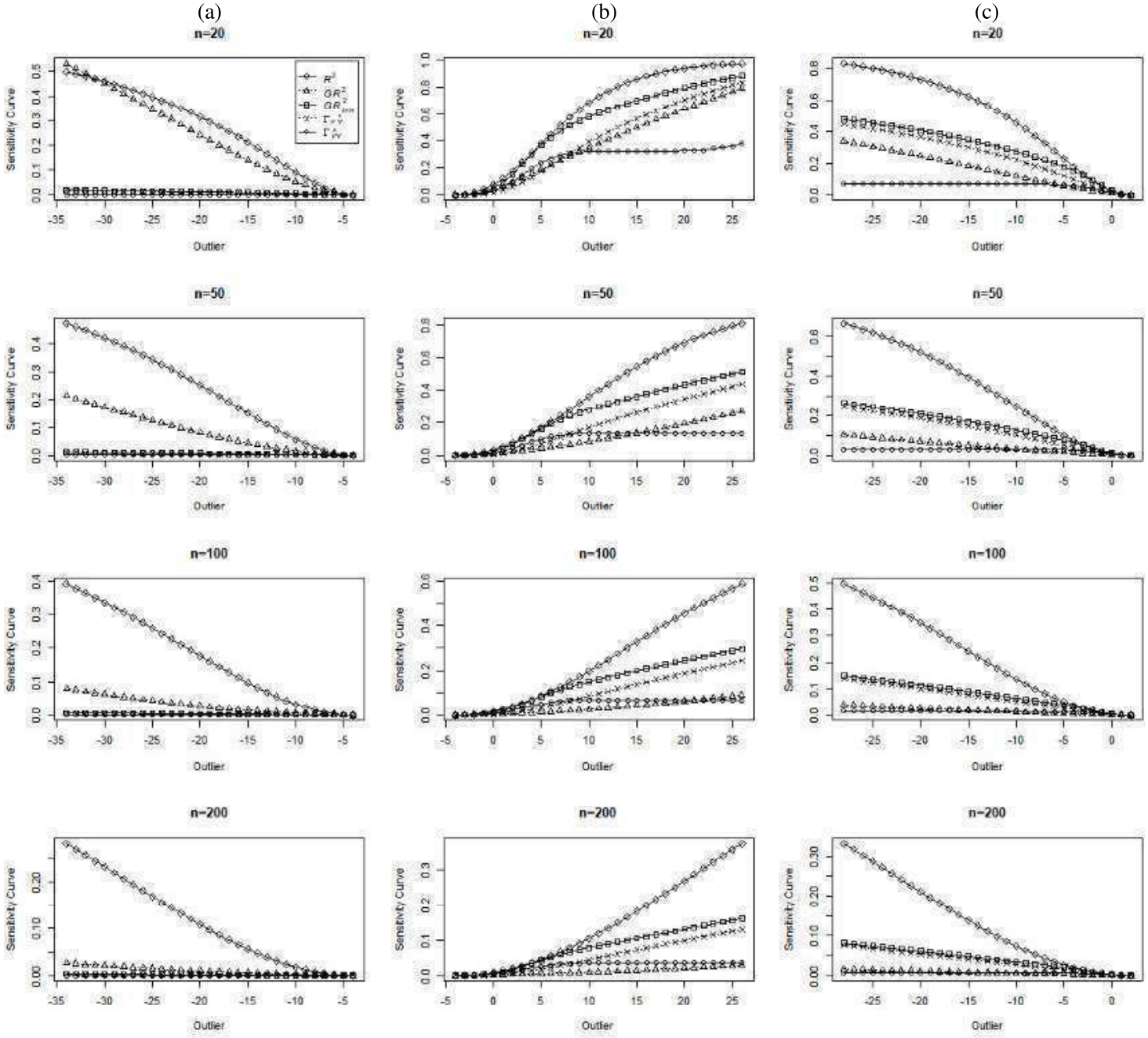

Figure 1. Sensitivity curves for the three cases (on columns (a) bottom left, (b) upper left and (c) bottom middle), by sample size (n=20, 50, 100, 200)

To summarize, based on our simulations we recommend the use of $\Gamma_{\hat{Y}Y}$. One possible explanation for its superiority is that in this measure Y is taken in its ranks, which in turn diminishes the effect of the outlier.

For completeness, we list two examples in Appendix B. In example 1, R^2, where $\hat{Y}$ is calculated by Gini regression (introduced in Section 2.2) is equal to 1.01, and in example 2, $\mathrm{GR} = -0.13$, $GR^2_{MN} = -0.06$ and $\Gamma_{\hat{Y}Y} = -0.37$. Obviously, it is preferred to have a measure which is bounded between 0 and 1. Exceeding 1 is not acceptable. However, negative values do not interfere because we are mainly interested in a good fit, that is, the closer to 1 – the better.

4. Conclusions

Gini regression was first introduced in 1992 by Olkin and Yitzhaki (Olkin and Yitzhaki, 1992). Since then it has been used by many researchers. The advantage of Gini regression over ordinary least squares was studied extensively, but to the best of our knowledge, there is no consensus as to how to evaluate its goodness of fit. There are several suggestions in the literature, but no comparison between those measures has been published.

In this paper we study (via simulation) the effect of one outlier on the goodness of fit measures (suggested in Olkin and Yitzhaki's (1992), Mussard and Ndiaye (2018), and Yitzhaki and Schechtman (2013)). We start with a simple linear regression model and add outliers, one at a time, at three locations: left end of the X-range, right end and the middle of the range. At each location we look at outliers above the line and below it. All together we look at 6 cases. Due to symmetry, three cases are discussed in detail. In addition, we look at the effect of the sample size on the results. We start with a small sample (n=20), and increase to n=50, 100 and 200. The criterion for comparison is the sensitivity curve (Hampel et al., 2011; Tukey, 1970).

As expected, as the sample size gets bigger, the effect of an outlier becomes smaller (because one outlier out of 20 observations has more effect than one out of 200). It turns out that $\Gamma_{\hat{Y}Y}$ (the Gini correlation between the Gini regression predictor ($\hat{Y}$) and the observed value, Y) is less sensitive to an outlier than the rest. One possible explanation is that only in this measure, Y appears only via $F(Y)$ (its cumulative distribution function). Therefore, the outlier appears in this measure only via its rank, which makes it robust to outliers. We note in passing that the measures suggested in Olkin and Yitzhaki's (1992), Mussard and Ndiaye (2018), and the second Gini correlation suggested in Yitzhaki and Schechtman (2013) can obtain negative values. Obviously, a bounded measure would be preferred, but because we are mainly interested in finding a good fit (i.e., values close to 1), negative values do not interfere.

Declarations

Funding: This research received no external funding.
Compliance with Ethical Standards:
Conflict of Interest: The authors declare that they have no conflict of interest.
Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Appendices

Appendix A - the sensitivity curve for a single outlier at the following 6 locations: (1) bottom left, (2) upper left, (3) bottom middle, (4) upper middle, (5) bottom right, and (6) upper right, for *n*=50

Figure A.1 illustrates the sensitivity curves for the 5 measures, with *n*=50, for outliers at 6 locations: (1) bottom left, (2) upper left, (3) bottom middle, (4) upper middle, (5) bottom right, and (6) upper right

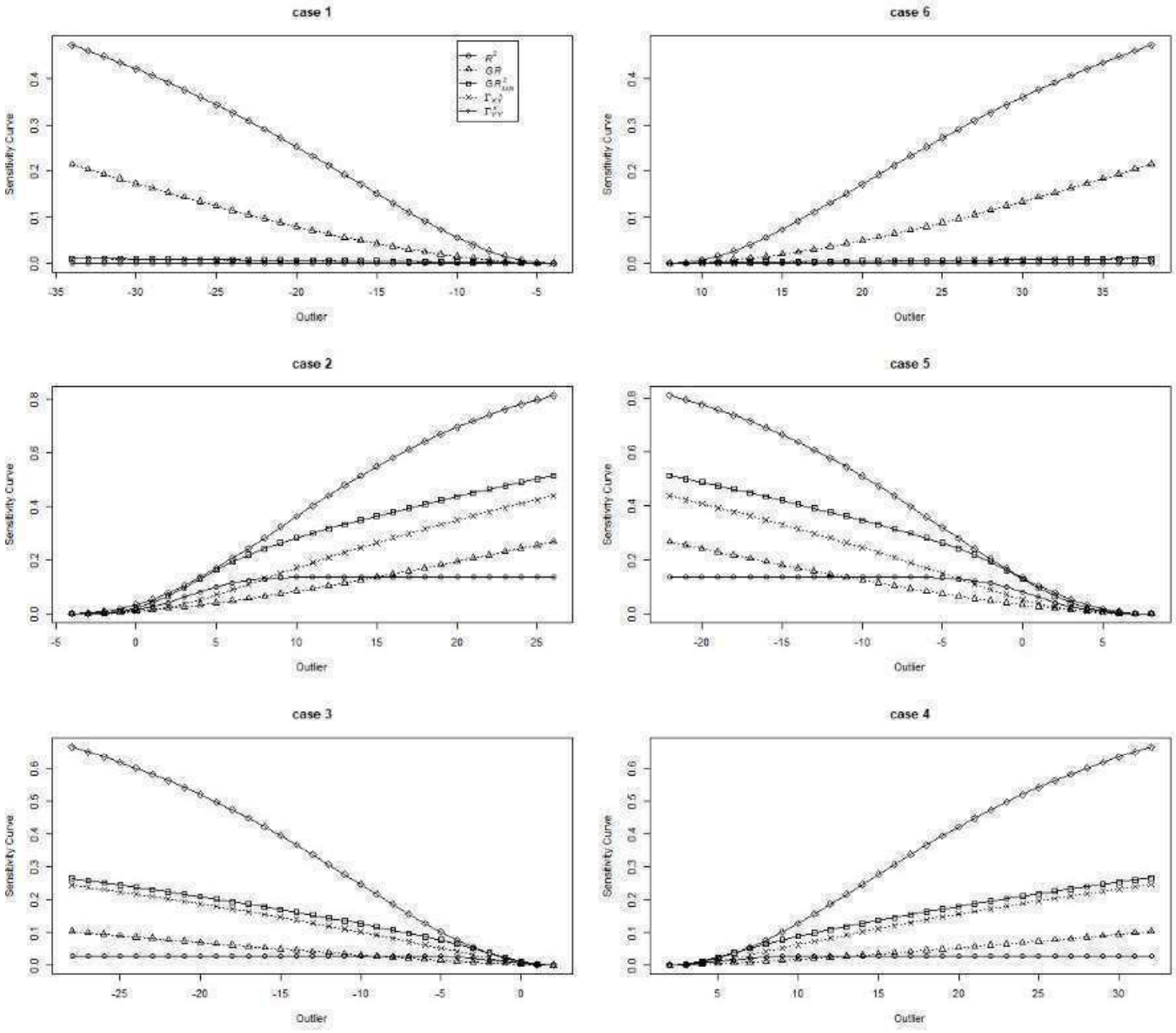

Figure A.1. Sensitivity curves for 6 cases (*n*=50): (1) bottom left, (2) upper left (3) bottom middle, (4) upper middle, (5) bottom right, and (6) upper right

Appendix B – Numerical examples

In Table A.1. we list two numerical examples. In example 1, the Pearson-based R^2 , where $\hat{Y}$ is calculated by Gini regression, is equal to 1.01. In example 2, some of the measures are negative: $\mathrm{GR} = -0.13$, $GR^2_{MN} = -0.06$ and $\Gamma_{\hat{Y}Y} = -0.37$.

Table A.1. Two numerical examples

	Example 1			Example 2	
	x	y		x	y
1	-1.309	-3.299	1	-0.393	-0.728
2	0.815	6.822	2	1.271	5.948
3	-1.112	-3.225	3	0.865	3.045
4	0.910	5.301	4	-0.095	3.320
5	-0.849	-2.394	5	-0.422	0.347
6	-0.243	2.379	6	0.355	3.418
7	-0.002	0.234	7	0.103	3.494
8	-0.634	-0.613	8	-0.600	1.349
9	-0.409	0.519	9	-1.084	9.191
10	1.223	7.296	10	0.765	4.550
11	-0.074	2.087	11	-1.221	2.648
12	-1.762	-1.444	12	0.487	3.079
13	0.547	5.204	13	0.829	4.545
14	0.503	3.824	14	-2.209	-3.566
15	0.191	1.909	15	1.242	5.242
16	0.481	2.204	16	0.717	3.808
17	-0.221	1.061	17	1.523	6.976
18	-2.332	-4.416	18	0.110	2.479
19	-0.428	-0.026	19	0.233	-0.814
20	2.945	9.150	20	-0.893	1.681
21	2.000	8.000	21	2.000	-22.000

References

Charpentier, A., Ka, N., Mussard, S., & Ndiaye, O. H. (2019). Gini Regressions and Heteroskedasticity. *Econometrics*, ***7*****(1)**, 4.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, ***19*****(4)**, 497–515.

Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika*, ***62*****(3)**, 531–545.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions*. John Wiley & Sons.

Lerman, R. I., & Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters*, ***15*****(3–4)**, 363–368. https://doi.org/10.1016/0165-1765(84)90126-5

Mussard, S., & Nadiaye, O. H. (2018). Vector autoregressive models: a Gini approach. *Physica A: Statistical Mechanics and Its Applications*, ***492***, 1967–1979.

Mussard, S., & Souissi-Benrejab, F. (2019). Gini-PLS Regressions. *Journal of Quantitative Economics*, ***17*****(3)**, 477–512.

Olkin, I., & Yitzhaki, S. (1992). Gini regression analysis. *International Statistical Review*, ***60***, 185–196.

Schechtman, E., & Yitzhaki, S. (1987). A measure of association based on Gini mean difference. *Communications in Statistics-Theory and Methods*, ***16*****(1)**, 207–231.

Tukey, J. W. (1970). *Exploratory data analysis: Limited preliminary Ed.* Addison-Wesley Publishing Company.

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

Yitzhaki, S., & Schechtman, E. (2013). *The Gini methodology*. Springer.