Ensayos sobre POLÍTICA ECONÓMICA



STATISTICAL INFERENCE FOR TESTING GINI COEFFICIENTS: AN APPLICATION FOR COLOMBIA

Luis Fernando Gamboa Andrés García-Suaza Jesús Otero

ENSAYOS SOBRE POLÍTICA ECONÓMICA, VOL. 28, NÚM. 62 EDICIÓN JUNIO 2010 PP. 226-238

Los derechos de reproducción de este documento son propiedad de la revista Ensayos Sobre Política Económica (*ESPE*). El documento puede ser reproducido libremente para uso académico, siempre y cuando no se obtenga lucro por este concepto y además, cada copia incluya la referencia bibliográfica de *ESPE*. El(los) autor(es) del documento puede(n) además poner en su propio website una versión electrónica del mismo, pero incluyendo la referencia bibliográfica de ESPE. La reproducción de esta revista para cualquier otro fin, o su colocación en cualquier otro website, requerirá autorización previa de su Editor de *ESPE*.

Inferência e testes estatísticos sobre o Coeficiente de Gini: Uma Aplicação para a Colômbia

Luis Fernando Gamboa Andrés García-Suaza Jesús Otero*

* Faculdade de Economía Universidade do Rosario Os autores agradecem os comentários e sugestões de Russell Davidson, Luis Eduardo Fajardo, Ana Maria Iregui(Editor), Jeremy Smith assim como os provenientes do avaliador anônimo. Os erros e omissões são responsabilidade dos autores.

Correios eletrônicos: luis.gamboa@urosario. edu.co; andres.garcia@ urosarioedu.co; jesus. otero@urosario.edu.co

Documento recebido no dia fevereiro 23,2010; versão final aceitada no dia maio 18, 2010. Este artigo estima coeficientes de Gini para a Colômbia, com seus correspondentes erros padrão, utilizando informação proveniente das Pesquisas de Opinião de Lares durante o período 1984-2005. Encontra-se um incremento estatísticamente significativo na desigualdade do salário por hora no período posterior à abertura econômica do começo da década dos noventa; para os anos posteriores à recessão de finais dos anos noventa, a evidência não é concluinte. Também se encontra que as diferenças observadas nos indicadores de desigualdade entre cidades não foram estatísticamente significativa em vários casos.

Classificação JEL: C12; D31; I32.

Palavras chave: Desigualdade, Coeficiente de Gini, bootstrap, Colômbia.

Inferencia y pruebas estadísticas sobre los Coeficientes de Gini: Una aplicación para Colombia

Luis Fernando Gamboa Andrés García-Suaza Jesús Otero*

Este artículo usa información recolectada a través del Sistema de Encuestas de Hogares en Colombia para el periodo 1984-2005 con el fin de estimar coeficientes de Gini y sus errores estándar correspondientes. Encontramos un aumento estadísticamente significativo en la medida de desigualdad salarial, consecuencia de las medidas de liberalización económica adoptadas al comienzo de los años noventa, así como evidencia mixta durante los años de recuperación que siguieron a la recesión económica de finales de esta misma década. Además, encontramos que en muchos casos las variaciones observadas entre los coeficientes de Gini de las diferentes ciudades y a través del tiempo no son significativas en términos estadísticos.

Clasificación JEL: C12; D31; I32.

Palabras clave: inequidad, coeficiente de Gini, bootstrap, Colombia.

* Facultad de Economía, Universidad del Rosario. Bogotá, Colombia.

Agradecemos a Russell Davidson, Luis Eduardo Fajardo, Ana María Iregui (Editora), Jeremy Smith y a un evaluador anónimo por sus comentarios y sugerencias.

Correos electrónicos: luis.gamboa@urosario. edu.co; andres.garcia@ urosarioedu.co; jesus. otero@urosario.edu.co Documento recibido el 23 de febrero de 2010; versión final aceptada el 18 de mayo de 2010.

Luis Fernando Gamboa Andrés García-Suaza Jesús Otero*

* Faculty of Economics, Universidad del Rosario, Bogotá, Colombia.

We would like to thank Russell Davidson, Luis Eduardo Fajardo, Ana María Iregui (Editor), Jeremy Smith and an anonymous referee for their useful comments and suggestions. The usual disclaimer applies.

E-mails: luis.gamboa@ urosario.edu.co; andres. garcia66@urosario.edu. co; jesus.otero@urosario. edu.co (corresponding author)

Document received: 23 february 2010; final version accepted: 18 may 2010. This paper uses Colombian household survey data collected over the 1984-2005 period to estimate Gini coefficients and their corresponding standard errors. We find a statistically significant increase in wage income inequality following the adoption of the liberalization measures during the early 1990s, and mixed evidence from the recovery years that followed the economic recession during the late 1990s. We also find that in several cases the observed differences in the Gini coefficients across cities have not been statistically significant.

JEL classification: C12; D31; I32

Keywords: inequality, Gini coefficient, bootstrap, Colombia.

I. INTRODUCTION

Measuring the evolution of income distribution over time and/or across regions and assessing the effect of policy measures on income concentration are topics of research that have historically received a great deal of attention. To address these topics, authors typically provide comparisons based on the ranking of estimated Gini coefficients without acknowledging the fact that, being a sample statistic, these coefficients have associated sampling distributions; see e.g., Baer and Maloney (1997), Caselli and Battini (2000), and Ezcurra and Pascual (2009).

A number of authors have considered different methodologies to estimate the standard error of the Gini coefficient: Zheng and Cushing (2001), Giles (2004 and 2006), Ogwang (2000, 2004 and 2006), and Modarres and Gastwirth (2006). However, in a recent paper Davidson (2009) points out that the estimators available in the literature are either mathematically complex to calculate or quite unreliable. For example, Davidson (2009) shows that the jackknife estimator of the variance is not a consistent estimator of the asymptotic variance of the Gini coefficient, and therefore does not give reliable inference. Davidson (2009) presents a procedure to compute an asymptotically correct standard error for the Gini coefficient based on a relatively simple expression. The work by Davidson has at least three main contributions. First, it provides a bias-corrected estimator of the Gini coefficient. Second, it derives an approximation for the standard error of the Gini coefficient that expresses it as a sum of independent and identically distributed (iid) random variables. Third, it illustrates how bootstrap methods can be used to yield reliable inference about the Gini coefficient.

This paper uses Colombian household survey data collected over the 1984-2005 period to estimate the Gini coefficient for labor income in the urban labor market, as well as for the labor markets in the main seven urban areas. Rankings of Gini coefficients based on income distributions for Colombia have been undertaken by Berry and Urrutia (1976) and Birchenall (2001, 2007), among others. In sharp contrast to this literature, in this paper we estimate standard errors on Gini coefficients. This enables us to test for statistical variation across urban areas and over time. The chosen sample period is interesting because the Colombian government instituted a series of major liberalizing reforms during the early 1990s, although this was followed by the deepest recession experienced by the country in the last century and the subsequent years of recovery.

The paper is organized as follows: Section II briefly describes the methodology used for the estimation of the Gini coefficient and its corresponding standard error. Section III describes the data set and summarizes the main results. Section IV concludes.

II METHODOLOGY

The Gini coefficient, defined as twice the area between the equidistribution line (i.e., the 45° line) and the Lorenz (1905) curve, is perhaps the most commonly used measure of inequality. It ranges between zero (perfect equality) and one (perfect inequality). Recently, Davidson (2009) expressed the Gini coefficient as:

$$\hat{G} = \frac{2}{\hat{\mu}n^2} \sum_{i=1}^n y_{(i)} \left(i - \frac{1}{2} \right) - 1, \tag{1}$$

where, $y_{(i)}$, i = 1, 2, ..., n, is the series of order statistics of the income variable y (that is, the original series sorted in increasing order), and $\hat{\mu}$ is the estimated mean of y. Davidson (2009) finds an approximate expression for the bias of \hat{G} , from which he derives the following bias-corrected estimator of the Gini coefficient, denoted \tilde{G} , which is given by:

$$\tilde{G} = \frac{n}{(n-1)}\hat{G}.$$
⁽²⁾

While the estimator (2) is still biased, its bias is of order smaller than n^{-1} . Equation (2) can be used to obtain an estimate of the standard error of \tilde{G} . Using:

$$\tilde{Z}_i = -(\tilde{G} + 1)y_{(i)} + 2(w_i - v_i),$$
(3)

where $w_i = (2i - 1) y_{(i)} / (2n)$ and $v_i = n^{-1} \sum_{j=1}^{i} y_{(j)}$. The standard error of the bias-corrected Gini coefficient is denoted as:

$$SE\left(\tilde{G}\right) = \sqrt{\frac{1}{(n\hat{\mu})^2} \sum_{i=1}^n (\tilde{Z}_i - \bar{Z})^2}.$$
(4)

Davidson (2009) shows, via simulation experiments, that the asymptotic distribution of the Gini coefficient is reliable even for sample sizes of around 100 observations. However, in case the underlying income distribution follows a lognormal distribution with a large variance or when the distribution has heavy tails, reliable inference can be obtained by applying the bootstrap method. In particular, Davidson (2009) suggests implementing the bootstrap method as follows. First, let

$$\tau \equiv \frac{(\tilde{G} - G_0)}{SE\left(\tilde{G}\right)},\tag{5}$$

be the test statistic required to test the null hypothesis that the bias-corrected Gini coefficient is equal to G_0 . Then, one generates b = 1, ..., B bootstrap samples of size n by resampling with replacement from the observed income data (which is also of size n). For bootstrap sample b, one computes a bootstrap statistic τ_b^* as in (5), but with G_0 replaced by \tilde{G} , that is the value of the statistic computed from the observed sample. This is required so that the hypothesis tested should be true of the bootstrap data-generating process. To calculate an interval at nominal confidence level $(1 - \alpha)$, one estimates the $\alpha/2$ and $1 - \alpha/2$ quantiles of the empirical distribution of the bootstrap statistics τ_b^* .

III DATA AND MAIN RESULTS

We use data from the nationwide household surveys periodically undertaken by the Departamento Administrativo Nacional de Estadística (DANE). Our period of

analysis, i.e., 1984 to 2005, is characterized by the implementation of two different surveys, namely the Encuesta Nacional de Hogares (ENH) and the Encuesta Continua de Hogares (ECH). The former was applied quarterly from 1979 to 2000, and up until 1983 it included the four main cities of Colombia: Bogotá, Medellín, Cali, and Barranquilla. In 1984 three more cities were added to the ENH: Bucaramanga, Manizales, and Pasto. In 2001, the ENH was superseded by the ECH, which is a monthly survey of thirteen cities: the original seven plus Ibagué, Montería, Cartagena, Pereira, Villavicencio, and Cúcuta¹.

The dataset used in the analysis consists of the hourly wage per worker (in constant prices of 2005) during the 1984-2005 period, which is used as a proxy for labor income. The choice of this variable has three important implications for the analysis. First, hourly wage per worker exhibits less variation than personal income. Indeed, over the period from 1984 to 2005 the coefficient of variation of the former ranges from 1.1 to 2.4, while the corresponding coefficient of variation of the latter ranges from 1.2 to 3.2. This implies that Gini coefficients based on hourly wage per worker will be lower than those based on personal income. Second, calculations not reported here suggest that problems of censored and truncated data are more frequent when dealing with personal income than with our measure of hourly wage per worker. For example, in 2001 and 2002 the percentage of individuals who do not report income amounts to approximately 43% and 44%, respectively, while the corresponding percentages for the individuals who do not report wage are 19% and 21%, respectively. Third, the use of hourly wage per worker offers the advantage that it controls for the fact that an individual may earn several wages from different jobs.

The unit of analysis is the employed individuals. This means that individuals who report having worked during the previous week but do not report labor income are excluded from the sample. One might be inclined to think that individuals in the upper tail of the income distribution do not tend to report their income. However, results not reported here indicate that there is no statistical difference between individuals who report labor income and those who do not report it, once one controls for human capital variables, such as education and experience. Furthermore, given that in this paper we are interested in assessing changes in Gini coefficients over time and across cities, it is likely that underreporting will be randomly distributed within the sample.

¹ The ECH also introduced changes in the phrasing of questions aimed at measuring labor market indicators, such as the concept of unemployment, and unpaid workers, etc. These methodological differences do not affect our measure of wage income.

The data for each year in the 1984-2005 period were obtained by aggregating the surveys of every given year. We use the seven main cities that are available throughout the sample period: Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Manizales, and Pasto. These account for more than seventy percent of the country's total urban population.

Table 1 reports bias-corrected Gini coefficients for the main seven cities and for the country².

This table also contains the corresponding standard errors based on equation (4), which can be used to construct confidence intervals using the quantiles of the standard normal distribution. The standard errors that result from implementing the bootstrap procedure outlined in the previous section (using 9.999 bootstrap replications) are reported in the Appendix. At this point it is also worth mentioning that the application of the jackknife method results is much larger estimates of the variance of the bias-corrected Gini coefficients; indeed, when using the data for all seven cities the estimated jackknife variance is almost 1.8 times the estimated asymptotic variance derived by the formula given in Davidson (2009)³.

Table 2 reports the number of times the bias-corrected Gini coefficients between pairs of cities are statistically the same over the sample period under consideration, 1984-2005. For example, when looking at Bucaramanga and Barranquilla, in 11 out of the 21 possible cases the coefficients between these two cities do not appear to be statistically different. This table shows that there are only three pairs of cities, namely Bogotá-Medellín, Medellín-Pasto and Bucaramanga-Pasto, for which the estimated coefficients always appear to be statistically different throughout the sample period.

² All the calculations were performed in the econometrics software RATS 6.1 and Stata SE 9.2.

³ In the case of the city of Pasto, the estimated jackknife variance is almost 3 times the estimated asymptotic variance. Jackknife estimates of the standard errors are not reported here for brevity, but are available from the authors upon request.

Table 1 Estimates and standard errors of the Gini coefficient

Voor	Total		Bogotá		Barrar	quilla	Bucarar	nanga	
Year	Gini	(s.e.)	Gini	(s.e.)	Gini	(s.e.)	Gini	(s.e.)	
1984	0.405	0.0021	0.413	0.0028	0.396	0.0087	0.387	0.0075	
1985	0.420	0.0032	0.429	0.0043	0.408	0.0076	0.400	0.0076	
1986	0.406	0.0033	0.428	0.0048	0.411	0.0070	0.379	0.0060	
1987	0.389	0.0024	0.416	0.0039	0.363	0.0062	0.376	0.0053	
1988	0.396	0.0022	0.432	0.0045	0.355	0.0053	0.396	0.0055	
1989	0.398	0.0032	0.430	0.0075	0.377	0.0053	0.406	0.0104	
1990	0.395	0.0022	0.434	0.0044	0.352	0.0061	0.389	0.0054	
1991	0.412	0.0028	0.462	0.0051	0.369	0.0044	0.388	0.0052	
1992	0.419	0.0037	0.453	0.0046	0.412	0.0221	0.376	0.0050	
1993	0.456	0.0057	0.513	0.0117	0.414	0.0112	0.396	0.0063	
1994	0.465	0.0046	0.513	0.0079	0.478	0.0163	0.395	0.0094	
1995	0.434	0.0033	0.478	0.0071	0.414	0.0055	0.393	0.0058	
1996	0.442	0.0033	0.466	0.0063	0.418	0.0050	0.426	0.0064	
1997	0.458	0.0034	0.517	0.0109	0.426	0.0056	0.432	0.0058	
1998	0.466	0.0029	0.514	0.0080	0.434	0.0053	0.445	0.0089	
1999	0.462	0.0027	0.503	0.0067	0.436	0.0048	0.409	0.0064	
2000	0.486	0.0043	0.538	0.0133	0.468	0.0071	0.464	0.0194	
2001	0.458	0.0026	0.505	0.0083	0.420	0.0054	0.449	0.0060	
2002	0.457	0.0048	0.530	0.0190	0.416	0.0051	0.403	0.0062	
2003	0.449	0.0025	0.490	0.0061	0.411	0.0083	0.426	0.0056	
2004	0.450	0.0037	0.511	0.0076	0.379	0.0064	0.418	0.0060	
2005	0.442	0.0025	0.502	0.0065	0.383	0.0077	0.444	0.0061	

Note: Standard errors are based on equation (4).

Table 3 compares the evolution of the Gini coefficients for each city and for the country, with respect to three different base years: 1984, 1990 and 1999. The first base year is chosen simply because it is the beginning of our sample period. The second base year allows us to compare with respect to the year when the government introduced a series of structural policy measures aimed at liberalizing Colombian trade and foreign exchange transactions, which were also accompanied by legislation to free the labor market while granting greater protection to union rights (see Urrutia (1994) for a review of these policy reforms). The third base year allows us to

234

Cali		Medellín		Manizales		Pasto	
Gini	(s.e.)	Gini	(s.e.)	Gini	(s.e.)	Gini	(s.e.)
0.402	0.0051	0.375	0.0056	0.436	0.0076	0.464	0.0096
0.424	0.0125	0.393	0.0088	0.437	0.0092	0.455	0.0058
0.392	0.0056	0.367	0.0113	0.403	0.0079	0.445	0.0069
0.409	0.0078	0.322	0.0044	0.404	0.0110	0.454	0.0078
0.395	0.0052	0.336	0.0038	0.401	0.0072	0.447	0.0060
0.398	0.0053	0.328	0.0052	0.393	0.0065	0.449	0.0069
0.397	0.0061	0.341	0.0039	0.374	0.0055	0.459	0.0062
0.418	0.0116	0.349	0.0049	0.379	0.0066	0.453	0.0070
0.418	0.0055	0.372	0.0042	0.407	0.0097	0.449	0.0063
0.475	0.0186	0.414	0.0104	0.406	0.0070	0.443	0.0062
0.425	0.0072	0.421	0.0106	0.427	0.0104	0.451	0.0060
0.431	0.0106	0.378	0.0048	0.409	0.0057	0.458	0.0064
0.408	0.0060	0.430	0.0100	0.446	0.0076	0.475	0.0109
0.441	0.0096	0.426	0.0074	0.451	0.0064	0.469	0.0061
0.451	0.0077	0.437	0.0067	0.456	0.0055	0.481	0.0052
0.476	0.0079	0.416	0.0053	0.466	0.0075	0.488	0.0085
0.456	0.0068	0.466	0.0075	0.469	0.0091	0.504	0.0057
0.449	0.0065	0.418	0.0067	0.454	0.0065	0.489	0.0047
0.453	0.0072	0.413	0.0056	0.441	0.0049	0.488	0.0047
0.429	0.0068	0.448	0.0064	0.433	0.0056	0.478	0.0054
0.446	0.0078	0.434	0.0067	0.454	0.0173	0.456	0.0049
0.449	0.0065	0.398	0.0053	0.407	0.0054	0.473	0.0059
0.449	0.0065	0.398	0.0053	0.407	0.0054	0.4/3	0.0059

provide a comparison with respect to the lowest point of the most serious recession recorded during the last century.

Let us consider first the results when using 1984 as base year. The cities of Barranquilla, Medellín and Manizales exhibit a downward trend in their Gini coefficients during the 1980s and early 1990s, which is subsequently reversed starting in the mid-1990s. In the case of Pasto, wage income distributions appear not to have changed with respect to the level observed in 1984. In the cases of Bogota and the aggregate of the seven cities, the corresponding Gini coefficients appear to have moved upwards.

> Using 1990 as base year, we find that most of the Gini coefficients exhibit an increase. This suggests that the liberalizing policy reforms of the early 1990s led to a worsening distribution of income. Lastly, when looking at the period that followed the deepest recession of the last century, evidence is somewhat mixed. The years of recovery do not appear to have had an effect on wage income distribution in 21 out of the 48 comparisons provided, whereas in 18 cases there is a statistically significant fall in the Gini coefficients.

Table 2

Number of times the Gini coefficients are equal (1984 - 2005)

City	Barranquilla	Bogotá	Cali	Medellín	Manizales	Pasto
Bucaramanga	11	1	12	14	10	0
Barranquilla		3	9	11	7	2
Bogotá			4	0	2	8
Cali				6	16	3
Medellín					7	0
Manizales						3

Note: The tests of hypotheses reported in Tables 2 and 3 are at the 5% level.

Overall, when assessing variations in the distributions of wage income with respect to 1990 and 1999, the picture that emerges is not particularly optimistic in the sense that most of the observed variations in the Gini coefficients are in the positive direction (reflecting a worsening in inequality). It appears that the best-case scenario is that which reflects no statistically significant variation at all.

IV. CONCLUDING REMARKS

This paper analyses the evolution of the Gini coefficient in Colombia across cities, as measured by the hourly wage per worker, over a period of more than two decades. To provide valid inference on the observed variations of the estimated Gini coefficients, we implement the Davidson (2009) methodology to compute asymptotically correct standard errors. The estimated standard errors were used to perform hypotheses tests on wage income distribution equality across cities and over time. Focusing first on the cross section dimension, we find several years during which the observed differences in the Gini coefficients at the city level do not appear to be statistically different from zero. This highlights the importance of taking into account the coefficient estimated standard errors when performing comparisons. As to the time se-

ries dimension, we compare the corresponding Gini coefficients for each city with the values observed in 1984, 1990 and 1999, and find that in most cases inequality has worsened.

Table 3

Statistically significant variations in Gini coefficients

Year	Total	Bucaramanga	Barranquilla	Bogotá	Cali	Medellín	Manizales	Pasto		
Variation measured with respect to 1984										
1985		-	-	+	-	-	-	-		
1986	-	-	-	+	-	-	$\mathbf{+}$	-		
1987	$\mathbf{+}$	-	+	-	-	+	$\mathbf{+}$	-		
1988	$\mathbf{+}$	-	\checkmark	+	-	\checkmark	\checkmark	-		
1989	-	-	-	+	-	+	+	-		
1990	$\mathbf{+}$	-	\checkmark	+	-	$\mathbf{+}$	$\mathbf{+}$	-		
1991		-	+	+	-	+	\checkmark	-		
1992		-	-	+		-	\checkmark	-		
1993		-	-	+	+	+	\checkmark	-		
1994		-		+			-	-		
1995	+	-	-	+	+	-	$\mathbf{+}$	-		
1996	+			+	-	+	-	-		
1997	+	+	+	+	+	+	-	-		
1998	+	↑	+	+	+	+	+	-		
1999	+	+	+	+	+	+	+	-		
2000	+	+	+	+	+	+	+	+		
2001	+	+	↑	+	+	+	-	+		
2002	+	-	+	+	+	+	-	+		
2003	1	+	-	+	+	+	-	-		
2004	+	+	-	+	+	+	-	-		
2005		↑	-	+	↑	†	+	-		
Variatio	on meas	ured with resp	pect to 1990							
1991		-		↑	-	-	-	-		
1992	+	-	+	+	+	†	+	-		
1993		-		↑	+	↑	+	-		
1994	+	-	+	+	+	↑	^	-		
1995	+	-	↑	↑	+	↑	+	-		
1996	+	+	+	+	-	↑	^	-		
1997	+	+	+	+	+	+	+	-		

Table 3 (continued)

Statistically significant variations in Gini coefficients

Variation measured with respect to 1990

1998								+
1999							↑	+
2000		^	^					+
2001								+
2002			↑			^	↑	+
2003								+
2004		^	^	^	↑	^	↑	-
2005								-
Variation	measured wit	h respect to	1999					
2000					-		-	-
2001	-	^	$\mathbf{+}$	-	$\mathbf{+}$	-	-	-
2002	-	-	$\mathbf{+}$	-	$\mathbf{+}$	-	$\mathbf{+}$	-
2003	$\mathbf{+}$	-	$\mathbf{+}$	-	$\mathbf{+}$		+	-
2004	$\mathbf{+}$	-	$\mathbf{+}$	-	$\mathbf{+}$		-	+
2005	$\mathbf{+}$		$\mathbf{+}$	-	$\mathbf{+}$	$\mathbf{+}$	+	-