Income Inequality, TFP, and Human Capital

Sequeira, Tiago and Santos, Marcelo and Ferreira-Lopes, Alexandra

Univ. Beira Interior and CEFAGE-UBI, Univ. Beira Interior and CEFAGE-UBI, ISCTE-IUL, BRU-IUL and CEFAGE-UBI

29 March 2014

Income Inequality, TFP, and Human Capital*

Tiago Neves Sequeira†
Univ. Beira Interior and CEFAGE-UBI

Marcelo Santos‡
Univ. Beira Interior and CEFAGE-UBI

Alexandra Ferreira-Lopes§
ISCTE-IUL, BRU-IUL, and CEFAGE-UBI

Abstract

A fruitful recent theoretical literature has related human capital and technological development with income (and wages) inequality. However, empirical assessments on the relationship are still scarce. We relate human capital and total factor productivity (TFP) with inequality and discover that, when countries are assumed as heterogeneous and dependent cross-sections, human capital is the most robust determinant of inequality, contributing to increase inequality, as predicted by theory. There is evidence of great heterogeneity on the effects of TFP and Openness across countries. These new empirical results open a wide avenue for theoretical research on the country-specific features conditioning the causal relationship from human capital, technology and trade to inequality.
Keywords: income inequality, human capital, technology.
JEL Codes: I32, O10, O33, O50.

1 Introduction

Understanding the causes of inequality is fundamental to indicate possible policy measures that ensure that the increased production and income of societies can be better shared among the whole population. Reducing inequality is important not just to achieve a fairer distribution of income and address the social concerns that widening disparities in income raise, but also to ensure a good environment for growth. As has been seen in some countries, these social concerns can lead to social instability. This inequality may itself limit the growth potential of

*This paper was funded by FCT - Fundação para a Ciência e a Tecnologia through project PTDC/EGE-ECO/112499/2009.

†Departamento de Gestão e Economia and CEFAGE-UBI. Universidade da Beira Interior. Estrada do Sineiro. 6200-209 Covilhã, Portugal. Corresponding author. e-mail: sequeira@ubi.pt.

‡Departamento de Gestão e Economia and CEFAGE-UBI. Universidade da Beira Interior. Estrada do Sineiro. 6200-209 Covilhã, Portugal.

§Instituto Universitário de Lisboa, ISCTE-IUL, ISCTE Business School Economics Department, BRU-IUL (Business Research Unit), Lisboa, Portugal, and CEFAGE-UBI.

economies as social, economic, and political instability is associated with slower growth. Even in democracies, an increase in inequality may contribute to elect politicians that are against openness and globalization, which may deter the world integration process which is known to have positive effect on the growth prospects of the economy.

There is a fruitful theoretical literature interested in explaining the rise of inequality in the second-half of the twentieth century (mainly in the USA) together with the rise in the supply of human capital. Skill biased technical change and capital-skill complementarity have been crucial to explain this phenomenon. Generally, according to this theory, skill-premia increase due to two effects. First, the skill premium would reflect the productivity difference between sectors. Second, with full capital mobility, factor price equalization requires capital to flow to the sector operating the new technology, and thus workers in the new technologies sectors are endowed with more capital, which boosts their relative wages (Acemoglu, 2002a, 2002b, 2003). A recent development has argued that the diffusion of IT - General Purpose Technologies - may have raised the demand for adaptable skilled workers and made vintages of capital more adaptable. Therefore, this increases the premium of workers that show a lower learning cost and can adapt quickly from one sector to another. These ideas have been formalized by Galor and Tsidon (1997), Greewood and Yorukoglu (1997), Caselli (1999), Galor and Moav (2000) and Aghion, Howitt, and Violante (2002). Theoretically, skill-biased technological change is explained by the proportion of skills (education) in the economy, and wage inequality (typically measured by the wage ratio between skilled and unskilled workers) is proportional to the proportion of skills in the economy. Education is thus seen in the theory as a determinant of more technical change (and consequently growth) and more inequality.

Whatever the explanation is for the rise in inequality and its relationship to technology and human capital, there is little quantitative literature on the issue, as pointed out by Hornstein, Krusel and Violante (2005:1361). In fact, empirical attempts to evaluate the relationship are mostly country-specific as, e.g. Ding et al. (2011) and Rattsø and Stokke (2013) dealing with the effect of technology, and in Birchenall (2001) dealing with the effect of human capital. Micro evidence on the relationship between education and income inequality is also mixed. While Martins and Pereira (2004) found a positive effect of education returns in inequality due to an increase in returns to education throughout the wage distribution for 16 European Countries, Wang (2011) found returns to education in China that are more pronounced for individuals in the lower tail of the earnings distribution than for those in the upper tail, in stark contrast to the results found in some developed countries.

We have only found two papers that evaluated this relationship using a large cross-section of countries. Barro (2000) presents fixed-effects estimations of equations of the Gini index on covariates such as GDP and GDP squared, schooling, democracy index, openness, rule of law index and several dummies. In his fixed-effects estimations, dummies for income or spending and secondary schooling are negatively related to inequality and higher schooling and openness are positively related to inequality (with significant coefficients). Primary schooling and the dummy for individual or household data are insignificantly related to the Gini coefficient. There is a strong inverted-U relationship with GDP (the so-called Kuznets curve) in Barro's estimations. Recently, Jaumotte, Lall, and Papageorgiou (2013) re-assessed the determinants of inequality. They focus on the effect of globalization on inequality but avoid the relationship between inequality and GDP. They conclude that trade globalization decreases inequality while financial globalization increase inequality. Moreover, information and communication technologies and credit deepening increases inequality while the share of industry in the economy decreases inequality. Interestingly, education variables and initial GDP (when included) are insignificantly related to inequality.

As can be noted, empirical evidence coming from a large cross-section of countries has quite ambiguous results regarding the determinants of inequality and does not confirm theories in crucial aspects such as the influence of education and technology. However, much criticism has affected data on inequality around the world. In fact, greater coverage across countries and over time is available from these sources only at the cost of significantly reduced comparability across observations. There are currently three different projects that collect and make publicly available inequality data for many countries and periods around the world: the Luxembourg Income Study (LIS), the dataset assembled by Deininger and Squire (1996) for the World Bank (WIID) - recently updated and upgraded by the WIDER (World Institute for Development Economic Analysis) project, and the most recent standardized World Income Inequality dataset (SWIID), by Solt (2009). The LIS, which was used by Jaumotte, Lall, and Papageorgiou (2013), has generated the most-comparable income inequality statistics currently available but covers relatively few countries and years. The Deininger and Squire dataset and its successors, used by Barro (2000), on the other hand, can be used to provide many more observations, but only at a substantial loss of comparability. Solt (2009) implemented a sequence of steps in order to standardize income inequality data and provide data with more ample coverage than the WIID but at the highest quality as in LIS. However, in the process of standardization, not all countries had the sufficient data in the original sources. To handle this, Solt (2009) also calculated a standard-error of each Gini coefficient to account for the remaining uncertainty in data.

This paper contributes to our knowledge of the relationship between human capital, technology and inequality in two crucial ways: first, it uses a large database on inequality, based on the Standardized World Income Inequality dataset, and combines it with the most recent data for human capital and TFP; second, for the first time, it takes into account country heterogeneity, cross-country dependence, and endogeneity to common factors in evaluating the effects of human capital and TFP on inequality. The exploration of a large dataset of over 150 countries across more than 50 years (since 1960) allowed us to explore issues such as panel heterogeneity, cross-country dependence and time-series features, such as stationarity and causality, which are absent from earlier contributions. Exploring the heterogeneity of results concerning the determinants of inequality is especially important since the effects of different inequality determinants may differ considerably from country to country. In fact, and to give a few examples, the effect of technology adoption may differ if the country is on the technological frontier or lagging behind; the effect of human capital may differ between countries where brain-drain is more evident than in others; and the effect of openness may depend crucially on the level of integration and on the market size of the country. In general, historical and institutional (e.g. labor market related) country-specific factors that are not simply captured by fixed-effects estimations, are in fact dealt through heterogeneous panel estimations.

Our main conclusions point out to a clearly significant, worldwide relevant, positive effect of human capital on inequality, an effect that is stronger for the developed world. On the contrary, our results indicate that the effects of technology and openness may be quite different from country to country, as well as dependent on different specifications. Overall, differences in inequality data and the common factors framework dismiss the existence of a Kuznets curve.

The remainder of the paper is organized as follows. Next, in Section 2 we describe our dataset. In Section 3 we describe our estimation strategy. In Section 4 we present our results, beginning with detailed evidence for cross-country dependence, stationarity, and causality and then showing the results from several different specifications based on heterogeneous panels methods. Section 5 concludes.

2 Sources and Data

We use data (mainly) from the Standardized World Income Inequality database (SWIID), version 4.0, from Solt (2009), for the Gini coefficient.[1] These include data on the Gini coefficient using post-taxes and post-transfers income (the net definition) and on the Gini coefficient using pre-taxes and pre-transfers income (the market definition), and the respective standard-errors by country and year. We use GDP *per capita*, openness, human capital index, and TFP index from Penn World Tables (PWT), version 8.0 (Feenstra et al., 2013).[2] Human capital in PWT 8.0 is measured by a 'Mincerian' combination of years of schooling (from Barro and Lee, 2013, version 1.3) and returns to education. The results from Psacharopoulos (1994) show that returns from schooling decrease across years of schooling. As the influence of human capital in inequality arguably changes through years of schooling (Barro's results show negative signs for primary and secondary schooling and positive signs for tertiary schooling) and returns from schooling are essential to understand income inequality, we think this variable is the most appropriate human capital measure to enter in inequality regressions. In fact, as human capital measures corrected for returns for education weights more lower levels of education, they correct underestimations of human capital in less developed countries. Lower levels of education in less developed countries may have more influence in decreasing wage inequality than they have in more developed countries. The human capital measure provided by the PWT 8.0 is the one with the highest coverage until now, as it not only corrects years of schooling by different returns by levels of education, but it is also interpolated to provide annual measures. It is worth noting that returns to education differ between levels of education but not between different countries or years as these alternatives would result in lower coverage.

TFP is available in PWT 8.0 both as a ratio to the USA=1 level and on constant national prices. We construct our index departing from a final TFP level (related to the USA) in 2011 and then deflating year by year using growth rates of the national currency measure of TFP. This allows us to have a PPP measure of TFP that is independent of the USA level in the time-series analyzed .[3] We compare some results with inequality data coming from the Word Income Inequality database (WIID2c).[4] In doing so, we followed some strict criteria to select data, separating Gini coefficients from net income, consumption and gross income and preferring data with wide coverage and higher quality.[5] Contrary to Barro (2000) but similar with Jaumotte, Lall, and Papageorgiou (2013), we used annual data.

[1] Available at $http : //thedata.harvard.edu/dvn/dv/fsolt/faces/study/StudyPage.xhtml?studyId = 36908$.

[2] Available at $http : //www.rug.nl/research/ggdc/data/penn - world - table$.

[3] We began with the year 2011 in order to maximize the available data for the TFP index.

[4] Available at $http : //www.wider.unu.edu/research/Database/en_GB/database/$.

[5] These criteria are detailed in a Technical Appendix, which can be provided by the authors.

Table 1: Descriptive Statistics

Variable	N. Obs	Mean	Std. Dev.	Min	Max
Gini (income net) WIID	1091	3.4720	0.2818	2.7081	4.1865
Gini (income gross) WIID	1143	3.6148	0.3070	2.7663	4.3516
Gini (consumption) WIID	419	3.6402	0.2273	2.8112	4.3027
Gini (net) SWIID	4597	3.5923	0.2960	2.7324	7.3871
Gini (market) SWIID	4597	3.7395	0.2234	2.8367	4.3740
Gini (net) - value/sd SWIID	4597	3.5613	0.9786	1.2658	9.5894
Gini (market) - value/sd SWIID	4597	3.2479	1.0049	1.0747	9.5410
Human Capital	6797	0.6905	0.3160	0.0198	1.2861
TFP	4994	0.5254	0.5287	-3.5389	1.1222
Openness	7760	1.1645	1.1020	-12.7415	3.2061
GDP per capita	7760	8.2779	1.1891	4.8890	10.9961

Notes: WIID is the World Income Inequality Database, version 2c, from the World Bank, updated by the WIDER - World Institute for Development Economic Analysis. SWIID is the Standardized World Income Inequality Database, from Solt (2009). Human Capital, TFP, Openness = (Exports+Imports)/GDP - and GDP per capita are from PWT 8.0. When value/sd is indicated it means that the Gini coefficient is divided by its standard-error, a measure to account for uncertainty in the data for each country-year pair. All variables are in natural logarithms.

We end up with an unbalanced panel database of 156 countries with an average of 31 years per country, from 1960 to 2011.[6] Table 1 shows descriptive statistics for the variables included in the analysis. The table shows the different coverage among the different income inequality measures, indicating that the measures coming from the SWIID have more than four times the number of observations than the measures coming from the WIID, yet less than the coverage given by the PWT to variables of TFP, human capital, openness, and GDP.

3 Estimation and Methods

Our first step in this section was a specification search, which we present in Appendix A. First, we ran fixed-effect regressions[7] for different Gini index measures, such as Gini from net income data (from the WIID2c), Gini from consumption data (from the WIID2c), Gini from gross income data (from the WIID2c), Gini from market data - pre-taxes, pre-transfers (from the SWIID 4.0), the definition of which is equivalent to Gini from net income data (from the WIID2c), and finally Gini from net data - post-taxes, post-transfers, our preferred measure. We ran those regressions in GDP *per capita* and GDP *per capita* squared, to account for the existence or not of the Kuznets curve, in openness ratio (imports plus exports as a share of GDP), in TFP and in human capital, including year dummies. Results of these regressions are in Tables A.1 and A.2 in Appendix A. Table A.1 shows regressions from the Gini indexes without correcting for quality or data uncertainty and Table A.2 shows regressions for the Gini indexes which are corrected for quality or data uncertainty. We conclude that the coefficients on GDP and GDP squared are significant only when the dependent variable is the Gini coefficient from net income data (from the WIID2c). Human capital decreases

[6] 31 observations per country is the average number of time-series per country considering the pool of the mentioned variables although some variables may include nearly 50 years per country.

[7] Fixed-effects estimation is common in earlier contributions and to better compare we used this in the specification search. However, in the main analysis we prefer to use a more adequate estimator class in our moderate T - moderate N panel data database.

inequality in regressions for the Gini coefficient from net income data (from the WIID2c) and in regressions for both variables from the SWIID - but only in the ones presented in Table A.1, when the Gini coefficients are not corrected for uncertainty. TFP has a significantly positive coefficient for regressions for the Gini coefficient from gross income data (from WIID). Openness is significantly related to inequality with a positive sign for net and gross income data (from the WIID) and for both variables from the SWIID - in this case only for the case in which Gini coefficients are not corrected for uncertainty in data. These regressions highlight the extremely different results obtained from considering different measures of inequality.[8] Given the wide rejection of the Kuznets curve hypothesis, we dropped GDP and GDP squared from our baseline specification. We otherwise consider the three additional variables - human capital, TFP, and openness - as they seem to be the variables that best summarize the covariates considered in the closest earlier articles and they seemed robust to some of our fixed-effects specifications.

Our estimation method hereinafter is the common factor framework for heterogeneous panels from Pesaran (2006) and followers. We note that also in this framework the Kuznets hypothesis is always rejected. Our baseline specification is thus as follows:

$$gini_{it} = \beta_1 hcap_{it} + \beta_2 TFP_{it} + \beta_3 Open_{it} + \lambda_i' \mathbf{f}_t + u_{it} \tag{1}$$

where *gini* is the natural logarithm of the Gini coefficient, TFP is the natural logarithm of a measure of total factor productivity, *hcap* is the the natural logarithm of the human capital variable, *Open* is the the natural logarithm of the openness ratio, $\mathbf{f}_t$ is the vector of unobservable common factors and λ_i' is the associated vector of factor loadings. As can be observed from (1), each coefficient is country-specific, thus allowing for complete heterogeneity in the estimation. Additionally, as each regressor can also depend on the common factor, the method is also robust to endogeneity of the observable factors toward the common factors determining inequality. As Pesaran and Tosetti (2011) explain, this method is robust to non-stationarity in both observables and non-observables and works well in the presence of weak and/or strong cross-sectionally correlated errors.[9] As the analysis in Jaumotte, Lall and Papageorgiou (2013) might indicate, we suspect that the Gini coefficients, financial openness, and technological development may well be non-stationary. Finally, we may consider that technology adoption is being determined by the same phenomena as inequality, say by common factors such as globalization or the entry of China into the world market, technology thus being an endogenous variable.[10] These are the reasons why we will apply the Pesaran (2006) estimator for heterogeneous panels.

[8] A comparison with the earlier papers from Barro (2000) and Jaumotte, Lall and Papageorgiou (2013) also leads to this conclusion although they use a different set of covariates.

[9] There are not many empirical applications with those heterogeneous panel methods. Notable exceptions are the recent papers from Markus Eberhardt and co-authors (Eberhardt and Presbitero, 2014; Eberhardt and Teal, 2013a, 2013b and Eberhardt, Helmers, and Strauss, 2013). Eberhardt and Teal (2011) explain why the standard cross-country regression framework and its panel cousins needs to be reconsidered.

[10] For complete arguments toward reconsideration of traditional econometric methods to study moderate-T dimensional panel data of countries, see Eberhardt and Teal (2011).

4 Empirical Results

Our results section begins by presenting evidence of the time-series properties of inequality. Due to the existence of quite unbalanced panels in Gini measures, clearer in those coming from the World Income Inequality Database (version 2c) than in those coming from its Standardized World Income Inequality Database (version 4.0) (from Solt, 2009), it is only possible to provide evidence on stationarity and causality for a sub-set of countries of the sample for Gini measures provided by the last database. Due to unbalance and holes in several time series, to perform some of those tests, we limit our variable of interest such that we include only countries with more than a given number of time-series observations (30). We consider both the Gini coefficient as provided by the source as well as an uncertainty-corrected version of the Gini coefficient which consists of dividing the coefficient by the standard-deviation, also provided by the source.[11]

These new data on inequality provide, for the first time, the means for analyzing time-series features in a reasonable set of countries. This analysis occupies Sub-Sections 4.1 and 4.2.[12] Then, in Section 4.3 we present evidence on the relationship between human capital, TFP and openness in inequality in a heterogeneous panel setup and several robustness analyses. Section 4.4. discusses the results.

4.1 Initial Analysis: cross-country dependence and stationarity

The standard literature on the panel data analysis assumes cross-sectional independence. However, there are several reasons why cross-sectional dependent error structure can arise in a large panel data of countries. Such cross-correlations can arise for a variety of reasons such as omitted common factors that affect the evolution of inequality, including technological cross-country spillovers, migration of high and low skilled workers and integration in international markets. As Pesaran and Tosetti (2011) write, "conditioning on variables specific to the cross-section units alone does not deliver cross-section error independence, an assumption required by the standard literature on panel data models", the one that has been applied in the existing analyses of the determinants of inequality. Table 2 shows results for the cross-sectional dependence test from Pesaran (2004) which tests the null of no cross-sectional dependence.

These tests constitute overwhelming evidence that the series of inequality (as well as their main determinants) are cross-country related, thus inducing bias on estimations assuming cross-country independence. It is interesting to note that the series with the highest cross-dependence test is human capital, following by openness. Also worth noting is that the uncertainty corrected measures of the Gini coefficient present higher values for the test than

[11]The uncertainty-corrected measure is $\frac{GINI}{sd(GINI)}$, where *GINI* is the Gini index provided by SWIID and *sd(GINI)* is the standard-deviation of the Gini index, also provided by the SWIID and that corrects for uncertainty or measurement error within the sources. Later on, on the Discussion section, we discuss the results obtained with an alternative uncertainty-corrected measure.

[12]It should be noted however, as stressed by Eberhardt and Teal (2011), that most of the unit-root and cointegration tests have low power in panels of moderate dimension such as the one under analysis. This does not invalidate that their results constitute important motivation to choose a heterogeneous common factor approach that is indeed appropriate to deal with moderate N, moderate T panels, typical in macroeconomic analysis.

Table 2: Cross-sectional dependence test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variable	Gini Net Income SWIID (>30)	Gini market SWIID (>30)	Gini Net Income SWIID (>30, ./sd)	Gini Market Income SWIID (>30, ./sd)	Human Capital	TFP	Open-ness
CD Test	23.33***	19.79***	96.40***	79.47***	554.05***	53.81***	240.32***
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of Countries	82	82	82	82	128	106	155

Note: >30 indicates that only cross-sections with more that 30 time-series observations are included. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. ./sd indicates when the Gini coefficient is divided by the source standard-deviation to account for data uncertainty. All variables are in natural logarithms.

the original Gini coefficients, indicating an increased correlation between countries in these uncertainty-corrected measures. Although we provided results from the Gini coefficient from the market approach (SWIID) in this Table 2, from now on we will concentrate on the most interesting variable: the Gini coefficient from post-tax and post-transfers income. This variable incorporates the effects of progressive tax systems and is close to a measure of inequality related to disposable income.[13]

Another issue to be dealt with is the integration level of the series, i.e. its stationarity or not. It is well-known that most macro time series are non-stationary even though the issue has received virtually no attention in traditional panel regression analyses (Phillips and Moon, 2000: 264). The graphic analysis in Jaumotte, Lall and Papageorgiou (2013: 277-283) is a means for observing non-stationarity of Gini coefficients and their determinants. Table 3 shows unit root tests for the same variables as before. We use the Pesaran (2007) Panel Unit Root test (CIPS) whose null is that the variable is I(1). The analysis of results – with the majority of the tests on the level variables not rejecting – points out the non-stationarity of the Gini coefficients and some of their determinants, with particularly clear results for human capital. The only determinant of inequality for which the tests clearly reject non-stationarity is Openness. These results are confirmed by the tests on the differenced variables, which clearly reject the unit root case.

This section provides clear motivation that the heterogeneous panels unobserved common factors framework from Pesaran (2006) and followers is appropriate to analyze inequality determinants. The availability of data in quality and quantity allow for its correct implementation.

[13]Variables linked with disposable income have also been the focus of earlier papers. Barro (2000) uses a dummy to account for differences from the net income and consumption definition and gross income definition. This dummy is highly significant indicating that these variables measure in fact different phenomena. Jaumotte, Lall and Papageorgiou (2013: 276) also express concern about jointly analyzing income and expenditure-based Gini indexes. Results obtained with the market Gini coefficient from the SWIID (and its uncertainty-corrected version), which can compare with the ones presented in the paper can be provided by the authors.

Table 3: Panel Unit-Root tests

		(1)	(2)	(3)	(4)	(5)
Variable	Lag	Gini Net Income SWIID (>30)	Gini Net Income SWIID (>30, ./sd)	Human Capital	TFP	Open-ness
		Pesaran (2007) Test without Trend				
Zt-stat	0	3.08	-10.29***	17.17	-3.30***	-6.58***
p-value		(0.999)	(0.000)	(1.000)	(0.000)	(0.000)
Zt-stat	1	-0.406	-7.70***	3.51	-3.37***	-5.44***
p-value		(0.342)	(0.000)	(1.000)	(0.000)	(0.000)
Zt-stat	2	-2.39***	-2.93***	3.80	-3.27***	-2.35***
p-value		(0.008)	(0.002)	(1.000)	(0.001)	(0.009)
Zt-stat	3	1.62	-2.091***	3.15	-2.51***	-1.45*
p-value		(0.948)	(0.018)	(0.999)	(0.006)	(0.073)
		Pesaran (2007) Test with Trend				
Zt-stat	0	6.17	-5.846***	14.49	0.70	-7.65***
p-value		(1.000)	(0.000)	(1.000)	(0.758)	(0.000)
Zt-stat	1	1.23	-3.451***	5.20	0.09	-5.66***
p-value		(0.109)	(0.000)	(1.000)	(0.535)	(0.000)
Zt-stat	2	-4.45***	2.685	6.04	0.28	-2.73***
p-value		(0.000)	(0.354)	(1.000)	(0.610)	(0.002)
Zt-stat	3	0.35	2.752	6.52	1.82	-1.65**
p-value		(0.635)	(0.997)	(1.000)	(0.965)	(0.049)
Number of Countries		82	82	128	106	155
N. of Observations		3224	3224	6694	4994	7760
Avr. N. of Obs.		40.5	40.5	55.4	51.5	53.9

Note: All variables are in natural logarithms. >30 indicates that only cross-sections with more that 30 time-series observations are included. ./sd indicates when the Gini coefficient is divided by the source standard-deviation to account for data uncertainty. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1.

The next section explores the causal relationship between inequality and human capital.

4.2 Initial Analysis: causality between education and inequality

Trade and productivity (or technology) as determinants of inequality have been widely studied and the causal relationship from openness and technology to inequality is well founded in theory (see e.g. Hornstein, Krusel and Violante, 2005, Chakrabarti, 2000, and Richardson, 1995). However, the causality path from human capital to inequality is not so well founded. Despite the tremendous emphasis on the role of human capital in the skill-biased technological change and general purpose technology literatures, there are some microeconomic arguments that come from the economics of education field suggesting that inequality may decrease incentives to educate and thus decrease human capital (Stocké et al, 2011 and Gutierres and Tanaka, 2009 are good examples that emphasize the causality channel from inequality to education). It is important then to evaluate evidence in our data from the causality channel between human capital and inequality. We do this using a cointegration test for the null of no cointegration, the Westerlund (2007) test. Table 4 presents the tests when the causality is evaluated between human capital and the uncertainty-corrected Gini coefficient. The intuition is as follows. If the null is rejected for a test in which the dependent variable is inequality and simultaneously the null is not rejected for a test in which the dependent variable is human capital, then human capital has a causal effect on inequality and inequality has no causal effect on human capital. The pattern of results clearly suggests a causal relationship from human capital to inequality and not the other way around. This is valid for both the uncertainty-corrected measure presented in Table 4 and for the uncorrected measure. As in previous tests, we use only cross-sections that have availability of time-series data of 30 or more periods.

The next sections present results for the influence of human capital, TFP, and openness on inequality using heterogeneous panels methods.

4.3 Results: baseline specification

In this section we present the results for our baseline specification in equation (1), using the different Gini coefficient measures from the two sources that allow for greater country coverage.

Results in Tables 5 and 6 show that on average those determinants are not quite significant which may mean that there is great heterogeneity concerning effects of human capital, TFP, and openness. However, when significant, TFP has a positive sign, confirming the recent theories and some evidence that points to technological progress as one of the major causes of rising inequality and Openness tending decrease inequality.[14] Human capital is significant only in the regression for the Gini coefficient (from the SWIID database) - with a negative sign when the Gini coefficient is not corrected for uncertainty and for the restricted sample

[14]Barro (2000) fixed-effects estimations also show a positive effect of Openness. For Jaumotte, Lall, and Papageorgiou (2013), only financial openness increases inequality while trade openness decreases inequality. Information and Communication Technology share tends to increase inequality in Jaumotte, Lall, and Papageorgiou (2013).

Table 4: Cointegration tests

		(1)	(5)	(6)	(7)	(8)
	Lag	Trend	Test Gt	Test Ga	Test Pt	Test Pa
Dependent Variable	Gini Coefficient net income (>30, ./sd) (from SIIWD)					
	1	No	-2.400***	-10.22***	-9.630***	-9.588***
p-value			(0.001)	(0.004)	(0.002)	(0.000)
	1	Yes	-2.653**	-12.64	-10.79	-11.343**
p-value			(0.049)	(0.332)	(0.156)	(0.033)
	2	No	-2.353***	-8.232	-7.195	-7.261***
p-value			(0.001)	(0.174)	(0.342)	(0.000)
	2	Yes	-2.689**	-10.952	-7.660	-8.500
p-value			(0.031)	(0.768)	(0.995)	(0.630)
Dependent Variable	Human Capital (from PWT 8.0)					
	1	No	-1.826	-3.711	-5.713	-1.453
p-value			(0.401)	(0.998)	(0.861)	(0.998)
	1	Yes	-1.990	-7.607	-9.765	-6.089
p-value			(0.985)	(0.999)	(0.565)	(0.985)
	2	No	-1.879	-3.855	-5.448	-1.406
p-value			(0.298)	(0.998)	(0.912)	(0.999)
	2	Yes	-1.807	-7.110	-8.696	-5.479
p-value			(0.999)	(1.000)	(0.917)	(0.996)

Note: All variables are in natural logarithms. >30 indicates that only cross-sections with more that 30 time-series observations are included. All tests include a constant. ./sd indicates when the Gini coefficient is divided by the source standard-deviation to account for data uncertainty. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Rejection of H0 in Ga and Gt tests should be taken as evidence of cointegration of at least one of the cross-sectional units. Rejection of H0 in Pa and Pt tests should therefore be taken as evidence of cointegration for the panel as a whole.

Table 5: Inequality, Human Capital, TFP, and Openness

	(1)	(2)	(3)	(4)
Dependent Variable: Gini Measure	Gini Net Income WIID	Gini Consumption WIID	Gini Net post-tax; post-transfer SWIID	Gini Net post-tax; post-transfer SWIID >30
hcap	0.204 (0.385)	-0.282 (0.271)	-0.204 (0.195)	-0.272** (0.050)
TFP	-0.045 (0.325)	-0.151 (0.283)	0.001 (0.965)	-0.038 (0.314)
Open	-0.045 (0.251)	0.042 (0.627)	0.011 (0.431)	0.009 (0.600)
N Observ.	937	171	3300	2593
Avr. N Obs. Min-Max	16.7 5-54	9.5 5-24	32 7-52	38.1 21-52
Number Countries	56	18	103	68
Wald	3.04	2.60	2.31	5.13
CD-test (res)	–	–	–	1.95* (0.052)
Stat-test (res)	–	–	–	rejects I(1)
sig. signs /countries for *hcap*	↗(3)↘(7)	↗(0)↘(2)	↗(19)↘(39)	↗(7)↘(27)
sig. signs /countries for *TFP*	↗(1)↘(5)	↗(1)↘(1)	↗(27)↘(28)	↗(17)↘(23)
sig. signs /countries for *Open*	↗(3)↘(5)	↗(2)↘(0)	↗(21)↘(20)	↗(16)↘(12)

Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The list of countries that enter in columns (3) and (4) are provided in the Appendix B.

Table 6: Inequality, Human Capital, TFP, and Openness (high quality data)

	(1)	(2)	(3)	(4)
Dependent Variable: Gini Measure	Gini Net Income WIID - high quality	Gini Consumption WIID - high quality	Gini Net post-tax; post-transfer SWIID (./sd)	Gini Net post-tax; post-transfer SWIID (>30, ./sd)
hcap	-0.371 (0.378)	0.132 (0.654)	2.406*** (0.001)	3.737*** (0.000)
TFP	-0.028 (0.841)	0.251* (0.058)	-0.116 (0.391)	-0.230 (0.196)
Open	-0.234** (0.020)	0.107 (0.115)	0.002 (0.963)	-0.009 (0.865)
N Observ.	529	88	3300	2593
Avr. N Obs. Min-Max	16.5 5-44	9.8 5-23	32 7-52	38.1 21-52
Number Countries	32	9	103	68
Wald	6.24*	6.27*	11.04**	21.64***
CD-test (res)	–	–	–	-0.28 (0.782)
Stat-test (res)	–	–	–	reject I(1)
sig. signs /countries for *hcap*	↗(4)↘(5)	↗(0)↘(0)	↗(43)↘(9)	↗(35)↘(3)
sig. signs /countries for TFP	↗(6)↘(7)	↗(1)↘(0)	↗(15)↘(19)	↗(9)↘(12)
sig. signs /countries for *Open*	↗(5)↘(9)	↗(0)↘(0)	↗(19)↘(6)	↗(12)↘(5)

Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. A constant is included in the regressions but omitted from the Table. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The lists of countries that enter in columns (3) and (4) are provided in Appendix B. ./(sd) indicates when the Gini coefficient is divided by the source standard-deviation to account for data uncertainty.

with longer time-series within panels - Table 5, column (3) - and with a positive sign when the Gini coefficient is corrected for uncertainty - Table 6, columns (3) and (4). In the former case, an increase in 1% in human capital would imply a decrease of 0.27% in the uncorrected Gini coefficient. In the latter, however, a 1% increase in human capital would increase the corrected Gini coefficient from 2.4% to 3.7%.[15]

In fact, significant results change considerably when the measure of inequality changes and when Gini coefficients are corrected for quality (in the case of the WIID measures) or take into account the uncertainty caused by less information for some country-year pairs (as in the case of the SWIID measures). The heterogeneity of effects are indeed high, as can be observed by the count of significant effects by country, provided in the Table. While in columns (1) and (2) countries with significant results are never more than 25% of the number of countries considered in regressions, in the regressions presented in columns (4) and (5), the number of countries with significant results for each variable are usually more than 50% of the number of countries included in the regressions. The number of countries with significantly positive coefficients and the number of countries with significantly negative coefficients for TFP and openness are relatively balanced, thus yielding in general non-significant averaged coefficients. The only exception is the openness case in columns (3) and (4) in Table 6. In fact, despite an averaged non-significant coefficient, there are many more countries that present a positive and significant coefficient for openness than those that present negatively significant coefficients (19 and 12 against 5 and 6).

For human capital coefficients, in the uncorrected SWIID Gini regressions, the number of countries with significant negative coefficients (39 and 27 respectively for columns (3) and (4)) are far more than the countries with positive and significant coefficients (19 and 7 respectively for columns (3) and (4)) in the last columns of Table 5. For human capital coefficients, in the corrected SWIID Gini regressions, the picture is now switched: the number of countries with significant positive coefficients (43 and 35 respectively for columns (3) and (4)) are far more than the countries with positive and significant coefficients (9 and 3 respectively for columns (3) and (4)) in the last columns of Table 6.

The overall significance of regressors is higher in regressions in Table 6 than in regressions in Table 5, as shown by higher significance of the Wald tests. It is also worth noting that the cross-independence of the residuals is not rejected in column (4) in Table 6. Although cross-independence of residuals cannot be rejected in column (4) of Table 5, the value of the test indicates that cross-dependence is now much lower than the level affecting the regressors, as shown in Table 2.

The information sets used by the regressions differ a grat deal. In fact, while for the WIID measures the number of observations never reaches 1000 and the included countries are at most 56, for the SWIID variables the number of observations is between 2500 and 3300 and the number of countries between 68 and 103. For these wider coverage measures of inequality, we can conclude for a much more robust effect of human capital than the effects of TFP or openness, which are highly heterogenous. Additionally, correcting for uncertainty in the information set used to construct inequality measures has been essential to uncover an effect of human capital on inequality which is consistent with the theoretical literature on the issue.

It is possible now to present an idea of the countries for which significant effects were

[15] Alternatively, it can be said that for the same level of precision of the Gini coefficient, a 1% increase in human capital would increase the corrected Gini coefficient from 2.4% to 3.7%.

detected. The lists of those countries for regressions of columns (3) and (4) are listed in Appendix B. In regressions presented in Table 6, columns (3) and (4), the great majority of countries for which human capital is statistically significant, the coefficient is positive. The exceptions are Bulgaria, Burundi, Central African Republic, Cyprus, Latvia, Mongolia, Namibia, Romania, and Swazilandia. A complete matching with relatively poor countries is not possible although some of the richest countries in the world present a significantly positive effect of human capital on inequality. Those are Australia, Canada, Finland, Hong Kong, Italy, Netherlands and Norway, just to mention some of them. Countries in which TFP tends to increase inequality are, among others, France, Germany, Japan, Botswana, Bulgaria, and Chile and those with a negative effect are, for example, Austria, Belgium, Croatia, Iran, Israel, Korea, Russia, and the USA. Openness tends to increase inequality in Austria, Canada, Colombia, Estonia, Greece, Indonesia, Japan, Russia, Phillipines and the USA and to decrease inequality in Bulgaria, Jordan, Portugal, Romania, Slovenia and Taiwan. Also in these cases, it is not possible to present an *a priori* association between those countries and the respective level of income or some other common feature that may characterize them. Next, we split our sample into rich and poor countries and systematically evaluate the effects of human capital, TFP, and openness in each of the samples. We used the sample median for real GDP per capita as the threshold to split the sample. Countries with an average of GDP *per capita* above the median would be classified as rich countries. Results are in Table 7 and show that the positive effect of human capital on inequality, once it is corrected for uncertainty in data, occurs mainly in rich countries. In these countries a 1% increase in human capital would imply that the corrected Gini coefficient increase from 3.2% to 4%.

Below, we present a set of robustness analysis to evaluate the effect of human capital and TFP on inequality, using the uncertainty-corrected measure of the Gini coefficient.

4.4 Robustness: alternative specifications

In the robustness analysis we have implemented slightly modified common correlated effects estimators as suggested in recent literature. We include in regressions one or more further covariates in the form of cross-section averages, which helps to identify the unobserved common factors (in the spirit of Pesaran, Smith and Yamagata, 2013 and following what Eberhardt and Presbitero, 2014 did in an empirical implementation). To this end, we consider openness as a cross-section average, seeking to identify the unobserved common factors as linked with globalization and global integration (e.g. the entrance of China in global markets affecting all the countries). Column (1) in Table 8 presents these results. In column (2) in the same table we present regressions in which we identify the common unobserved factors as, not only globalization and integration (using the variable openness as cross-section average) but also technological spillovers (using the variable TFP as cross-section average). In column (3) we add to the set of possible unobserved common factors, production spillovers, including GDP per capita as a cross-section average. In column (4) we consider only openness as cross-section average and eliminate TFP from the regression. This regression aims to show that the robustness of the negative effect of human capital on inequality is not dependent on the presence of TFP, and thus, not dependent on the way this particular TFP measure is calculated. In this robustness analysis we consider as dependent variable the Gini coefficient

Table 7: Inequality, Human Capital, TFP, and Openness (Rich versus Poor countries)

	Rich Sample		Poor Sample	
	(1)	(2)	(3)	(4)
Dependent Variable: Gini Measure	Gini Net post-tax; post-transfer SWIID (./sd)	Gini Net post-tax; post-transfer SWIID (>30, ./sd)	Gini Net post-tax; post-transfer SWIID (./sd)	Gini Net post-tax; post-transfer SWIID (>30, ./sd)
hcap	4.043***	3.157***	1.169	0.518
	(0.002)	(0.005)	(0.239)	(0.487)
TFP	-0.127	-0.251	-0.041	-0.030
	(0.656)	(0.444)	(0.717)	(0.860)
Open	0.032	-0.119	0.001	-0.078
	(0.784)	(0.242)	(0.985)	(0.315)
N Observ.	1657	1431	1643	1162
Avr. N Obs.	36.8	40.9	28.3	35.2
Min-Max	12-52	22-52	7-48	21-48
Number Countries	45	35	58	33
Wald	9.77**	9.98**	1.52	1.52
CD-test (res)	–	-1.40 (0.162)	–	-0.79 (0.430)
Stat-test (res)	–	reject I(1)	–	reject I(1)

Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. A constant is included in the regressions but omitted from the Table. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The list of countries that enter in columns (3) and (4) are provided in the Appendix B. ./(sd) indicates when the Gini coefficient is divided by the source standard-deviation to account for data uncertainty.

(net definition) from SWIID, using only cross-sections with more than (or equal to) 30 time-series observations. This is done to allow for diagnostic testing. We will also describe the results obtained with the same variable from all the cross-sections (independently of time-series coverage).

In regressions in which production spillovers are not considered as cross-country common factor - columns (1), (2) and (4) - the effect of human capital is highly significant meaning that a 1% increase in human capital would imply a rise in the level of inequality that is around 3.8%. From these, regressions in which productivity spillovers are considered, columns (1) and (2) present residuals that show no evidence of stationarity or cross-country dependence. Regression residuals from column (4) regression present some evidence of cross-country dependence (yet much lower than in the regressors) and no evidence of stationarity. In fact, as in Eberhardt and Prebistero (2014), the introduction of additional cross-country averages in regressions helps to obtain cross-country independence of residuals. In the regression that includes production spillovers as a possible common factor - column (3) - the effect of human capital decreases quantitatively but maintains the high level of significance. In this case, a 1% increase in human capital would imply a rise in the level of inequality of around 1.9%. Additionally residuals show no evidence for cross-country dependence or stationarity. Wald tests point to high significance of the regressors. The majority of countries present significant coefficients for human capital (from 22 to 44, of which not more than 9 are significantly negative). A relatively high number of countries (27) also present significant coefficients for TFP - in column (1) - although in this case there is a balance between significantly positive and significantly negative results. The most significant individual change that occurred in those regressions that abstain from considering openness as a country-specific determinant of inequality, is the entrance of the USA to the set of countries for which a significantly positive effect of human capital occurs, a fact common to all the regressions in Table 8.

Regressions that include all the cross-sections (and not only those with high time-series coverage, as those in the Tables) would generally confirm those results. Regressions corresponding to those in columns (1), (2) and (4) slightly decrease the effect of human capital to a coefficient from 2.5 to 3.17 (with a high significance corresponding to p-values of 0.000). Regression corresponding to that in column (3) decreases the quantitative effect and the level of significance (to a value near 0.8 and a significance level of near 0.25).

4.5 Discussion

In this section we critically discuss our results and also present some information about additional tests that are not presented in the paper but that are available upon request. We present evidence on the effects of human capital, TFP and openness on inequality. To that end, we used a recent measure of inequality with high coverage (Solt, 2009) and also recently developed estimators that allow for country heterogeneity and are robust to country dependence, stationarity and endogeneity toward unobserved common factors (generally described in the survey from Eberhardt and Teal, 2011). We found that there is great heterogeneity concerning the effects of TFP and openness on inequality. There are countries with positive effects, those with negative effects and even others that present insignificant effects. Thus, theories that are not based on country heterogeneity to explain the relationship between technology,

Table 8: Inequality, Human Capital, TFP, and Openness

Dependent Variable	Gini Coefficient net income (./sd, >30) (from SIIWD)			
Vars. only as CS Avr.	Open	Open; TFP	Open; TFP; GDP p.c.	Open; without TFP
	(1)	(2)	(3)	(4)
hcap	3.801*** (0.000)	3.854*** (0.000)	1.984*** (0.001)	3.716*** (0.000)
TFP	-0.204 (0.248)	–	–	–
Open	–	–	–	–
N Observ.	2593	2855	2855	2855
Avr. N Obs. Min-Max	38.1 21-52	38.6 21-52	38.6 21-52	38.6 21-52
Number Countries	68	74	74	74
Wald	78.80***	97.11***	75.50***	133.90***
CD-test (res)	-0.20 (0.839)	0.81 (0.420)	-1.01 (0.314)	1.89* (0.058)
Stat-test (res)	Reject I(1)	Reject I(1)	Reject I(1)	Reject I(1)
sig. signs /countries for *hcap*	↗(37)↘(2)	↗(44)↘(6)	↗(22)↘(9)	↗(42)↘(8)
sig. signs /countries for *TFP*	↗(12)↘(15)	–	–	–
sig. signs /countries for *Open*	–	–	–	–

Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. A constant is included in the regressions but omitted from the Table. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The list of countries that enter in columns (3) and (4) are provided in the Appendix B. Vars. only as CS Avr. means variables that enter regressions only as cross-section average but not as country-specific variable. ./sd indicates when the Gini coefficient is divided by the source standard-deviation to account for data uncertainty.

openness and inequality miss an important part of the story. Institutions and history may be behind those heterogeneous effects. We also found a positive robust effect of human capital on inequality. This does not dismiss that some heterogeneous effects between different countries are also present. However an overwhelming majority of countries present positive effects such that the global effect is positive and significant among several different specifications. We also discovered that this influence of higher human capital in higher inequality is totally dependent on correcting the Gini coefficient for its standard-error. According to Solt (2009) the provided standard-error for the Gini coefficient aims to correct the remaining uncertainty in the estimations for the inequality measure. The standard-error measures the remaining error due to lack or poorer information available for some country-year pairs. Interestingly, ignoring this correction would yield a negative and significant effect of human capital on inequality, thus implying allegedly that human capital investments would decrease inequality, a result that would be in opposition to the most recent theories of the skill-biased technological change or general purpose technologies. A deep analysis of the data reveals that the negative sign of the coefficient for the uncorrected Gini index is due to poorer precision in Gini coefficients. For instance, restricting the regression of column (3) in Table 5 to values for the standard-error above the median would yield a significantly negative coefficient of -0.788 (with a p-value of 0.000) and doing the same to the regression of column (4) in the same Table would yield a coefficient of -0.596 (with a p-value of 0.010). Thus, there is a clear need to account for these differences in quality of the source data when assessing the determinants of inequality.

There are two main issues that might compromise our results: (1) the use of a certain measure of human capital and (2) the correction of the Gini measure with the source standard-error to account for different data quality across the world. Would it be possible that this effect is linked with the human capital variable used in this paper? In fact, measurement of human capital has always been somewhat controversial in the literature. The measure of human capital that is most used in the literature is that of Barro and Lee (2001), which has been criticized by e.g. Cohen and Soto (2007) due to measurement errors and sources. In fact, Cohen and Soto (2007) argued to have crucially increased the data quality when compared to their predecessors. Barro and Lee (2013), in the version 1.3. of the database, updated the data to incorporate the criticism. PWT 8.0 human capital variable used in this paper builds on Barro and Lee database, version 1.3. Additionally, the authors of PWT 8.0 filled in the years between the 5 year intervals provided by Barro and Lee, using linear interpolation and corrected the years of schooling to different returns from schooling by level of education following a Mincerian approach. There are, of course, some limitations of this measure, especially the fact that it does not distinguish the returns from schooling by country and by year. An exploration of the returns to schooling variability in a human capital measure would certainly be obtained at the cost of reducing the country coverage and measurement error. Thus, the human capital variable from PWT 8.0 is the human capital data with widest coverage, and thus the only that consistently allow for the use of heterogeneous panel data methods. In order to investigate whether the interpolation approach would have eliminated the significance of our results, we ran regressions that eliminated the interpolated observations. This greatly decreased the number of observations available for each regression from nearly 3200 observations to nearly 500 observations. Nevertheless, all regressions corresponding to specifications presented earlier in Table 8 maintain the highly significant positive signed human capital coefficient, with statistical significance of 5% or less.

The human capital variable construction and the results give us confidence that the obtained results must be common to any correct measure of human capital given that it has the wide time-series and cross-country coverage as does this one. As a consequence our strong effect of human capital on inequality has a non-negligible policy effect. Until now, and given the results in Barro (2000), the common wisdom has been that if some education increases inequality, it should be the higher levels of education. However, by construction, the employed measure of human capital strongly weights lower levels of education (due to higher returns for lower levels of education). Thus, the effect of education on inequality should also be due to lower levels of education. This has policy relevance as politicians should be aware of this effect in promoting education, even at the lower levels. Notwithstanding, this effect is absent from the poorer countries, which indicates no negative influence of education on inequality. Thus, generally, in poorer countries, policy may enhance education with no caution about rising inequality.

The second issue is related to the correction of the Gini coefficient. We did that by simply dividing the Gini coefficient by the standard-error, as explained above. This standard-error oscillates in the sample from 0.0016 to 15.43, which gives an idea of the difference in quality remaining in data and suggests the need to account for these quality heterogeneity. In fact, 25% of the observations present a standard-error below 0.5. Dividing the Gini coefficient by this standard-error would greatly magnify Gini coefficients with high precision. A correction that would not present that problem would be the division of the Gini coefficient by (1+standard-error).[16] With this, a high precision Gini coefficient - with a standard-error close to 0 - would not be increased so much, although a low precision coefficient would be decreased. The high significance of human capital positive coefficients hardly changes with this modification in the corrected Gini index in all the different specifications we present in the paper (corresponding to specifications in Tables 6 - columns (3) and (4)- Tables 7 and 8). The only expected difference in results is quantitative. With this alternative variable, a 1% increase in human capital would increase inequality by between 0.62% to 1.52% (compared to 1.98% to 3.85% with the baseline measure). The causal relationship between human capital and inequality in re gressions corresponding to specifications in Table 8, but in which all the cross-sections (and not restricted to the ones with larger time-series) are included, is also robust to the mentioned change in the definition of the corrected Gini coefficient.

5 Conclusion

There is scarce quantitative literature on the determinants of inequality. We contribute to that literature by evaluating potential determinants of inequality in a large panel data of countries. earlier attempts have faced problems with the coverage and quality of the income inequality data. We compare results using different inequality measures and conclude for (i) dismissal of a Kuznets curve and (ii) quite different results according to the different inequality measures used. We then begin to use a recent standardized measure of the Gini coefficient,

[16] The alternative proposed uncertainty-corrected measure is thus $\frac{GINI}{1+sd(GINI)}$, where *GINI* is the Gini index provided by SWIID and *sd(GINI)* is the standard-deviation of the Gini index, also provided by the SWIID and that corrects for uncertainty or measurement error within the sources. Results are provided in Appendix C.

due to Solt (2009) to evaluate human capital, TFP and openness as possible determinants of inequality. We conclude that this measure also needs to be corrected for differences in original data precision. Failure to do so would determine crucially different and misleading results concerning the influence of human capital on inequality. Fortunately, Solt (2009) also provides the means to implement such correction.

We found that inequality data, as well as other macroeconomic variables, are subject to cross-country dependence and stationarity and so, newly developed econometric methods designed to analyze moderate T, moderate N panels should be employed (Eberhardt and Teal, 2010). We proceeded along this line and implemented cointegration tests to evaluate the causality between human capital and inequality. Results indicate a strong channel from human capital to inequality.

Regressions based on heterogeneous panels methods indicate that there is great heterogeneity concerning the effects of TFP and openness on inequality. There are countries with positive effects, those with negative effects and even others that present insignificant effects. This yields overall non-significant effects of TFP and openness on inequality. We also found a positive robust effect of human capital on inequality once the Gini coefficient is corrected for differences in precision. This does not dismiss that some heterogeneous effects between different countries are also present. However, an overwhelming majority of countries present positive and significant effects which results is a very strong effect of human capital on inequality. Contrary to what may have been the current wisdom until now, it is not only tertiary education that tends to cause higher inequality, but the effect is highlighted with a measure that strongly weights lower levels of education.

These results suggest that theories that are not based on country heterogeneity to explain the relationship between technology, openness, and inequality may be unrealistic. In fact, institutions and history may be behind the heterogeneous effects of human capital, technology, and openness on inequality detected. Additionally, contrary to earlier evidence, the results in this paper suggest that human capital may be seen as the most important worldwide determinant of inequality.

References

[1] Acemoglu, D. (2002a). Technical change, inequality and the labor market. *Journal of Economic Literature*, 40:7–72.

[2] Acemoglu, D. (2002b). Directed technical change. *Review of Economic Studies*, 69:781–810.

[3] Acemoglu, D. (2003). Technology and Inequality. *NBER Reporter*.

[4] Aghion, P., Howitt, P., and Violante, G. (2002). General purpose technology and wage inequality. *Journal of Economic Growth*, 7(4):315–345.

[5] Barro, R. J. (2000). Inequality and Growth in a Panel of Countries. *Journal of Economic Growth*, 5(1):5–32.

[6] Barro, R. J. and Lee, J.-W. (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers, Oxford University Press*, 53(3):541–63.

[7] Barro, R. J. and Lee, J.-W. (2013). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104:184–198.

[8] Birchenall, J. A. (2001). Income distribution , human capital and economic growth in Colombia. *Journal of Development Economics*, 66(1):271–287.

[9] Caselli, F. (1999). Technological revolutions. *American Economic Review*, 89:78–102.

[10] Chakrabarti, A. (2000). Does Trade Cause Inequality ? *Journal of economic development*, 25(2):1–21.

[11] Cohen, D. and Soto, M. (2007). Growth and human capital: good data, good results. *Journal of Economic Growth*, 12(1):51–76.

[12] Deininger, K. and Squire, L. (1996). New Data Set Measuring Income Inequality. *World Bank Economic Review*, 10:565–591.

[13] Ding, S., Meriluoto, L., Reed, W. R., Tao, D., and Wu, H. (2011). The impact of agricultural technology adoption on income inequality in rural China: Evidence from southern Yunnan Province. *China Economic Review*, 22(3):344–356.

[14] Eberhardt, M., Helmers, C., and Strauss, H. (2013). Do spillovers matter when estimating private returns to RD? *The Review of Economics and Statistics*, 95(2):436–448.

[15] Eberhardt, M. and Presbitero, A. (2014). This Time They Are Different: Heterogeneity and Nonlinearity in the Relationship Between Debt and Growth. *version 13th February 2014*.

[16] Eberhardt, M. and Teal, F. (2011). Econometrics for Grumblers: a New Look At the Literature on Cross-Country Growth Empirics. *Journal of Economic Surveys*, 25(1):109–155.

[17] Eberhardt, M. and Teal, F. (2013a). No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis. *Oxford Bulletin of Economics and Statistics*, 75(6):914–939.

[18] Eberhardt, M. and Teal, F. (2013b). Structural Change and Cross-Country Growth Empirics. *World Bank Economic Review*, 27(2):229–271.

[19] Galor, O. and Moav, O. (2000). Ability biased technological transition, wage inequality within and across groups, and economic growth. *Quarterly Journal of Economics*, 115:469–497.

[20] Galor, O. and Tsiddon, D. (1997). Technological progress, mobility, and economic growth. *American Economic Review*, 87:362–382.

[21] Greenwood, J. and Yorukoglu, M. (1997). 1974. *Carnegie-Rochester Conference Series on Public Policy 46*, pages 49–96.

[22] Gutiérrez, C. and Tanaka, R. (2009). Inequality and education decisions in developing countries. *The Journal of Economic Inequality*, 7(1):55–81.

[23] Hornstein, A., Krusell, P., and Violante, G. L. (2005). Chapter 20 The Effects of Technical Change on Labor Market Inequalities. *Handbook of Economic Growth*, 1:1275–1370.

[24] Jaumotte, F., Lall, S., and Papageorgiou, C. (2013). Rising Income Inequality: Technology, or Trade and Financial Globalization? *IMF Economic Review*, 61(2):271–309.

[25] Martins, P. and Pereira, P. (2004). Does Education Reduce Wage Inequality? Quantile Regression Evidence From 16 Countries. *Labour Economics*, 11:355–371.

[26] Pesaran, M. H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels. *IZA Discussion Papers 1240.*

[27] Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012.

[28] Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2):265–312.

[29] Pesaran, M. H., Smith, V., and Yamagata, T. (2013). Panel Unit Root Tests in the Presence of a Multifactor Error Structure. *Journal of Econometrics*, 175(2):94–115.

[30] Pesaran, M. H. and Tosetti, E. (2011). Large Panels with Common Factors and Spatial Correlations. *Journal of Econometrics*, 161(2):182–202.

[31] Phillips, P. and Moon, H. (2000). Nonstationary panel data analysis: an overview of some recent developments. *Econometric Reviews*, 19(3):263–286.

[32] Psacharopoulos, G. (1994). Returns to investment in education:A global update. *World-Development*, 22(9):1325–1343.

[33] Ratts, J. and Stokke, H. (2013). Regional Convergence of Income and Education: Investigation of Distribution Dynamics. *Urban Studies, Published online before print August 19, 2013, doi: 10.1177/0042098013498625.*

[34] Richardson, J. D. (1995). Income Inequality and Trade: How to Think, What to Conclude. *Journal of Economic Perspectives*, 9(3):33–55.

[35] Solt, F. (2009). Standardizing the World Income Inequality Database. *Social Science Quarterly*, 90(2):231–242.

[36] Stocké, V., Blossfeld, H.-P., Hoenig, K., and Sixt, M. (2011). 7 Social inequality and educational decisions in the life course. *Zeitschrift für Erziehungswissenschaft*, 14(2):103–119.

[37] Wang, L. (2011). How Does Education Affect the Earnings Distribution in Urban China? *IZA Discussion Paper No. 6173.*

[38] Westerlund, J. (2007). Testing for Error Correction in Panel Data. *Oxford Bulletin of Economics and Statistics*, 69(6):709–748.

A Appendix: Specification Search

A.1 Without quality correction

Table A.1: Specification Search: without quality correction

Dependent Variable	lgini_in	lgini_c	lgini_ig	lgini_net	lgini_market
lrgdp_pc	1.85***	-0.13	0.64	0.15	-0.27
	(.000)	(.783)	(.122)	(.559)	(.287)
lgdp2	-.1***	0.01	-0.04	-0.005	0.02
	(.000)	(.310)	(.100)	(.723)	(.210)
lhc	-.61***	0.29	-0.36	-.41**	-.29*
	(.007)	(.399)	(.351)	(.010)	(.074)
ltfp	0.07	-0.02	.13*	0.01	0.08
	(.464)	(.852)	(.080)	(.878)	(.171)
lOpen	.03***	0.06	.04***	.04***	.04***
	(.000)	(.187)	(.001)	(.000)	(.000)
N Observ.	974	281	863	3310	3310
N of Countries	73	66	82	106	106

Notes: All variables are in natural logarithms. Values between parentheses are p-values from robust standard errors. Regressions include a complete set of time-dummies that are not shown in the table. Level of significance: *** for p-value<0.01; ** for p-value<0.05; * for p-value<0.1.

A.2 With quality correction

Table A.2: Specification Search: with quality correction

Dependent Variable	lgini_in	lgini_c	lgini_ig	lgini_net_st	lgini_market_st
lrgdp_pc	1.06**	-.89	1.07	1.17	1.15
	(.026)	(.634)	(.246)	(.192)	(.152)
lgdp2	-.06**	.08	-.07	-.06	-.05
	(.027)	(.485)	(.217)	(.207)	(.519)
lhc	-.52***	.08	-.53	-.3	-.37
	(.007)	(.893)	(.194)	(.639)	(.586)
ltfp	.15	.06	.3***	-.09	-.23
	(.134)	(.690)	(.000)	(.558)	(.528)
lOpen	.05***	.23***	.07	.04	.03
	(.008)	(.004)	(.312)	(.234)	(.123)
N Observ.	550	44	181	3310	3310
N of Countries	40	13	23	106	106

Notes: All variables are in natural logarithms. Values between parentheses are p-values from robust standard errors. Regressions include a complete set of time-dummies that are not shown in the table. Level of significance: *** for p-value<0.01; ** for p-value<0.05; * for p-value<0.1.

B Appendix: Lists of Countries

This section lists the countries used in the main regressions in the paper (Tables 5 and 6 - columns (3) and (4), Table 8).

B.1 Sample in Tables 5 and 6, columns (3)

Argentina, Armenia, Australia, Austria, Barbados, Belgium, Bolivia, Botswana, Brazil, Bulgaria, Burundi, Cameroon, Canada, Central African Republic, Chile, China, Colombia, Costa Rica, Cote d'Ivoire, Croatia, Cyprus, Czech Republic, Denmark, Dominican Republic, Ecuador, Egypt, Estonia, Fiji, Finland, France, Germany, Greece, Guatemala, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Korea, Republic of, Kyrgyz Republic, Latvia, Lesotho, Lithuania, Luxembourg, Malaysia, Malta, Mauritania, Mauritius, Mexico, Moldova, Mongolia, Morocco, Mozambique, Namibia, Netherlands, New Zealand, Niger, Norway, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russian Federation, Rwanda, Senegal, Serbia, Sierra Leone, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, Swaziland, Sweden, Switzerland, Taiwan, Tajikistan, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United Kingdom, United States, Uruguay, Venezuela, Zimbabwe.

B.2 Sample in Tables 5 and 6, columns (4), and Table 8, column (1)

Argentina, Australia, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Cote d'Ivoire, Denmark, Egypt, Estonia, Fiji, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Korea, Republic of, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Mauritius, Mexico, Moldova, Morocco, Netherlands, New Zealand, Norway, Panama, Peru, Philippines, Poland, Portugal, Russian Federation, Sierra Leone, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Tanzania, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United Kingdom, United States, Uruguay, Venezuela.

B.3 Sample in Tables 8, columns (2), (3) and (4)

Argentina, Australia, Bangladesh, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Cote d'Ivoire, Denmark, Egypt, El Salvador, Estonia, Fiji, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Korea, Republic of, Kyrgyz Republic, Latvia, Lithuania, Malawi, Malaysia, Mauritius, Mexico, Moldova, Morocco, Nepal, Netherlands, New Zealand, Norway, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Russian Federation, Sierra Leone, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Tanzania, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, United Kingdom, United States, Uruguay, Venezuela, Zambia.

C Appendix: Alternative Corrected Gini index

Table C.1: Inequality, Human Capital, TFP, and Openness

	(1)	(2)
Dependent Variable: Gini Measure	Gini Net post-tax; post-transfer SWIID ./(1+sd)	Gini Net post-tax; post-transfer SWIID ./(1+sd),>30
hcap	1.10***	1.44***
	(.004)	(.000)
TFP	.006	-0.058
	(.931)	(0.498)
Open	.02	0.02
	(.460)	(0.461)
N Observ.	3300	2593
Avr. N Obs.	32	38.1
Min-Max	7-52	21-52
Number Countries	103	68
Wald	9.01**	15.39***
CD-test (res)	–	1.10 (0.272)
Stat-test (res)	–	rejects I(1)
sig. signs /countries for *hcap*	↗(38)↘(12)	↗(31)↘(4)
sig. signs /countries for *TFP*	↗(19)↘(19)	↗(11)↘(16)
sig. signs /countries for *Open*	↗(17)↘(5)	↗(15)↘(4)

Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. A constant is included in the regressions but omitted from the Table. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The lists of countries that enter in columns (3) and (4) are provided in the Appendix B. ./(1+sd) indicates when the Gini coefficient is divided by 1 plus the source standard-deviation to account for data uncertainty.

Table C.2: Inequality, Human Capital, TFP, and Openness

Dependent Variable	Gini Coefficient net income (./(1+sd), >30) (from SWIID)			
Vars. only as CS Avr.	Open	Open; TFP	Open; TFP; GDP p.c.	Open; without TFP
	(1)	(2)	(3)	(4)
hcap	1.31***	1.54***	0.62**	1.52***
	(0.001)	(0.000)	(0.035)	(0.000)
TFP	-0.064	–	–	–
	(0.441)			
Open	–	–	–	–
N Observ.	2593	2855	2855	2855
Avr. N Obs.	38.1	38.6	38.6	38.6
Min-Max	21-52	21-52	21-52	21-52
Number Countries	68	74	74	74
Wald	55.98***	68.92***	34.68***	34.68***
CD-test (res)	1.15	1.14	0.21	0.39
	(0.250)	(0.254)	(0.834)	(0.694)
Stat-test (res)	Reject I(1)	Reject I(1)	Reject I(1)	Reject I(1)
sig. signs /countries for *hcap*	↗(13)↘(3)	↗(41)↘(9)	↗(19)↘(8)	↗(42)↘(9)
sig. signs /countries for *TFP*	↗(13)↘(18)	–	–	–
sig. signs /countries for *Open*	–	–	–	–

Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. A constant is included in the regressions but omitted from the Table. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The lists of countries that enter in columns (3) and (4) are provided in the Appendix B. Vars. only as CS Avr. means variables that only enter regression as cross-section average but not as country-specific variable. ./(1+sd) indicates when the Gini coefficient is divided by 1 plus the source standard-deviation to account for data uncertainty.

Table C.3: Inequality, Human Capital, TFP, and Openness (Rich versus Poor countries)

	Rich Sample		Poor Sample	
	(1)	(2)	(3)	(4)
Dependent Variable: Gini Measure	Gini Net post-tax; post-transfer SWIID (./sd)	Gini Net post-tax; post-transfer SWIID (>30, ./sd)	Gini Net post-tax; post-transfer SWIID (./sd)	Gini Net post-tax; post-transfer SWIID (>30, ./sd)
hcap	1.49***	1.29***	0.76	0.499
	(0.003)	(0.004)	(0.170)	(0.451)
TFP	0.02	-0.03	-0.046	-0.097
	(0.853)	(0.792)	(0.563)	(0.396)
Open	0.01	-0.04	0.024	-0.016
	(0.777)	(0.534)	(0.395)	(0.712)
N Observ.	1657	1431	1643	1162
Avr. N Obs.	36.8	40.9	28.3	35.2
Min-Max	12-52	22-52	7-48	21-48
Number Countries	45	35	58	33
Wald	9.22**	8.62**	2.94	1.43
CD-test (res)	–	1.02 (0.307)	–	-0.06 (0.955)
Stat-test (res)	–	reject I(1)	–	reject I(1)

Note: Note: Dependent Variables natural logarithm of the Gini coefficient. All variables are in natural logarithms. A constant is included in the regressions but omitted from the Table. Values between parentheses below coefficients are p-values from robust (clustered) standard errors. Level of significance: *** for p-value<0.01; **for p-value<0.05;* for p-value<0.1. Wald test is a joint significance test for the regressors. CD-test is a Pesaran (2004) cross-section dependence test on the null of cross-section independence done on the residuals from the regression (p-value presented between parentheses). Stat-test is the Pesaran (2007) unit root test made on the residuals. This test used 3 lags and rejects I(1) means that in all lags the test of unit root rejects. The lists of countries that enter in columns (3) and (4) are provided in the Appendix B. ./(1+sd) indicates when the Gini coefficient is divided by 1 plus the source standard-deviation to account for data uncertainty.