Economic Complexity, Human Capital and Income Inequality: A Cross-Country Analysis

Lee, Kang-Kook and Vu, Trung

Ritsumeikan University, Shiga, Japan, University of Otago, Dunedin, New Zealand

27 June 2019

Economic Complexity, Human Capital and Income Inequality: A Cross-Country Analysis[†]

Kang-Kook Lee

Faculty of Economics, Ritsumeikan University, Shiga, Japan

Trung V. Vu[‡]

Department of Economics, University of Otago, Dunedin, New Zealand

Abstract

This paper investigates the relationship between economic complexity, a measure of economic structures, and income inequality. Using cross-country OLS regression analysis, we show that countries with economic structures geared toward complex products enjoy a lower level of inequality. Human capital is found to magnify this correlation. Different measures of human capital also have differentiated interaction effects. Concerns about the endogeneity bias of OLS estimates motivate us to estimate a dynamic panel data model, using a system GMM estimator. We find that an increase in economic complexity provokes higher inequality.

Key words: Economic complexity, human capital, income inequality, system GMM.

JEL Classification: F16, F40, O11, O15

[†] We are grateful to David Flath for helpful discussions and numerous inputs on earlier drafts of this manuscript. Our gratitude also goes to two anonymous referees and the journal co-editor for their constructive comments that helped improve the quality of this paper. We thank Dorian Owen for several useful suggestions. The second author thanks the Japanese government for financial support provided in the form of a Monbukagakusho scholarship. The usual disclaimer applies.

[‡] Correspondence address: Department of Economics, University of Otago, PO Box 56, Dunedin 9054, New Zealand. Email: trung.vu@postgrad.otago.ac.nz or trungvu.ueb@gmail.com

1. Introduction

Since the 1950s, there has been an intense discussion on the relationship between economic development and income inequality, following the path-breaking paper of Kuznets (Kuznets, 1955). This line of research attempts to verify empirically the well-known inverted U-shaped "Kuznets curve". But findings remain inconclusive, with some papers providing empirical support for the Kuznets hypothesis (Ahluwalia, 1976; Barro, 2008), others finding support for only the downward sloping part of the curve (Perotti, 1996; Galbraith, 2007; Palma, 2011), and still others refuting the systematic relationship between economic growth and the level of income inequality (Deininger and Squire, 1996).

A glaring limitation of the above strand of literature is that scholars have typically made use of aggregate indicators of economic development, such as GDP or the total contribution of agriculture, manufacturing or services to GDP (Hartmann et al., 2017). Economists, however, contend that an aggregate monetary measure of economic size is an inadequate proxy for economic development when it comes to explaining variations in income inequality (Kuznets, 1973; Stiglitz et al., 2010; Sbardella et al., 2017; Hartmann et al., 2017). The basic reasoning is that they ignore the conventional wisdom about the heterogeneous economic outcomes, depending on what a country produces and exports. Because these indicators are calculated based upon broad categories, they fail to capture effectively the complexity level of production.[1] By contrast, economic performance is not necessarily uniform across a diverse range of products, which justifies a disaggregated approach.

Indeed, the relationship between a country's economic structures, reflecting what it produces and exports, and economic growth and income inequality, has long been recognized in the literature. Specifically, Lewis (1955), Rostow (1959), Kuznets and Murphy (1966), Kaldor (1967) and Chenery and Taylor (1968), among others, emphasize the role of economic transformation, that is the process of moving from activities with lower productivity to those with higher productivity. Accordingly, specialization in different activities is associated with differentiated economic outcomes. Countries whose economic structures are geared toward sophisticated commodities develop faster than those specializing in simple products (Felipe et al., 2012).

[1] See Hartmann et al. (2017) for more discussions.

This paper departs from the current literature by investigating the relationship between economic complexity and income inequality. We build upon the recent literature which argues that income distribution depends not only on economic growth but also on the type of growth (Hartmann et al., 2017). This line of research calls for better measures of economic development, rather than an aggregated indicator of income (Costanza et al., 2009; Stiglitz et al., 2010). Accordingly, economic development involves not only increases in economic size, but also brings about variations in technologies, human capital, and institutions, etc. All these factors are defined as non-tradeable capabilities, which determine a country's productivity level (Hidalgo and Hausmann, 2009). They are captured in the economic complexity index, which measures productive capabilities embedded in economic structures. The index is constructed based on the method proposed by Hidalgo and Hausmann (2009). An economy is complex if it can produce a diverse range of products (diversity) that are not widely produced by many countries (ubiquity). Complexity reveals information about a country's human capital, technology, and institutions. It is strongly correlated with economic performance (Hausmann et al., 2007; Hidalgo and Hausmann, 2009). Nevertheless, little is known about the extent to which complexity affects income inequality.

Examining the relationship between economic complexity and inequality, our study contributes to the literature in several dimensions. First, we employ an extended dataset of income inequality with the broadest coverage of countries and years to examine the distributional effects of economic complexity. Second, we demonstrate that countries with higher economic complexity enjoy a lower level of inequality, using cross-country OLS regression. Furthermore, human capital magnifies this negative correlation. Breaking down the sample based on a country's income level, we find that secondary education and tertiary education have differentiated interaction effects on inequality. We also note that this strong correlation does not necessarily imply causality due to endogeneity concerns. Third, we estimate a dynamic panel data model, using a system GMM estimator that caters for the potential endogeneity bias. We differ from previous research by separating the short- and longer-run effects of complexity on inequality. An immediate and a longer-term increase in complexity is found to worsen income distribution.

The paper proceeds as follows. The second section discusses economic complexity and its relationship with income inequality. Next, we present our econometric methods. The fourth section examines empirical findings. The study concludes by summarizing the main results.

2. Economic complexity and income inequality

2.1. What is economic complexity?

Economic complexity, as constructed by Hidalgo and Hausmann (2009), is a measure of the set of productive capabilities available in a country. Here, capabilities are defined as non-tradeable inputs required to make a product (Hidalgo and Hausmann, 2009). These inputs, for instance, include human capital, technology, institutions, and the legal system. At the firm level, capabilities encompass the set of "know-how" or working practices, which are created when individuals interact and work closely together in a group. Broadly speaking, capabilities at firm level capture the organizational capacity, associated with the formation, management, and operation of production activities (Felipe et al., 2012). The diversity of these capabilities, as measured by economic complexity, determines a country's productivity level. Thus, economic complexity helps explain cross-country divergence in income per capita (Hidalgo and Hausmann, 2009).

Hidalgo and Hausmann (2009) construct an economic complexity index based on the so-called "Method of Reflections". First, the authors calculate the Revealed Comparative Advantage (RCA) index to determine whether a country is effectively exporting a given product.[2] Second, they define two concepts, "diversity" and "ubiquity", as the number of products a country can export with RCA and the number of countries that have an advantage in exporting a given product, respectively. Third, a country is considered to be complex if it can export a diverse range of products with RCA (high diversity) while a product is viewed as complex if it is not exported popularly by many countries (low ubiquity). The basic reasoning is that producing a given commodity requires the set of productive capabilities. Hence, the diversity of these capabilities is revealed by the number of products a country exports with RCA. At the same time, complex products that require many hard-to-find capabilities are exported by only few countries. Finally, economic complexity is obtained by combining information on country diversity and product ubiquity. A country is complex if it can export a diverse range of products (high diversity) and their products are not popularly exported by other countries (low ubiquity).

The calculation can be shown below, following Hidalgo and Hausmann (2009).

[2] The RCA indicator, developed by Balassa (1965), is the ratio of the share of a country's exports of a given product in the total export basket to the overall share across countries. A country is considered as a significant exporter of a given commodity if its RCA is greater than or equal to unity($RCA \geq 1$).

$$DIVERSITY = k_{c,0} = \sum_p RCA_{cp} \quad [1]$$

$$UBIQUITY = k_{p,0} = \sum_c RCA_{cp} \quad [2]$$

where c stands for the country, p denotes the commodity. RCA_{cp} is an indicator variable, taking the value of 1 if a country has a revealed comparative advantage in product p, and 0 otherwise.

However, these two indicators, considered separately, are imprecise measures of economic complexity. For instance, a country may export a diverse range of products, simply because of its economic size. In this case, product diversity may provide biased information on the availability of capabilities within a country. To address this bias, Hidalgo and Hausmann (2009) propose the method of reflection to calculate economic complexity by computing jointly and iteratively the mean value of the diversity and ubiquity, calculated in the previous iteration. The formulae can be represented as follow.

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_p RCA_{cp} k_{p,N-1} \quad [3]$$

$$k_{p,N} = \frac{1}{k_{p,0}} \sum_c RCA_{cp} k_{c,N-1} \quad [4]$$

where N stands for the number of iterations.

Substituting Eq. [4] into [3], we obtain the following equation (Hausmann et al., 2014).

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_p RCA_{cp} \frac{1}{k_{p,0}} \sum_{c'} RCA_{c'p}.k_{c',N-2} = \sum_{c'} k_{c',N-2} \sum_p \frac{RCA_{cp} RCA_{c'p}}{k_{c,0} k_{p,0}}$$

This can be rewritten as:

$$k_{c,N} = \sum_{c'} \widetilde{M}_{cc'} k_{c',N-2}$$

in which

$$\widetilde{M}_{cc'} = \sum_p \frac{RCA_{cp} RCA_{c'p}}{k_{c,0} k_{p,0}}$$

$\widetilde{M}_{cc'}$ is basically a matrix connecting countries exporting similar products. To discount common products, the matrix is weighted by the inverse of the ubiquity of a product. Then it is normalized by the diversity of a country (Hartmann et al., 2017).

Finally, economic complexity is computed as:

$$ECI_c = \frac{K_c - \langle K \rangle}{std(K)}$$

where K_c is the eigenvector of $\widetilde{M}_{cc'}$, associated with the second largest eigenvalue. The vector associated with the largest eigenvalue is a vector of ones (Hausmann et al., 2014; Caldarelli et al., 2012).

2.2. Why does economic complexity matter for income inequality?

Hartmann et al. (2017) is the only paper that documents a relationship between economic complexity and income inequality. Using OLS and panel fixed-effects regression, the authors demonstrate that economic complexity lowers income inequality. Diverging from this study, we show that countries with higher economic complexity enjoy a lower level of inequality, but an increase in complexity is associated with higher inequality within a country.

First, there are several mechanisms by which economic complexity improves a country's income distribution. Individuals living in a complex economy have more employment and occupational opportunities, and a vast majority of disseminated skills and knowledge (Hartmann, 2014). Economic structures geared toward a diverse range of sophisticated products are associated with a relatively flat occupational structure (Hartmann et al., 2017). Furthermore, workers in complex industries have higher bargaining power, lowering income disparity. Complex economic structures are also associated with better institutions and higher unionization, which reduces income inequality (Hartmann, 2014).

Considering a country whose economic structures are geared toward simple commodities (e.g., raw materials and agricultural products), inequality tends to proliferate because of the following reasons. First, the income of the majority of workers in this country depends on economic activities characterized by decreasing returns to scale and low productivity. By contrast, a small fraction of the population enjoys a higher income, stemming from limited activities with higher productivity. Second, the diffusion of limited knowledge and skills would be occupied by small groups of individuals, leaving them with an economic premium. Third, individuals at the bottom of the income distribution are constrained by limited learning and occupational opportunities. This further increases the income gap. Finally, the presence of one main sector producing primary commodities is associated with a vertical hierarchy in the occupational structure to manage a huge cadre of unskilled labors. By contrast, complex economic structures, characterized by products with increasing returns to scale, lower inequality by enhancing the lifetime earnings of workers who produce them (Constantine, 2017). A diverse range of opportunities also provides the poor with several means of moving up in the social stratification (Hidalgo, 2015; Hartmann et al., 2017). Therefore, economic complexity helps to lower income inequality.

Second, we argue that an increase in economic complexity can be associated with a higher level of inequality within a country. When economic structures become more complex, there appears a growing demand for qualified workers (Constantine, 2017). This is attributable to changes in the set of required capabilities (Hodgson, 2003). Specifically, this process results in some new capabilities emerging and becoming increasingly essential, in accordance with the growing richness of the complex economy. This is described as a creative-destruction process, which leads to the emergence of new sectors and the disappearance of traditional ones (Hartmann, 2014). In this process, workers may find it difficult to alter from one specialism to another (Hartmann, 2014). There would be opportunities for the low-income group to climb up the career ladder, but continual retraining would be of great importance in acquiring remunerative employment. It is much easier and less costly for skilled workers to advance because they are endowed with greater capabilities. Skilled workers with the capacity to adapt to the changing sets of capabilities eventually secure an improved economic premium and become active agents in society. For this reason, we argue that growing economic complexity benefits skilled workers more than unskilled workers, exacerbating the income gap within a country.

Finally, we suppose that human capital magnifies the negative correlation between economic complexity and inequality. Specifically, human capital plays an important role in tackling inequality when a country experiences structural transformation toward complex industries. Endogenous growth theory, attempting to explain cross-country income disparities, observes that human capital is an essential determinant of growth (Romer, 2012). Furthermore, the literature has long acknowledged that economic development is the process of structural change in a country's productive structures (Rostow, 1959; Lewis, 1955). This occurs through resource transformation from simple to complex industries, achieved by accumulating new productive capabilities. In this process, human capital allows a country to learn and acquire more capabilities faster. Consequently, human capital helps a country to engage in the production of sophisticated commodities, thereby enhancing its economic complexity (Hidalgo and Hausmann, 2009; Hausmann et al., 2014). Improvements in education also allow individuals to take advantage of diverse opportunities due to growing economic complexity. It, therefore, lowers the possibility that complexity worsens a country's income distribution due to skill-bias. Hence, we argue that human capital is the key factor that intertwines with economic complexity in affecting income inequality.

3. Econometric models

3.1. Cross-country models

A cross-country model is used to test the hypothesis that more complex economies have more equal income distribution. We calculate the average value for each variable across the period from 1980 to 2014 to obtain cross-sectional data for 96 countries.[3] The dependent variable is the Gini coefficient, a measure of the income inequality level. This index is estimated using post-tax, post-transfer measures of household income in the Standardized World Income Inequality Database. The main variable of interest, economic complexity, is taken from MIT's Observatory of Economic Complexity (atlas.media.mit.edu).

First, income inequality is regressed on economic complexity, controlling for the level of income, its square, years of schooling, population, trade openness, and FDI. The model specification is expressed as:

$$GINI_i = \beta_0 + \beta_1 ECI_i + \beta_2 lnGDPPC_i + \beta_3 lnGDPPCSQ_i + \beta_4 Schooling_i + \beta_5 X_i + \varepsilon_i \quad [5]$$

where i stands for country i, $GINI$ represents the Gini coefficient for disposable income, ECI represents the Economic Complexity index, $lnGDPPC$ and $lnGDPPCSQ$ are the natural log of GDP per capita and its squared term, respectively, $Schooling$ denotes years of schooling, X is a vector of other control variables, ε represents the error term.

Second, the effects of economic complexity on income inequality are conditional on human capital, as argued earlier. Therefore, an additional model is specified, including interaction terms between economic complexity and different measures of human capital. The model specification is presented below.

$$GINI_i = \beta_0 + \beta_1 ECI_i + \beta_2 lnGDPPC_i + \beta_3 lnGDPPCSQ_i + \beta_4 Schooling_i + \beta_5 ECI \times Schooling_i + \varepsilon_i \quad [6]$$

As the results show in the following section, most of the control variables included in Eq. [5] are statistically insignificant. The basic reasoning is that some of the distributional effect of those factors (e.g., trade openness and FDI) may be captured by ECI. Hence, we do not include

[3] Data for years of schooling are available for only 96 countries. Hence, the number of observations drops to 96 when it is included in the model.

them in Eq. [6] and [7]. However, our conclusions remain largely insensitive to including those variables.

Finally, we estimate Eq. [6] using two different samples, based on the level of income, to test whether different measures of human capital have differentiated effects in developing and developed countries.

Estimating Eq. [5] and [6] requires some attention to reverse causality that may run from income inequality to economic complexity. A higher level of income inequality impedes the development of education, health outcomes, investment, and social consensus (Persson and Tabellini, 1994; Benabou, 1996; Alesina and Perotti, 1996; Easterly, 2007; Berg and Ostry, 2013). In addition, economic complexity is a correlate of a country's innovative outcomes, its increasing innovative capacity, and the implementation of those innovations (Sweet and Eterovic Maggio, 2015). A high level of income inequality, which deters economic development, can also lower the innovation creation process by limiting the demand for innovative products and services (Zweimüller, 2000). We may well think that income inequality may impede the incentives for innovation. These are essential for a country to obtain new productive capabilities, which allow it to engage in more sophisticated production activities. Thus, the estimates of causal effects of complexity on inequality may reflect reverse causation. This potential bias is ignored by Hartmann et al. (2017), which is the only study linking complexity and inequality. The consistency of estimates requires using an instrumental variable that affects income inequality only through economic complexity. This is a challenging task, given that the main determinants of economic complexity have direct effects on income inequality (e.g., government consumption, the abundance of arable land, and capital per worker).[4] This motivates us to estimate a dynamic panel data model in the following section.

In short, we estimate Eq. [5] and [6] using OLS, albeit recognizing that some endogeneity bias may exist. Hence, the strong correlation between economic complexity and income inequality, as demonstrated later, may not necessarily imply causality.

3.2. A dynamic panel data model

To test the effects of an increase in economic complexity on income inequality in a dynamic model, this section analyses an unbalanced panel dataset, comprising 113 countries with 5-year

[4] See Daude et al. (2016) and Zhu and Fu (2013).

averages from 1965 to 2014. The motivation for taking 5-year averages for all time-variant variables is that we can reduce short-run fluctuations due to business cycles. In addition, our main variables of interest, including economic complexity, GDP and inequality, change slowly year-over-year. This paper follows Blundell and Bond (1998) to estimate the following model.

$$GINI_{i,t} = \beta_0 + \beta_1 GINI_{i,t-1} + \beta_2 ECI_{i,t} + \beta_3 ECI_{i,t-1} + \beta_4 lnGDPPC_{i,t} + \beta_5 lnGDPPCSQ_{i,t} + \beta_6 Schooling_{i,t} + \beta_7 Population_{i,t} + \delta_i + \gamma_t + \varepsilon_{i,t} \quad [7]$$

where i denotes country i (i=1, 2,…113). The time dimension runs from 1 to 10, denoting 10 periods of 5-year averages from 1965 to 2014. $\mathrm{GINI_{it}}$ represents the disposable income Gini coefficients of country i in period t. $\mathrm{GINI_{it-1}}$ reflects the dynamic characteristics of the model. $\mathrm{Schooling_{it}}$ is years of schooling of country i in period t.[5] $\mathrm{Population_{it}}$ is the total population; δ_i is the country fixed effects; γ_t stands for time effects; ε_{it} is the error term; and β is the vector of estimated coefficients.[6]

The reason why we include lagged inequality as an explanatory variable in Eq. [7] is that aggregate indicators, such as GDP or inequality, exhibit high time persistence. Indeed, serial correlation of income inequality across time has been well established in the literature (Banerjee and Duflo, 2003; Kurita and Kurosaki, 2011). This motivates us to estimate a linear dynamic panel data.[7]

We include both current and lagged economic complexity to capture the short- and long-run effects of an increase in economic complexity on income inequality.[8] In particular, the current economic complexity reflects the distributional effects of an immediate increase in economic complexity. We also compute the long-run effects, following Wooldridge (2013, page 635). It is necessary to conduct formal tests of selecting the appropriate lag length of ECI when calculating the long-run effect. However, this task is challenging given that our time-

---

[5] Data for years of schooling is obtained from Barro and Lee (2013), but this data is available at 5-year intervals. Therefore, we assigned the educational data at the beginning year of every period. For instance, data on years of schooling in 1965 will be used for the first period (1965-1969), and data on years of schooling in 2010 will be used for the final period (2010 -2014). Years of schooling represent the stock of human capital, accumulated over every 5 years. Hence, we can match them with 5-year averages data to estimate Eq. [7].

[6] Unlike the OLS models, the panel data models do not include institutional variables because data is relatively sparse.

[7] Income inequality varies slowly within countries over years. This suggests that there exist some unobserved factors explaining this time persistence. In this case, fixed-effects estimates are biased if these factors are correlated with our explanatory variables. Addressing this issue requires including the past income inequality levels as an explanatory variable.

[8] We sincerely thank an anonymous referee for this suggestion.

series dimension is relatively short. We also tried including the second and third lag of ECI, and the estimated coefficients are positive but statistically insignificant. Including further lags of ECI also results in the model suffering from instrument proliferation when we have too many endogenous regressors. For these reasons, we include only a one-period lag of ECI.

Estimation of Eq. [7] requires the following issues to be addressed. First, the lagged dependent variable, included as an explanatory variable, is endogenous. Second, estimating the distributional effects of economic complexity, GDP, and human capital suffers from endogeneity bias due to measurement errors, omitted variable bias, and reverse causality, as argued earlier. The solution to these problems is to use an exogenous instrument that needs to satisfy the exclusion restrictions. However, this is a challenging task as we discussed in the previous section. Motivated by this issue, we estimate Eq. [7] using internal instruments, which are the higher order lags of our endogenous variables.[9]

Arellano and Bond (1991) propose the use of first differences of the variables to remove bias arising from unobserved country heterogeneity. The first difference equation can be expressed as follows.

$$\Delta GINI_{i,t} = \beta_1 \Delta GINI_{i,t-1} + \beta_2 \Delta ECI_{i,t} + \beta_3 \Delta ECI_{i,t-1} + \beta_4 \Delta X_{i,t} + \Delta \varepsilon_{i,t} \qquad [8]$$

in which Δ stands for the first difference operator. It is easy to recognize that $\Delta \mathrm{GINI}_{\mathrm{it}-1}$ is endogenous in this case because

$$E(\Delta GINI_{i,t-1} \Delta \varepsilon_{i,t}) = E[(GINI_{i,t-1} - GINI_{i,t-2})(\varepsilon_{i,t} - \varepsilon_{i,t-1})] \neq 0$$

Under the assumption that the error terms in Eq. [7] are serially uncorrelated ($\mathrm{cov}(\varepsilon_{i,t}, \varepsilon_{i,t-j}) = 0$ if $j \neq 0$), the moment conditions for the difference GMM is specified as follows

$$E[GINI_{i,t-j} \Delta \varepsilon_{i,t}] = 0 \text{ for } t = 3,4, \dots T \text{ and } j \geq 2 \qquad [9]$$

Hence, two periods or further lags of the dependent variable are valid instruments for $\Delta GINI_{i,t-1}$ in Eq. [8].[10]

We apply the same logic to address the second problem caused by reverse causality running from inequality to complexity. Under the assumption that $E[ECI_{i,t} \varepsilon_{i,j}] \neq 0$ for $j \leq t$ and $E[ECI_{i,t} \varepsilon_{i,j}] = 0$ for $j > t$, the moment condition is specified as

$$E[ECI_{i,t-j} \Delta \varepsilon_{i,t}] = 0 \text{ for } t = 3,4, \dots T \text{ and } j \geq 2 \qquad [10]$$

[9] We treat all variables on the right-hand side, except population, as endogenous variables.
[10] See Arellano and Bond (1991) for more detailed discussions.

Hence, two periods or further lags of economic complexity are valid instruments for both $\Delta \mathrm{ECI}_{\mathrm{i,t}}$ and $\Delta \mathrm{ECI}_{\mathrm{i,t-1}}$. The difference GMM employs two sets of moment condition [9] and [10] to estimate Eq. [7]

However, the difference estimation method suffers from the problem of weak instruments, given the high time-persistence of the dependent variable and the short time dimension (Arellano and Bover, 1995; Blundell and Bond, 1998). This problem can be addressed by using the system GMM estimator, developed by Blundell and Bond (1998). The basic idea is that we specify additional moment conditions by adding the level form equation to the difference equation. The lagged differences of the explanatory variables are used as instruments for the equations in levels. The additional moment conditions can be described as below:

$$E\left[\Delta GINI_{i,t-1}\varepsilon_{i,t}\right] = 0 \text{ for } t = 3,4, \dots T \quad [11]$$

$$E\left[\Delta ECI_{i,t-1}\varepsilon_{i,t}\right] = 0 \text{ for } t = 3,4, \dots T \quad [12]$$

In short, the system GMM employs the moment conditions [9], [10], [11], and [12] to estimate Eq. [7]. Because our dependent variable is time persisting, as discussed above, the system GMM estimator is more suitable. Furthermore, it can address the potential endogeneity of our variables by using appropriate lags of endogenous variables as valid instruments.

Finally, we perform several specification tests to check the validity of our results. First, the Arellano-Bond AR(2) test is used to test the absence of second-order autocorrelation. Failure to reject the null of the AR(2) test implies no autocorrelation in the second-differenced errors, validating the moment conditions. In this case, we can perform the system GMM estimation without changing the instrument sets. By contrast, rejection of the null of the AR(2) test supports the presence of autocorrelation in the error term of order one or higher. This requires re-estimating the models using further lags (Cameron and Trivedi, 2005). For this reason, we perform the Arellano-Bond AR(3) test to check the absence of third-order autocorrelation in the error term. Failure to reject the null validates our instruments used in the modified system GMM estimator. Second, we check the Hansen J-test of over-identifying restrictions. The null hypothesis is that the over-identifying restrictions are valid, of which the failure to reject substantiates the overall validity of the instruments and the model. Furthermore, the system GMM estimator may suffer from instrument proliferation problems because it employs a larger set of instruments, relative to difference GMM. We solve this problem by limiting the number of lags of the endogenous variables.[11]

4. Results and discussions

4.1. OLS estimation

The OLS estimation results are presented in Table 1 to Table 3. We find a strong negative effect of economic complexity on income inequality. It can be observed from Table 1 that the estimated coefficients of economic complexity are negative and statistically significant at the 1% level in all models. This means that countries whose productive structures are more complex have a lower level of income inequality, ceteris-paribus. This empirical finding is in line with the OLS estimates of Hartmann et al. (2017). It is also consistent with the argument that higher complexity provide individuals with more occupational choices, learning opportunities, a higher sustained income, a flat hierarchical occupational structure, thus lowering income inequality. The potential determinants of inequality, including trade openness, FDI, education and institutions, are statistically insignificant when ECI is employ to explain cross-country differences in inequality. This shows that economic complexity is a strong predictor of income inequality across countries.

(Insert Table 1 about here)

Using the cross-country average data from 1980 to 2014 for 96 countries, this paper provides empirical support for the validity of the Kuznets proposition after controlling for economic complexity. The coefficients for GDP per capita have a positive sign, and are statistically significant at the 1% level. Also, the estimated coefficients for the squared term of GDP per capita are negative and statistically significant at the 1% level. Hence, the inverted U-shaped curve is recovered, using this sample of 96 countries.

(Insert Table 2 about here)

Table 2 examines the effects of the interaction between economic complexity and human capital on income inequality. The effects of ECI on inequality at the mean value of the log of schooling can be computed from column (1) of Table 1 as:

[11] Specifically, we use the third and the fourth lags of the endogenous variables. The number of instruments should be ideally smaller than the number of cross-sectional units, which is the number of countries in our context (Roodman, 2009). We use this criterion to select the appropriate lags.

$$\frac{\partial GINI}{\partial ECI} = 0.191 - 0.110Ln(Schooling) = 0.191 - 0.110 * 1.932 = -0.021$$

This is consistent with our findings in Table 1. Hence, it further supports the negative correlation between economic complexity and inequality, using the OLS estimation method.

The estimated coefficients of interaction terms between complexity and human capital are found to be negative and statistically significant at the 1% level. The strong significance and the negative sign of the estimated coefficients suggest that human capital considerably bolsters the negative effect of complexity on income inequality. Our results are robust to using different measures of human capital. Countries that manage to complement social policies, which improve the quality of human capital, with industrial policies, which diversify toward more sophisticated products, will have a lower level of income inequality. Hartmann et al. (2016) argue that high-performing Asian economies (HPAEs) such as South Korea, Malaysia, and Thailand have managed to diversify their economic structures into more complex industrial products, creating wider employment opportunities. At the same time, they invested substantially in education that prepared workers for new occupational choices. Consequently, those countries maintained a lower level of income inequality from 1970 until 2010, along with growing economic complexity. By combining social and economic policies, the benefits of their growing level of economic complexity were spread into the whole society, evidenced by decreasing income inequality (Stiglitz, 1996).

(Insert Table 3 about here)

We further perform different regressions for two sub-samples of developing and developed countries. It is interesting to observe that the estimated coefficients for secondary education and its interaction with complexity are noticeably higher than for tertiary education as shown in Table 2 and 3. Estimation results indicate that secondary education has much stronger interaction effects on inequality than tertiary education, in both high-income countries and low- and middle-income countries. More importantly, the interaction between education and complexity has a positive effect on inequality in high-income countries, but a negative effect in developing countries. Hence, the negative joint effect when pooling data across countries (as shown in Table 2) is mainly driven by the group of developing economies. The reasons for this are various.

First, the link between income inequality and education expansion, measured by increasing years of schooling or by decreasing education inequality, has been subject to numerous empirical tests. Barro (2000) considers the effects of different levels of education on income inequality, and finds that secondary education is negatively associated with inequality while tertiary education is positively correlated with inequality. Examining the distributional implications of education in Greece, Tsakloglou and Antoninis (1999) find that public transfers in primary and secondary education are associated with a more equal distribution of income, while transfers in tertiary education have limited distributional effects. The explanation for such a finding is that investments in secondary education benefit the middle quintiles (the majority of households) while the impact of tertiary education is found only in the top quintiles. Hence, transfers in secondary education have much higher and stronger effects in decreasing inequality than transfers in tertiary education. Furthermore, stronger distributional effects of secondary education are found in other studies (Abdullah et al., 2015).

Second, these results are not surprising when considering that the majority of the sample consists of developing countries. Over recent decades, most developing economies with high economic growth rates have experienced rapid structural transformation, which is dominated by resource reallocation away from agriculture and toward manufacturing. According to Atkin (2016), this industrialization process is mainly driven by low-skill manufacturing exports. Thus, there has been a growing demand for low-skill manufacturing labor in most developing countries. To the extent that an improvement in secondary education responds to this increased demand and improves productive capabilities, the joint effect of secondary education and complexity is more pronounced in developing countries. The negative interaction effect is also mainly driven by developing countries.

Third, Zhu and Li (2017) argue that the stronger interaction effect of secondary education in developing countries can be explained by looking at the method of computing the economic complexity index. The first step of measuring the complexity level is to select products in which a country can produce and export with revealed comparative advantage. In developing countries, it is far more common to reveal advantage in low-skill manufacturing and labor-intensive production activities, which require lower levels of education. Therefore, the labor market of developing economies is likely to absorb workers with secondary education. Also, complexity can be bolstered by employing workers with secondary education in the expanding industries.

Finally, the effect on income distribution of the interaction between complexity and human capital is found to be positive in high-income countries. This means that the higher quality of education, together with growing economic complexity, drives widening income disparities. While the structural transformation in developing countries has been driven by the resource movement toward manufacturing, it has been accompanied by service-based activities in developed economies. Wage variation in the manufacturing sector is usually lower than that in the service sector. Meanwhile, manufacturing jobs are characterized by the presence of labor unions and collective bargaining that act as an impediment to worsening income inequality. Hence, the structural transformation in high-income countries from labor-intensive manufacturing to more service-intensive activities is associated with higher inequality. Furthermore, we argue this could be due to the saturation of high-quality human capital in developed countries. Highly complex economies usually require a smaller workforce as industries tend to be more highly automated and technologically advanced. Therefore, workers with highly advanced skills or advanced degrees will be rewarded by significantly higher income, but these industries also need fewer workers to drive the value chain. Therefore, higher inequality and income differences should be seen between workers with exceptional skills or advanced degrees compared with others. This suggests that the interaction effects of economic complexity and education on income inequality, as argued before, is more important in developing countries than in advanced countries.

We also test whether the quality of institutions magnifies the correlation between economic complexity and income inequality. Indeed, institutions act as a catalyst for new economic activities (North, 1990). They are also beneficial for investment in productive capabilities, including new technologies and skills (North, 1990), and the so-called "self-discovery" process (Hausmann and Rodrik, 2003). In addition, the quality of institutions promotes innovations through creating entrepreneurship, and disseminating knowledge and technologies (Schumpeter, 1934; Brown, 1999). Countries with better institutions, therefore, acquire productive capabilities faster, thus improving their economic complexity and income distribution. The estimated coefficients are significantly negative, supporting this argument. However, they are statistically insignificant when we include the interaction between human capital and ECI in the model. This indicates that human capital is the most important factor that interacts with complexity to affect inequality.[12]

4.2. System GMM estimates

[12] To conserve space, we do not report the results of the interaction between ECI and institutions. However, they are available upon request.

Estimation results of the dynamic panel data, using a system GMM estimator, are presented in Table 4. The failure to reject the Hansen J-test substantiates the model specifications and the validity of our instruments. However, the AR(2) test shows that the error terms are serially correlated. For this reason, we re-estimate our model, using an adjusted instrument set. The AR(3) test indicates the absence of serial correlation in the error term, validating our instruments.[13]

(Insert Table 4 about here)

The estimated coefficients of the lagged dependent variable are relatively large in terms of magnitude, and are highly significant as shown in Table 4. This implies that income inequality shows high time persistence, which further justifies the relevance of the system GMM estimation method.

The estimated coefficients of ECI are positive and statistically significant at the 1% level. This implicates that an immediate increase in economic complexity provokes higher income inequality. Specifically, a one-unit of standard deviation of complexity is associated with approximately a 0.02-unit increase in inequality, as shown in column 6 of Table 4. It should be noted that this result is opposite to that of Hartmann et al. (2017), as discussed below.

Although the estimated coefficients of the lagged ECI are not statistically significant in several models, they are statistically significant at the 5% level after we control for schooling. Following Wooldridge (2013, page 635), we compute the long-run effects of an increase in economic complexity on income inequality, given the estimated coefficients in column 6 of Table 4.[14]

[13] We used the third and fourth lags of endogenous instruments to avoid the problem of instrument proliferation in system GMM estimator. This choice is also motivated by the criterion that the number of instruments should be ideally smaller than the number of countries (Roodman, 2009) and the AR(3) test results. Our main findings are robust to limiting lags of endogenous variables used as instruments.

[14] We can re-write Eq. [7] as:

$GINI_t - \beta_1 GINI_{t-1} = \beta_0 + \beta_2 ECI_t + \beta_3 ECI_{t-1} + \beta_4 \mathrm{lnGDPPC}_{i,t} + \beta_5 \mathrm{lnGDPPCSQ}_{i,t} + \beta_6 \mathrm{Schooling}_{i,t} + \beta_7 \mathrm{Population}_{i,t} + \delta_i + \gamma_t + \varepsilon_{i,t}$.

This is equivalent to:

$GINI_t(1 - \beta_1 L) = \beta_0 + ECI_t(\beta_2 + \beta_3 L) + \beta_4 \mathrm{lnGDPPC}_{i,t} + \beta_5 \mathrm{lnGDPPCSQ}_{i,t} + \beta_6 \mathrm{Schooling}_{i,t} + \beta_7 \mathrm{Population}_{i,t} + \delta_i + \gamma_t + \varepsilon_{i,t}$ where L denote the lag operator (i.e. $LGINI_t = GINI_{t-1}$).

In long-run equilibrium, L=1 ($GINI_t = GINI_{t-1}$). Hence, $GINI_t(1 - \beta_1) = \beta_0 + ECI_t(\beta_2 + \beta_3) + \beta_4 \mathrm{lnGDPPC}_{i,t} + \beta_5 \mathrm{lnGDPPCSQ}_{i,t} + \beta_6 \mathrm{Schooling}_{i,t} + \beta_7 \mathrm{Population}_{i,t} + \delta_i + \gamma_t + \varepsilon_{i,t}$.

The partial effects of an increase in ECI on inequality is calculated as below:

$$\left.\frac{\partial GINI}{\partial ECI}\right|_{LR} = \frac{\beta_2 + \beta_3}{1 - \beta_1}$$

$$\left.\frac{\partial GINI}{\partial ECI}\right|_{LR} = \frac{\beta_2 + \beta_3}{1 - \beta_1} = \frac{0.022 - 0.014}{1 - 0.931} = 0.116 > 0$$

Hence, a long-term increase in economic complexity is associated with an increase in the degree of income inequality. Specifically, a one-unit increase in ECI in the long-run results in a 0.116-unit increase in inequality. The magnitude of the long-term effects, however, are larger than the short-term effects.

Our results suggest that when the economy manages to diversify its production activities and engage in the production of more complex products, the distribution of income becomes less equal. This empirical finding supports the arguments of Hodgson (2003) and Hartmann (2014). In particular, an increase in economic complexity is achieved when a country acquires new productive capabilities that allow it to make more complex products. On the one hand, this process brings about new employment and education opportunities, which tends to lower inequality. On the other hand, it can also create a situation of "winner-take-all" that intensifies income inequality within and between countries because low-skilled workers and low-income countries are at a greater disadvantage. Hodgson (2003) argues that the growing diversity and sophistication of products are associated with the emergence of new specialisms. Hartmann (2014) refers to this as the "creative-destruction" process in which new specialisms emerge and replace outdated ones. Workers who have advanced and transferable skills can learn quickly, and are exposed to less risk when adapting to this structural change. Hence, they tend to attract an economic premium.

Furthermore, economic complexity exacerbates inequality when it results in the problem of structural dependence and underdevelopment of a country's less developed areas (Myrdal, 1957; Hartmann, 2014). The growing diversity of complex industries in developed regions attracts inflows of qualified workers from a "periphery" of impoverished and underdeveloped areas, enriching the complexity levels of the former while worsening the economic issues of the latter. Consequently, the income disparity among regions increases. If we consider the issue from a cross-national perspective, growing complexity, which might be facilitated by automation and artificial intelligence, lowers the demand for low-skilled jobs in developing countries. This happens because developed nations are bringing back manufacturing and industrial jobs in which low-income countries traditionally have a comparative advantage. Thus, income disparity between countries escalates.

See Wooldridge (2013) for discussions on the long-run propensity in distributed lag models.

It is important to compare our empirical findings with those of Hartmann et al. (2017). The authors perform a cross-country fixed-effects panel regression with 10-year averages from 1963 to 2008, and find that an increase in complexity lowers inequality. We also estimate a static panel data, using 5-year and 10-year averages, but employ another dataset of income inequality with the broadest coverage of countries and years. Fixed-effects estimates show that the estimated coefficients of ECI is consistently positive, but they are statistically insignificant in some cases.[15] However, as discussed earlier, fixed-effects estimates may be biased due to the potential endogeneity of complexity and the persistence of the dependent variable. Addressing these problems, we estimate a dynamic panel data, using a system GMM estimator. Furthermore, our paper also differs from Hartmann et al. (2017) in demonstrating that an increase in economic complexity is associated with higher degrees of inequality in both short- and long-run, using the dynamic panel data estimation. The shortcoming of Hartmann et al. (2017) is that this paper does not include the lagged ECI, thus failing to capture the long-run effects of an increase in complexity on inequality. Using a new dataset of inequality with the broadest coverage of countries and years, we find that an increase in economic complexity is associated with a higher level of income inequality, controlling for the potential endogeneity bias.

To sum up, we estimate a dynamic panel data, using a system GMM estimator. Results show that an increase in economic complexity results in higher income inequality in both short- and long-run. The magnitude of the long-run effects is larger than that of the short-run effects.

5. Conclusions

This paper attempts to shed some light on the relationship between economic complexity and income inequality. In doing so, we estimate cross-country OLS regressions and a dynamic panel data model.

The cross-sectional OLS regressions show that countries with higher economic complexity enjoy lower levels of income inequality. This finding indicates that economic complexity is a strong predictor of income inequality. Furthermore, human capital magnifies the negative correlation between complexity and inequality. Countries endowed with better and improved human capital are able to enhance economic structures. Consequently, this reinforces the distributional effects of economic complexity, leaving them with a lower level of inequality.

[15] Fixed-effects estimation results are not reported to save space, but are available upon request.

We also demonstrate that secondary education plays a more important role in interacting with economic complexity, relative to tertiary education. Although these findings document a strong correlation, they do not necessarily imply causality because of potential endogeneity concerns.

Motivated by the possible endogeneity bias of OLS regressions, we estimate a dynamic panel data to consider the time-varying effects and address the potential endogeneity bias, using a system GMM estimator. We also include economic complexity and its lagged term to capture the short- and long-term effects. We find that an increase in economic complexity is associated with a higher level of income inequality, which is in contrast with the OLS estimates. This is consistent with the notion that when the economy experiences structural changes toward more sophisticated products, the degree of income inequality increases. In particular, the estimated coefficients of lagged economic complexity are significantly negative. The calculation of the long-term effects shows a positive correlation between an increase in complexity and the level of inequality. Hence, an increase in economic complexity is associated with higher income inequality.

Reference

Abdullah, A., Doucouliagos, H. & Manning, E. (2015) "Does Education Reduce Income Inequality? A Meta-Regression Analysis", *Journal of Economic Surveys,* Vol. 29, No. 2, pp. 301-316.

Ahluwalia, M. S. (1976) "Income Distribution and Development: Some Stylized Facts", *American Economic Review,* Vol. 66, No. 2, pp. 128-135.

Alesina, A. & Perotti, R. (1996) "Income Distribution, Political Instability, and Investment", *European Economic Review,* Vol. 40, No. 6, pp. 1203-1228.

Arellano, M. & Bond, S. (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies,* Vol. 58, No. 2, pp. 277-297.

Arellano, M. & Bover, O. (1995) "Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics,* Vol. 68, No. 1, pp. 29-51.

Atkin, D. (2016) "Endogenous Skill Acquisition and Export Manufacturing in Mexico", *American Economic Review,* Vol. 106, No. 8, pp. 2046-85.

Balassa, B. (1965) "Trade Liberalisation and “Revealed” Comparative Advantage", *Manchester School,* Vol. 33, No. 2, pp. 99-123.

Banerjee, A. V. & Duflo, E. (2003) "Inequality and Growth: What Can the Data Say?", *Journal of Economic Growth,* Vol. 8, No. 3, pp. 267-299.

Barro, R. J. (2000) "Inequality and Growth in a Panel of Countries", *Journal of Economic Growth,* Vol. 5, No. 1, pp. 5-32.

Barro, R. J. (2008) "Inequality and Growth Revisited". Asian Development Bank: ADB Working paper series on regional economic integration.

Barro, R. J. & Lee, J. W. (2013) "A New Data Set of Educational Attainment in the World, 1950–2010", *Journal of Development Economics,* Vol. 104, pp. 184-198.

Benabou, R. (1996) "Inequality and Growth", *NBER Macroeconomics Annual,* Vol. 11, pp. 11-74.

Berg, A. G. & Ostry, J. D. (2013) "Inequality and Unsustainable Growth: Two Sides of the Same Coin?", *International Organisations Research Journal,* Vol. 8, No. 4, pp. 77-99.

Blundell, R. & Bond, S. (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics,* Vol. 87, No. 1, pp. 115-143.

Brown, P. (1999) "Globalisation and the Political Economy of High Skills", *Journal of Education and Work,* Vol. 12, No. 3, pp. 233-251.

Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A. & Tacchella, A. (2012) "A Network Analysis of Countries' Export Flows: Firm Grounds for the Building Blocks of the Economy", *Plos One,* Vol. 7, No. 10, pp. e47278.

Cameron, A. C. & Trivedi, P. K. (2005) *Microeconometrics: Methods and Applications,* Cambridge: Cambridge University Press.

Chenery, H. B. & Taylor, L. (1968) "Development Patterns: Among Countries and over Time", *Review of Economics and Statistics,* Vol. 50, No. 4, pp. 391-416.

Constantine, C. (2017) "Economic Structures, Institutions and Economic Performance", *Journal of Economic Structures,* Vol. 6, No. 1, pp. 1-18.

Costanza, R., Hart, M., Talberth, J. & Posner, S. (2009) "Beyond GDP: The Need for New Measures of Progress", *The Pardee Papers*.

Dahlberg, S., Holmberg, S., Rothstein, B., Khomenko, A. & Svensson, R. (2016) "The Quality of Government Basic Dataset, Version Jan16", University of Gothenburg: The Quality of Government Institute. University of Gothenburg: The Quality of Government Institute.

Daude, C., Nagengast, A. & Perea, J. R. (2016) "Productive Capabilities: An Empirical Analysis of Their Drivers", *Journal of International Trade & Economic Development,* Vol. 25, No. 4, pp. 504-535.

Deininger, K. & Squire, L. (1996) "A New Data Set Measuring Income Inequality", *World Bank Economic Review,* Vol. 10, No. 3, pp. 565-591.

Easterly, W. (2007) "Inequality Does Cause Underdevelopment: Insights from a New Instrument", *Journal of Development Economics,* Vol. 84, No. 2, pp. 755-776.

Felipe, J., Kumar, U., Abdon, A. & Bacate, M. (2012) "Product Complexity and Economic Development", *Structural Change and Economic Dynamics,* Vol. 23, No. 1, pp. 36-68.

Galbraith, J. K. (2007) "Global Inequality and Global Macroeconomics", *Journal of Policy Modeling,* Vol. 29, No. 4, pp. 587-607.

Hartmann, D. (2014) *Economic Complexity and Human Development: How Economic Diversification and Social Networks Affect Human Agency and Welfare,* New York: Routledge.

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M. & Hidalgo, C. A. (2017) "Linking Economic Complexity, Institutions, and Income Inequality", *World Development,* Vol. 93, No. Supplement C, pp. 75-93.

Hartmann, D., Jara-Figueroa, C., Guevara, M., Simoes, A. & Hidalgo, C. S. A. (2016) "The Structural Constraints of Income Inequality in Latin America", *Integration & Trade Journal No. 40*.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A. & Yildirim, M. A. (2014) *The Atlas of Economic Complexity: Mapping Paths to Prosperity,* Cambridge: MIT Press.

Hausmann, R., Hwang, J. & Rodrik, D. (2007) "What You Export Matters", *Journal of Economic Growth,* Vol. 12, No. 1, pp. 1-25.

Hausmann, R. & Rodrik, D. (2003) "Economic Development as Self-Discovery", *Journal of Development Economics,* Vol. 72, No. 2, pp. 603-633.

Hidalgo, C. (2015) *Why Information Grows: The Evolution of Order, from Atoms to Economies,* New York: Basic Books.

Hidalgo, C. A. & Hausmann, R. (2009) "The Building Blocks of Economic Complexity", *Proceedings of the National Academy of Sciences,* Vol. 106, No. 26, pp. 10570-10575.

Hodgson, G. M. (2003) "Capitalism, Complexity, and Inequality", *Journal of Economic Issues,* Vol. 37, No. 2, pp. 471-478.

Kaldor, N. (1967) *Strategic Factors in Economic Development,* New York: Cornell University Press.

Kurita, K. & Kurosaki, T. (2011) "Dynamics of Growth, Poverty and Inequality: A Panel Analysis of Regional Data from Thailand and the Philippines", *Asian Economic Journal,* Vol. 25, No. 1, pp. 3-33.

Kuznets, S. (1955) "Economic Growth and Income Inequality", *American Economic Review,* Vol. 45, No. 1, pp. 1-28.

Kuznets, S. (1973) "Modern Economic Growth: Findings and Reflections", *American Economic Review,* Vol. 63, No. 3, pp. 247-258.

Kuznets, S. & Murphy, J. T. (1966) *Modern Economic Growth: Rate, Structure, and Spread,* New Haven: Yale University Press.

Lewis, W. (1955) *The Theory of Economic Growth,* London: George Allen & Unwin.

Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions,* London: Gerald Duckworth.

North, D. (1990) *Institution, Institutional Change and Economic Performance,* Cambridge Cambridge University Press.

Palma, J. G. (2011) "Homogeneous Middles Vs. Heterogeneous Tails, and the End of the 'Inverted-U': It's All About the Share of the Rich", *Development and Change,* Vol. 42, No. 1, pp. 87-153.

Perotti, R. (1996) "Growth, Income Distribution, and Democracy: What the Data Say", *Journal of Economic growth,* Vol. 1, No. 2, pp. 149-187.

Persson, T. & Tabellini, G. (1994) "Is Inequality Harmful for Growth?", *American Economic Review,* Vol. 84, No. 3, pp. 600-621.

Romer, D. (2012) *Advanced Macroeconomics,* New York: McGraw-Hill.

Roodman, D. (2009) "A Note on the Theme of Too Many Instruments", *Oxford Bulletin of Economics and Statistics,* Vol. 71, No. 1, pp. 135-158.

Rostow, W. W. (1959) "The Stages of Economic Growth", *Economic History Review,* Vol. 12, No. 1, pp. 1-16.

Sbardella, A., Pugliese, E. & Pietronero, L. (2017) "Economic Development and Wage Inequality: A Complex System Analysis", *PloS One,* Vol. 12, No. 9, pp. e0182774.

Schumpeter, J. A. (1934) *The Theory of Economic Development – an Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle,* Cambridge, MA: Havard University Press.

Stiglitz, J. E. (1996) "Some Lessons from the East Asian Miracle", *World Bank Research Observer,* Vol. 11, No. 2, pp. 151-177.

Stiglitz, J. E., Sen, A. & Fitoussi, J.-P. (2010) "Report by the Commission on the Measurement of Economic Performance and Social Progress". Paris: Commission on the Measurement of Economic Performance and Social Progress.

Sweet, C. M. & Eterovic Maggio, D. S. (2015) "Do Stronger Intellectual Property Rights Increase Innovation?", *World Development,* Vol. 66, No. Supplement C, pp. 665-677.

Tsakloglou, P. & Antoninis, M. (1999) "On the Distributional Impact of Public Education: Evidence from Greece", *Economics of Education Review,* Vol. 18, No. 4, pp. 439-452.

Wooldridge, J. M. (2013) *Introductory Econometrics: A Modern Approach,* Mason: South-Western Cengage Learning.

Zhu, S. & Fu, X. (2013) "Drivers of Export Upgrading", *World Development,* Vol. 51, pp. 221-233.

Zhu, S. & Li, R. (2017) "Economic Complexity, Human Capital and Economic Growth: Empirical Research Based on Cross-Country Panel Data", *Applied Economics,* Vol. 49, No. 38, pp. 3815-3828.

Zweimüller, J. (2000) "Schumpeterian Entrepreneurs Meet Engel's Law: The Impact of Inequality on Innovation-Driven Growth", *Journal of Economic Growth,* Vol. 5, No. 2, pp. 185-206.

Table 1. Basic OLS estimates of the effects of economic complexity on income inequality.

VARIABLES	(1) GINI	(2) GINI	(3) GINI	(4) GINI	(5) GINI
ECI	-0.043***	-0.041***	-0.046***	-0.045***	-0.044***
	(0.013)	(0.014)	(0.015)	(0.016)	(0.016)
Ln(GDPPC)	0.280***	0.269***	0.274***	0.275***	0.275***
	(0.065)	(0.075)	(0.077)	(0.077)	(0.077)
Ln(GDPPCSQ)	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Ln(Schooling)	-0.028	-0.027	-0.025	-0.024	-0.025
	(0.028)	(0.029)	(0.029)	(0.029)	(0.029)
Rule of law		-0.005	-0.002	-0.002	-0.002
		(0.013)	(0.013)	(0.014)	(0.014)
Ln(Population)			0.004	0.003	0.003
			(0.005)	(0.006)	(0.006)
Ln(Trade)				-0.006	-0.009
				(0.020)	(0.024)
Ln(FDI)					0.002
					(0.011)
Constant	-0.732***	-0.694**	-0.794**	-0.742**	-0.738**
	(0.259)	(0.288)	(0.324)	(0.354)	(0.358)
Observations	96	95	95	95	95
R-squared	0.477	0.481	0.485	0.485	0.486

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2. OLS estimates of interaction effects.

VARIABLES	(1) GINI	(2) GINI	(3) GINI
ECI	0.191***	-0.001	-0.071***
	(0.042)	(0.022)	(0.019)
Ln(GDPPC)	0.156**	0.192***	0.192**
	(0.060)	(0.070)	(0.078)
Ln(GDPPCSQ)	-0.009***	-0.011***	-0.011**
	(0.003)	(0.004)	(0.004)
Ln(Schooling)	-0.077***		
	(0.026)		
ECI*Ln(Schooling)	-0.110***		
	(0.020)		
Ln(Secondary)		-0.042**	
		(0.017)	
ECI*Ln(Secondary)		-0.042**	
		(0.016)	
Ln(Tertiary)			-0.018
			(0.013)
ECI*Ln(Tertiary)			-0.023**
			(0.011)
Constant	-0.092	-0.379	-0.428
	(0.243)	(0.295)	(0.351)
Observations	96	96	96
R-squared	0.578	0.512	0.502

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3. OLS estimates (by sub-samples).

VARIABLES	High-income countries		Low- and middle-income countries	
	(1)	(2)	(3)	(4)
	GINI	GINI	GINI	GINI
ECI	-0.269***	-0.015	0.043**	-0.091**
	(0.055)	(0.017)	(0.019)	(0.038)
Ln(GDPPC)	0.536	0.434	0.247*	0.182
	(0.503)	(0.602)	(0.139)	(0.145)
Ln(GDPPCSQ)	-0.027	-0.021	-0.014	-0.010
	(0.025)	(0.030)	(0.009)	(0.009)
Ln(Secondary)	-0.140***		-0.082***	
	(0.028)		(0.021)	
ECI*Ln(Secondary)	0.171***		-0.085***	
	(0.041)		(0.021)	
Ln(Tertiary)		-0.075***		-0.041**
		(0.027)		(0.017)
ECI*Ln(Tertiary)		0.061**		-0.047***
		(0.023)		(0.016)
Constant	-2.081	-1.877	-0.579	-0.439
	(2.519)	(3.031)	(0.528)	(0.571)
Observations	36	36	58	58
R-squared	0.607	0.544	0.257	0.198

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Two-step system GMM estimates.

VARIABLES	(1) GINI	(2) GINI	(3) GINI	(4) GINI	(5) GINI	(6) GINI
$GINI_{t-1}$	0.996***	0.975***	0.975***	0.942***	0.969***	0.931***
	(0.032)	(0.026)	(0.032)	(0.026)	(0.030)	(0.025)
ECI	0.010**	0.018***	0.010***	0.021***	0.009**	0.022***
	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.006)
ECI_{t-1}		-0.009		-0.014**		-0.014**
		(0.006)		(0.006)		(0.007)
Ln(GDPPC)	-0.007	-0.005	-0.033	-0.032**	-0.032	-0.039*
	(0.015)	(0.014)	(0.025)	(0.015)	(0.021)	(0.021)
Ln(GDPPCSQ)	0.000	-0.000	0.001	0.001*	0.001	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Schooling)			0.003	0.008	0.007	0.011*
			(0.005)	(0.005)	(0.006)	(0.006)
Ln(Population)					-0.001	-0.001
					(0.001)	(0.001)
Constant	0.055	0.056	0.177	0.180***	0.190*	0.226**
	(0.062)	(0.061)	(0.109)	(0.062)	(0.107)	(0.101)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	633	633	582	582	582	582
Number of countries	113	113	98	98	98	98
AR(2) (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
AR(3) (p-value)	0.407	0.680	0.503	0.729	0.479	0.787
Hansen test (p-value)	0.597	0.510	0.391	0.718	0.775	0.643
No. of instruments	57	92	71	82	58	82

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1. List of countries

Low- and middle- income countries	High-income countries
Albania, Algeria, Argentina, Azerbaijan, Bangladesh, Belarus, Bolivia, Botswana, Brazil, Bulgaria, Cambodia, Cameroon, China, Colombia, Congo, Costa Rica, Côte d'Ivoire, Croatia, Democratic Republic of Congo, Dominican Republic, Ecuador, Egypt, El Salvador, Gabon, Georgia, Ghana, Guatemala, Guinea, Honduras, India, Indonesia, Iran, Jamaica, Jordan, Kazakhstan, Kenya, Lao PRD, Lebanon, Macedonia, Madagascar, Malaysia, Mauritania, Mexico, Moldova, Mongolia, Morocco, Namibia, Mozambique, Nicaragua, Nigeria, Pakistan, Panama, Paraguay, Peru, Philippines, Romania, Russia, Senegal, Serbia, South Africa, Sri Lanka, Sudan, Syrian Arab Republic, Tanzania, Thailand, Togo, Tunisia, Turkey, Turkmenistan, Ukraine, Vietnam, Yemen, Zambia, Zimbabwe.	Australia, Austria, Belgium, Canada, Chile, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong (China), Hungary, Ireland, Israel, Italy, Japan, Korea, Kuwait, Latvia, Lithuania, Netherlands, New Zealand, Norway, Poland, Portugal, Qatar, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Trinidad and Tobago, United Kingdom, United States, Uruguay.

Note: the classification of country groups is based on World Bank's definitions https://datahelpdesk.worldbank.org/knowledgebase/articles/906519

Table A2. Summary statistics of variables (OLS regressions)

Variable	Obs	Mean	Std. Dev.	Min	Max
GINI	109	0.380943	0.083856	0.230411	0.57253
ECI	109	0.077387	0.951225	-2.04546	2.350117
Ln(GDPPC)	109	8.570385	1.40703	5.540155	11.20966
Ln(GDPPCSQ)	109	75.41308	24.20833	30.69332	125.6566
Schooling	96	7.462445	2.578148	1.214192	12.51064
Secondary	96	2.499143	1.225113	0.132902	5.297678
Tertiary	96	0.352094	0.264049	0.008360	1.299796
Trade	109	78.76388	45.2895	21.35973	355.3419
FDI	109	3.675215	3.722337	0.123863	27.03412
Population	109	4.99E+07	1.54E+08	905957.1	1.21E+09
Rule of law	108	0.123102	1.025547	-1.91632	2.100273

Table A3. Summary statistics of variables (panel data regressions)

Variable	Obs	Mean	Std. Dev.	Min	Max
GINI	749	0.379715	0.085317	0.206	0.58384
ECI	749	0.151807	0.996357	-2.37329	2.544056
Population	749	5.56E+07	1.59E+08	660450.6	1.35E+09
Ln(GDPPC)	749	8.591261	1.442885	5.255596	11.40545
Ln(GDPPCSQ)	749	75.8889	24.71798	27.62129	130.0842
Schooling	683	7.283101	2.821076	0.727634	13.18264

Table A4. List of variables and data sources

Variables	Description	Data source	Period	
			OLS (average value)	System GMM (5-year average, with gaps)
GINI	Estimate of GINI index of inequality using household disposable income	Standardized World Income Inequality	1980-2014	1965-2014
ECI	Economic Complexity index	MIT's Observatory of Economic Complexity (atlas.media.mit.edu)	1980-2014	1965-2014
GDPPC	GDP per capita (constant 2010 US$)	World Bank's World Development Indicators database	1980-2014	1965-2014
GDPPCSQ	The square of GDP per capita (constant 2010 US$)	World Bank's World Development Indicators database	1980-2014	1965-2014
Schooling	Average years of schooling	Barro and Lee (2013)	1980-2010	1965-2010, with gaps

Secondary	Average years of secondary schooling	Barro and Lee (2013)	1980-2010	1965-2010, with gaps
Tertiary	Average years of tertiary schooling	Barro and Lee (2013)	1980-2010	1965-2010, with gaps
Trade	The sum of exports and imports as a share of GDP	World Bank's World Development Indicators database	1980-2014	1965-2014
FDI	The net inflows of foreign direct investment as a percentage of GDP	World Bank's World Development Indicators database	1980-2014	1965-2014
Population	The total population (in millions)	World Bank's World Development Indicators database	1980-2014	1965-2014
Rule of law	This index comprises of various indicators which capture the extent to which different agents of society have confidence in and abide by its rules	The quality of government basic dataset Dahlberg et al. (2016)	2009	N/A