Human Capital and Inequality: A Cointegration Analysis for Colombia for the last 29 years

Ramirez Chaparro, Maria Nathalia and Chacón Mejía, Catalina

International Institute of Social Studies, Universidad Santo Tomás Bucaramanga

2021

Nathalia Ramírez Ch
Catalina Chacón M

1 Introduction

As is well known, one of the main problems that afflicts thousands of people with greater intensity is the economic inequality they live with, as day after day their quality of life is detrimentally affected by the enrichment of a small but powerful percentage of millionaires, located in the high spheres of power. Precisely, this economic inequality has come hand in hand with a strong concentration of wealth, which has been explained by multiple factors, among which stand out the low levels of education, financial and/or political instability, the incipient availability of goods and services, among other issues that have been addressed by multiple authors (Ordoñez, 2017; Piketty, 2013; Stiglitz, 2012; Atkinson, 2012). However, although it has had a boom among economists, debates persist between how to measure it, and how to understand it, essential axes to combat it. But also, there is still no consensus on what factors are related to its evolution or how to explain it. This being the reason for this essay, which seeks to see the behavior of inequality (from the Gini index approach), and adult literacy, variables that together can be explanatory of the other, contextualizing it in the evolution it has had in the Colombian reality.

The analysis is necessary, not only because of the academic debate that revolves around the issues of inequality, but also because it contrasts with the reality of the region and the country. Nowadays, Latin America is one of the most unequal regions in the world in terms of income inequality (not to mention other indicators) (Alvaredo and Gasparini 2015). And as Alicia Barcena (Secretary of ECLAC) mentioned in the presentation of the *Social Panorama of Latin America* (2020), there is no doubt that the costs of inequality have become unsustainable and that it is necessary to rebuild with equality and sustainability, where the task is long and aims at building better welfare states. Well, according to the most recent general figures presented by ECLAC (2016), the average inequality measured with the Gini coefficient in the region was 0.491, highlighting Uruguay as the country with the lowest inequality, in contrast with the most unequal countries such as: Colombia, Brazil and Guatemala.

Taking into account the condition of temporality present in the article, it is worth mentioning that in the region the first decades of the 21st century have had "acceptable" behaviors (ECLAC 2015), with decreases even that do not make the current figure negligible. In the case of Colombia, inequality (Gini coefficient) decreased by 9% between 2002 and 2015. However, the trend is not clear, because although there have been improvements in certain years, it has also been elevated in others (2003, 2008 and 2013).

Regarding literacy, in the Colombian context, at the beginning of the millennium (2001) the illiteracy rate was 7.38%, so the government proposed that by 2015 the goal was to reach 3.69%, however, this did not happen. And later in the National Development Plan 2010 - 2014 they would establish 5.7% as the goal for the illiteracy rate. On the other hand, the Ten-Year Education Plan 2006-2016 established the goal of totally eradicating illiteracy in urban areas and reducing it to 2% in rural areas, goals that were not met and governmental and institutional planning did not take place. However, it is worth evaluating the evolution of the literacy rate, which for the Colombian Ministry of Education (2015) includes literacy and basic education for young people and adults who for whatever reason did not enter the educational service or dropped out of it prematurely.

Now, having mentioned the variables that will be used in the article, it is worth linking the human capital approach and its relationship with inequality, which can be related from the capabilities approach of Sen (2000), who explains capabilities as the basis for the development of people, and as a way of guaranteeing a "better life". In this sense, people's capabilities are not defined by money, but by access to facilities and opportunities, of which education is highlighted in this particular case (other capabilities are related to variables such as freedom, democracy and housing). Such is, some works such as Walker (2012) and D'Agata (2007), among others, reaffirm the need for opportunities for the generation of capabilities, which lead to and are related to the growth and development of countries, and therefore a decrease in inequality. Thus, based on this, this article seeks to look at this relationship, taking adult literacy as a proxy of human capital in the particular context of Colombia. Also, the relationship between human capital and inequality could be seen from Becker's theory, who proposed the causal relationship between human capital and productivity (with respect to income), on the grounds that the greater the inequality in access to human capital, the lower the productivity and therefore the lower the economic growth (Weiss 2015).

of this article (nor do the data for Colombia allow for this type of analysis), for the moment, we only consider the possible relationship between the variables, and see how they interact over time.

In the development of the research, time series were used for the period between 1990 and 2019, using data from the World Bank (2019). To verify the existence of cointegration in the long and short term, Augmented Dickey-Fuller tests and the error correction model (ECV) were used, and the possible existence of causal links between variables was analyzed with the vector autoregressive model (VAR). Thus, the aim was to see whether human capital drives the reduction in inequality in Colombia in the short and long term. The paper is divided into four sections, starting with the conceptual and background description, followed by the methodology employed, and finally the concluding results.

2 Theoretical framework and background

As mentioned in the introduction, Amartya Sen (2000), from his capabilities approach, proposes an alternative way of understanding the development and freedom of people in the different spheres that intersect in their daily lives, conditions that allow individuals to have or not function and capabilities that, under this theoretical framework, allow them to develop. Thus, several authors have related the absence of this type of capabilities to inequality, so that not having the possibility of developing these capabilities compared to others who do, conditions them with the absence of well-being and multiple precariousness (Urquijo 2014). Then, following the work of authors such as Rambe and Mosweunyane (2017) it could be thought that if people are given opportunities, inequality gaps can be reduced, and in this case, the opportunity would be reflected in education.

Within empirical research, works such as those of Bumman and Lensink (2016), Zhao, Wu and He 2017) and Signorelli (2016) show that the development of capabilities, reflected in human capital formation largely explains inequality and intergenerational mobility in terms of income, although under the analysis of high-income countries. Regarding research in developing countries (as is the case of Colombia), they relate inequality as an impediment to productive sectors, which evidently affects capital accumulation, and in turn, they allude as another affectation, the formation of human capital (Chakraborty and Gupta 2009).

3 Data and Methodology

3.1 Data

The time series used were taken from the World Bank (2019) database, taking the variables Gini Coefficient and Adult Literacy Rate, over the time period 1990 to 2019, with annual information for each variable, for a total of 29 years.

Table 1. Description of the variables

Variable	*Symbol*	Description
Inequality	I_t	measures the distribution of income among individuals or households within an economy, and how equal or unequal that distribution is or is not.
Human Capital	H_t	Capacity and skills that people acquire through (their)investments in education and training

Although income inequality has different ways of measurement as suggested by Atkinson (1975), any indicator of inequality implies normative judgments about the weights assigned to differences in income and distribution, for ease of data and consensus, the Gini Coefficient was chosen (even being aware of its limitations compared to other indexes such as entropy or Atkinson, which are yet to be better developed). On the other hand, the proxy for human capital for Colombia was the adult literacy rate, because even though there is now a developed human capital index, its temporality is very recent. The literacy rate, on the other hand, allows us to know in these 29 years the number of literate people aged 15 and over, expressed as a percentage of the total population of people aged 15 and over.

3.2 Methodology
As mentioned above, the capabilities theory and Sen's econometric contributions, I based my model on it.

$$D_t = \beta_0 + \beta_1 I_t + \varepsilon_t$$

Where Dt is the dependent variable, Ht is the independent variable (human capital) and "t is the error term. (human capital) and "t is the error term. The sub-index t = 1990-2019 indicates the time.

$$\Delta I_t = \beta_0 + \beta_1 \sum_{i=0}^{n} \Delta I_{t-i} + \beta_2 \sum_{i=0}^{n} H_{t-i} + \varepsilon_t$$

$$\Delta H_t = \beta_3 + \beta_4 \sum_{i=0}^{n} \Delta H_{t-i} + \beta_5 \sum_{i=0}^{n} I_{t-i} + \varepsilon_t$$

In order to examine the long-run relationship between the variables, a vector autoregressive model (VAR) and the cointegration test are applied to the equations and the error correction test to the equations.

4 Results

Initially, logarithms were generated so that the variables would be uniform, and with the logarithms the model would be run, and then the residuals would be checked to see if they were stationary or not. This, taking into account that the cointegration theory is based on the fact that in order to consider that two variables are cointegrated, the residuals must be stationary, so logarithms of the variables were generated. After that, the program (Rstudio) was asked to take the data as a time series, reaffirming that it is an annual time series object from 1990-2019. Once the time series condition was given, the variation of the data can be observed in graph 1.

Graph 1. Time series of the variables Inequality (Gini) and Human Capital (Adult Literacy rate)
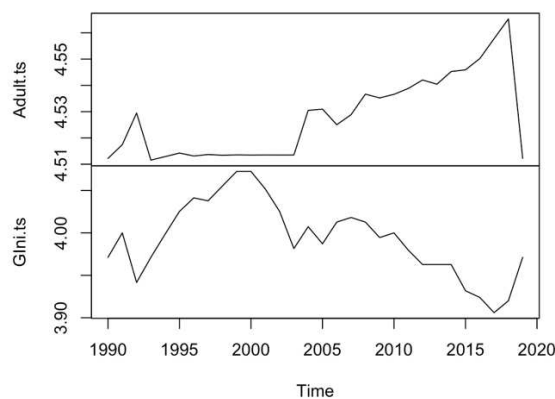
To achieve stationarity, it was differentiated once, and as an advantage of R, it is the ndiff command that allows to specify the number of differences needed to achieve stationarity, in that sense, it also resulted in 1, thus corroborating the need for only 1 difference

Graph 2. Time series of the stationarized variables
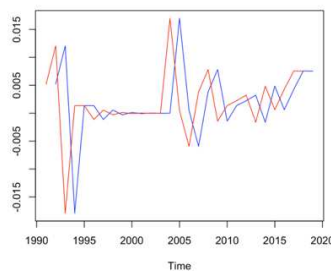
As can be seen in graph 2, the mean is quite constant, which assures that the variables are stationary and the best forecasts can be obtained, since graphically it appears that their distribution and parameters do not vary over time, nor do they appear to be following a trend.

Once the stationarity was confirmed, a regression model was generated to check if the variables were cointegrated, when generating the first model, its expression would be:

LlnGINI1=5.55129-0.256360(lnADULT)***

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.55129    0.18309  30.320  < 2e-16 ***
GIni.ts     -0.25636    0.04585  -5.592 5.52e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01092 on 28 degrees of freedom
Multiple R-squared:  0.5276,    Adjusted R-squared:  0.5107
F-statistic: 31.27 on 1 and 28 DF,  p-value: 5.522e-06
```

Once this was done, the residuals were generated, which graphically show that they are not stationary, as can be seen in graph 3, where they are dispersed and, taking into account Engle and Granger (1987), cointegration is a process between two variables that are not stationary in time, which exists when the linear combination between them is stationary given that both present the same order of integration, a property that must be fulfilled. both have the same order of integration, a property that must be fulfilled.

Graph. 3 Residuals of the regression model
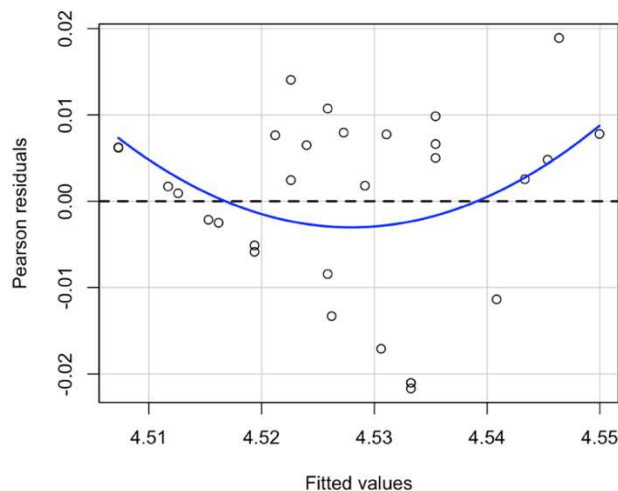
Therefore, so far the variables show that they are not cointegrated. To test the cointegration of the variables, the Augmented Dickey-Fuller test was chosen, which allows us to check the stationarity of the residuals. It gives a result of 0.7944 which is greater than 0.05 indicating that there is no stationarity in the errors, therefore the variables are not cointegrated. However, following the recommendations of some econometrics texts such as Gujarati and Porter (1999), is to add the trend variable, which once created, is added to the model. As a result, the variables are now significant, and the p-value is less than $0.05(6.440 \times 10^{-4})$. Thus, it was necessary to add the trend variable to get a better model. Model that yielded the coefficients shown below.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.1412044  0.1670723  30.772  < 2e-16 ***
tendencia    0.0009431  0.0002080   4.535 0.000106 ***
GIni.ts     -0.1573230  0.0414048  -3.800 0.000750 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.008376 on 27 degrees of freedom
Multiple R-squared:  0.7318,    Adjusted R-squared:  0.7119
F-statistic: 36.84 on 2 and 27 DF,  p-value: 1.924e-08
```

However, the analysis can be made easier graphically. Thus, as can be seen in the Graph 4., the modeling errors improve and now appear to be stationary.

Graph 4. Residuals of the regression model, including the trend of the series
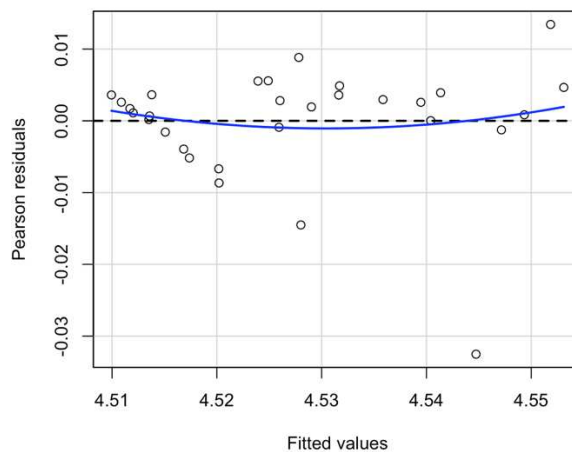
However, in order to test, the Augmented Dickey-Fuller test was performed again, in order to reaffirm the stationarity of the residuals. This test, as it is already known, investigates the existence of unit roots, but unlike Dickey-Fuller, in ADF the error term is correlated.

```
Value of test-statistic is: -3.5988 4.7815 7.0117

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Based on the critical value and the t-statistics value at 10pct the value falls in the region of rejecting H0, which is worth remembering H0: unit root and H1: No unit root, at 5 percent also rejects H0, i.e. at 5 and 10% the errors are stationary. This proves that by adding the trend to the model, the errors become stationary. And once the errors are stationary, we can say that the variables are cointegrated, which means that there is an equilibrium between the variables in the long run, that is, there is an equilibrium between Inequality and Human Capital in the long run.

Now, to prove that the model is valid in the long run, it is necessary to generalize the error correction model. Then, to generate the error correction model, the differences of the logarithm of the two variables with the time format (ts) were calculated.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.128e-19  2.000e-03   0.000   1.0000
dlnGini     -2.084e-01  8.138e-02  -2.561   0.0163 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01077 on 27 degrees of freedom
Multiple R-squared:  0.1955,    Adjusted R-squared:  0.1657
F-statistic:  6.56 on 1 and 27 DF,  p-value: 0.01633
```

From the above, we can see that the differences in Gini are significant. Taking this into account, the residuals of the new model were generated again, because for error correction modeling, the variables with differences are added and the error is added with a lag, that is, with a delay. After generating the residuals of the new model, the lag of the residuals was calculated to generate the error correction model (ECM).

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0003891  0.0019820   0.196  0.84595
dlnGini     -0.2543792  0.0804420  -3.162  0.00408 **
res4_1      -0.5529346  0.2912557  -1.898  0.06924 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01023 on 25 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.3227,    Adjusted R-squared:  0.2686
F-statistic: 5.957 on 2 and 25 DF,  p-value: 0.007664
```

Taking out the model, the difference of the natural logarithm of Gini and the residuals was significant, so the cointegration of the variables is valid in the short run. In turn, the parameter of the cointegration vector of the errors is negative (which complies with what was expected to happen), and in absolute terms it is less than 1 [0.55]. Therefore, the relationship is accepted both in the short and long term, i.e., in the model the cointegration is valid, because the residuals with a lag were significant in the error correction model.

To determine the causal order, Granger causalities were generated, using the lmtes, and dynml libraries, which allow for diagnostic checking in linear regression models. Furthermore, some generic tools for inference in parametric models that are provided.

```
Model 1: dlnAdult ~ Lags(dlnAdult, 1:1) + Lags(dlnGini, 1:1)
Model 2: dlnAdult ~ Lags(dlnAdult, 1:1)
  Res.Df Df     F Pr(>F)
1     25
2     26 -1 0.147 0.7047
> |
```

Now with the cross variables

```
Model 1: dlnGini ~ Lags(dlnGini, 1:1) + Lags(dlnAdult, 1:1)
Model 2: dlnGini ~ Lags(dlnGini, 1:1)
  Res.Df Df     F Pr(>F)
1     25
2     26 -1 0.7928 0.3817
    |
```

As can be seen, we are accepting the null hypothesis, i.e. literacy does not cause in the granger sense inequality, because our result is greater than .05, and if we do it the other way around, neither does it. Therefore, this means that there is no Granger causality of the variables in the Colombian context, although they are cointegrated in the short and long run, they do not cause each other.

Following the VAR process, as shown in Annex 1, it would be understood that the order of the lags would be 9, resulting in the graph below.
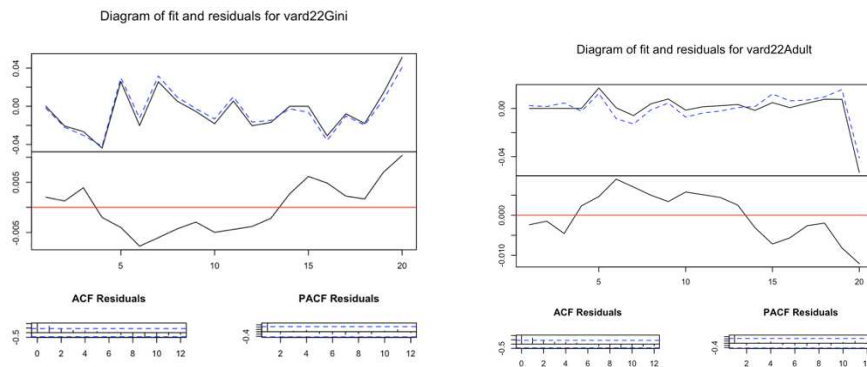
However, so far the results have been significant, the serial autocorrelation test of the residuals could not be performed optimally, taking into account that the database is 29 years old, and is insufficient for the matrix analysis. It could be said that , when the existence of integration with the first difference of inequality and human capital is verified, the equilibrium error term is obtained, which is used to estimate the error correction model (ECV), to determine the existence of equilibrium in the short term between the variables. This would mean that changes in human capital formation would lead to favorable results in inequality in a few years, i.e., that individuals could have access to a better quality of life by applying the knowledge acquired during their formative years. However, due to the lack of data, the analysis could not be complete, so it would not be correct to make such assertions in advance.

References

Alvaredo, F. and Gasparini, L., 2015. Recent trends in inequality and poverty in developing countries. *Handbook of income distribution*, *2*, pp.697-805.

Atkinson, A.B., 1975. The economics of inequality.

Atkinson, A.B., 2015. *Inequality*. Harvard University Press.

Bumann, S. and Lensink, R., 2016. Capital account liberalization and income inequality. *Journal of International Money and Finance*, *61*, pp.143-162.

Chakraborty, B. and Gupta, M.R., 2009. Human capital, inequality, endogenous growth and educational subsidy: A theoretical analysis. *Research in Economics*, *63*(2), pp.77-90.

D'Agata, A., 2007. Endogenizing Sen's capabilities: An adaptive dynamic analysis. The Journal of Socio-Economics, 36(2), pp.177-190.

ECLAC., 2015. Anuario estadístico de Am rica Latina y el Caribe. Santiago de Chile: Comisión Económica para America Latina y el Caribe, Naciones Unidas.

ECLAC., 2016. *Bases de datos y publicaciones estadísticas. Estadísticas e indicadores*. Disponible en http:// estadisticas.cepal.org.

ECLAC., 2020. *Panorama Social de América Latina 2020* Cepal.

Engle, R.F. and Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pp.251-276.

Gujarati, D.N. and Porter, D.C., 1999. *Essentials of econometrics* (Vol. 2). Singapore: Irwin/McGraw-Hill.

Mincer, J., 1974. Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

Ordóñez, J.L.Z., 2017. De la función económica del cheque, del cheque común al de pago diferido. *JURÍDICAS CUC*, *13*(1), pp.183-198.

Piketty, T., 2013. Capital in the 21st Century. *Cambridge, MA: President and Fellows, Harvard College*.

Sen, A., 2000. Development as freedom. DEVELOPMENT IN PRACTICE-OXFORD-, 10(2), pp.258-258.

Signorelli, M., 2016. Human capital investment and inequality of opportunities in China. *China Economic Review*, *100*(38), p.A1.

Stiglitz, J.E., 2012. *El precio de la desigualdad: el 1% de población tiene lo que el 99% necesita*. Taurus.

Urquijo, M., 2014. La teoría de las capacidades en Amartya Sen. *Revista Edetania*, *46*, pp.63-80.

Walker, M., 2012. A capital or capabilities education narrative in a world of staggering inequalities?. International Journal of Educational Development, 32(3), pp.384-393.

Weiss, Y., 2015. Gary Becker on human capital. Journal of Demographic Economics, 81(1), pp.27-31.

Annex 1. VAR process

```
AIC(n) HQ(n) SC(n) FPE(n)
   9    9    9    10

$criteria
                 1            2            3            4            5            6            7
AIC(n) -1.604970e+01 -1.617470e+01 -1.606130e+01 -1.580285e+01 -1.551974e+01 -
1.555047e+01 -1.517145e+01
HQ(n)  -1.599923e+01 -1.609057e+01 -1.594352e+01 -1.565143e+01 -1.533467e+01 -
1.533174e+01 -1.491908e+01
SC(n)  -1.575146e+01 -1.567762e+01 -1.536540e+01 -1.490812e+01 -1.442618e+01 -
1.425808e+01 -1.368023e+01
FPE(n) 1.076508e-07 9.692425e-08 1.138250e-07 1.615659e-07 2.524581e-07 3.250275e-07
7.915956e-07
                 8    9   10
AIC(n) -1.616258e+01 -Inf -Inf
HQ(n)  -1.587656e+01 -Inf -Inf
SC(n)  -1.447253e+01 -Inf -Inf
FPE(n) 8.647838e-07 NaN    0
```