The Relationship Between Growth and Economic Complexity: Evidence from Southeastern and Central Europe

Stojkoski, Viktor and Kocarev, Ljupco

Macedonian Academy of Sciences and Arts

2017

The Relationship Between Growth and Economic Complexity: Evidence from Southeastern and Central Europe

Viktor Stojkoski*[1] and Ljupco Kocarev[1,2]

[1]Research Center for Computer Science and Information Technologies, Macedonian Academy of Sciences and Arts, Skopje, Macedonia,

[2]Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, Skopje, Macedonia

March 22, 2017

Abstract

The index of economic complexity is created by analyzing the relations between countries and the products they export. Constructed in such way, it defines the basis for the theory of economic complexity, which reflects the knowledge embedded in the productive structure of an economy. Exactly this knowledge is at the core of the endogenous theory of economic growth. Until now, all econometric analyses for the relationship between economic complexity and growth were done by implementing methods in which each country is valued equally. However, the countries are heterogeneous – they exhibit individual characteristics that directly encourage the complexity, and are in tight relation with growth. Therefore, in this paper the analysis is faced towards one region – Southeastern and Central Europe, and, in the spirit of the endogenous theory, a model is created which adequately captures the long run, as well as the short run relationship between the two variables. The results show that the economic

*Corresponding author. E-mail: vstojkoski@manu.edu.mk

complexity is a statistically significant explanatory variable of growth on the long run, and thus, it creates enormous economic implications. Contrarily, on the short run the productive knowledge has no effect on the income changes in Southeastern and Central Europe. All of this implies that the economic complexity reveals a structure which promotes development of long run strategies in the countries for inventing products. These strategies serve for the purpose of accumulating new capabilities that will help in creating and maintaining long term prosperity and economic growth.

Keywords: Economic Complexity, Economic Growth, South Eastern and Central Europe, Cointegration, Error Correction Model

1 Introduction

Economic Complexity studies the productive structure embedded in the goods and services that an economy creates. This structure reveals the economy's productive capabilities, i.e. it's productive knowledge. The productive knowledge can be formally defined as *"the sum of the complex cooperation between the individuals, institutions and policies in a society"* Hidalgo and Hausmann, (2009). This definition implies that the the economic complexity serves as a significant promoter of the wealth of the nations. In fact, we can even conjecture that the disparities in the level of complexity create divergence in the rates of economic growth among the countries, Hidalgo and Hausmann, (2009).

The observation regarding the relationship between economic growth and economic complexity was implicitly conceived in Hidalgo, Klinger, et al., (2007). In this paper, the authors create a model for the so-called *product space*, and by analyzing the network of similarities between the exported products they reveal that products which are exported mainly by rich countries are located centrally in the network, while products exported by countries with lower income are in the periphery. Based on this work, Hidalgo and Hausmann, (2009) construct the building blocks for economic complexity. By interpreting international trade data as a bipartite network between countries and products, the authors create a measure for the complexity of an economy. Since this measure is able to explain the differences in income across countries and is a significant predictor of their future growth, Hidalgo and Hausmann conclude that the *countries converge to the level of income which can be supported by their complexity*. According to Hidalgo and Hausmann, this conclusion shows that development policies should focus on creating conditions that will stimulate the growth of economic complexity. Detailed explanation for the implications created by economic complexity is presented in Hausmann et al., (2014). In this work, a precise definition for the measure (now called Index of Economic Complexity - ECI) is given - it *quantifies the composition of the productive output of a country and reflects the structures that encourage knowledge*. As such, ECI can be used to empirically test the importance of knowledge accumulation and diversification of products for economic growth. On the one hand, it underpins ancient macroeconomic theories of economic growth, such as Adam Smith's idea for division of labor Smith, (1776); or their modern counterparts, for instance the theory of endogenous growth P. M. Romer, (1994). On the other hand, through the information contained in the empirical data, economic complexity questions the validity of the Ricardian theory of comparative advantage Ricardo, (1817) and Kremer's O-ring Model for economic development Kremer, (1993). Of course, economic complexity has implications which create an alternative view for the productive structure of an economy.

All these discoveries for the potential impact of economic complexity on growth contribute to the huge actualization of the topic. In fact, in recent years a multitude of papers have appeared which present the notion of economic complexity as an approximation of the knowledge and human capital factors, and study the effect of the productive structure in the spirit of endogenous growth theory. Nevertheless, most of them overlook the differences between countries. Instead, they focus on the basic meaning and consequences of improving or neglecting the productive structure. This is mostly due to the fact that all papers for quantifying the relationship between economic growth and complexity use econometric techniques in which the data are distributed in the form of panel. The biggest advantage of this structure is that it helps in dealing with the problem of small data samples. However, in possession of sufficient data (regardless of whether it is structured as time-series or cross-sectional), that are enough for producing statistical conclusions, the introduction of a new dimension (and creating a panel) opens a wide range information which make the conclusions more consistent, more efficient and of course, more general. The newly introduced generality represents a double-edged sword because, on the one hand, it helps to make the conclusions more concise and subject of a wider audience, but on the other hand, this way the differences between the different groups or periods are overlooked. This feature of panel structured data is an obvious problem in modeling the effect of economic complexity on the growth, since there is heterogeneity across countries Cristelli, Tacchella, and Pietronero, (2015). The heterogeneity appears because the countries occupy different positions in the product space - developed countries populate almost all of it, and as the development decreases the occupied area decrements exponentially. Hence, it can be argued that there is a different relationship between the economic complexity and the growth across economies. To tackle the problem of heterogeneity Hidalgo and Hausmann, (2009) applied a model of panel regression with fixed effects across cross sections. The term of cross sectional (periodic) fixed effects implies a decomposition of the random error into a sectional (periodic) random error and a residual that changes through time and space Brooks, (2014). By applying this technique Hidalgo and Hausmann control for the level of development across countries. However, cross-fixed effects are not a practical solution for heterogeneity because they neglect the potential differences in the countries' ability to create new products. Therefore, Hausmann et al., (2014) avoid this approach, and the authors resort to periodic fixed effects and adding a new variable that explains the heterogeneity of countries in the product space. This specification creates a robust model that consistently explains the differences in the long run economic growth of the countries. However their model is not without shortcomings. Tacchella et al., (2012) and Cristelli, Gabrielli, et al., (2013) criticized Hidalgo and Hausmanns' regression approach due to the inconsistent magnitude of the relationship through time - their main e conclusion is that simple regression analysis has poor explanatory and predictive power. Although, to some extent, the arguments presented in the aforementioned papers are justified (in respect to the inconsistent relationship through time), the conclusion that standard regression analysis is unable to explain the relationship between the two variables is absurd - regression is the main technique used by scholars of economic growth for validating hypotheses about the effect of certain variables.

The heterogeneity among the countries can be used as an explanation for the unstable relationship over time. Particularly, we can assume that there is a variety of external factors that affect the diffusion of a country in the product space. Therefore, we claim that when modeling the relationship between economic growth and complexity it is better to limit the analysis to one country or to a set of countries which exhibit similar characteristics. The techniques of cointegration and error correction are especially suited for such models because they produce consistent and efficient estimates for both, the long run and short run relationship between the phenomena.

In this paper, we relax the assumption of heterogeneity and thoroughly analyze the heterogeneous relationship between economic complexity and growth in Southeastern and Central Europe. Our focus is on these countries as research on the subject of the determinants of endogenous growth in this European region has been mostly neglected. This can be concluded from a brief literature review - almost all papers that study the endogenous growth in Southeastern and Central Europe belong to one of the two most common groups of papers: (i) papers relating growth with foreign direct investment (FDI) and (ii) papers relating growth with various financial indicators. Although all papers, in a way, relate economic growth with various economic variables that help explain its endogeneity, none of them explore the direct effect of the factors of technology and knowledge. To our knowledge, the only papers that differ from this scheme are Silaghi et al., (2014) and Hartman, (2013). Closest to our approach is the one presented in first paper, where the authors explore the short run and long run relationship between economic growth and investment in research and development in 10 countries of Central and Eastern Europe, and reveal that the variables are positively correlated even when controlling for other variables that could possibly affect growth. Nevertheless, they employ inconsistent methodology and fail to include other countries with similar characteristics. This paper furthermore distinguishes from the others as it applies consistent methodology of panel cointegration and error correction modeling to adequately quantify the heterogeneous long run and short run relationship between technology approximated through the productive knowledge and the economic growth in Southeastern and Central Europe.

The paper is structured as follows. We start by describing our econometric model, structured to identify the long run and short run effect of economic complexity on economic growth. Then, we continue with presenting the data, which indirectly uncovers the implication of economic complexity. The Results Section is divided into two parts - The Long Run Relationship and the Short Run Relationship. The first part presents the long run effect of the productive structure, it quantifies the magnitude and provides a discussion on the possible directions for the long run productive policies in South Eastern and Central Europe. In the second part, we quantify the short run relationship, and answer two questions which the literature on economic complexity has not answered yet, "How does economic complexity affect the short run growth?", and "How quickly do the countries converge to the long run equilibrium, when there is a shock in the relationship?". In the last section we conclude our findings.

2 Econometric Model

Our econometric model is very similar to the standard production function used in most papers, see for e.g. Herzer and Vollmer, (2012) and Mankiw, D. Romer, and Weil, (1990). In particular, we specify a Cobb-Douglas-esque form, Douglas, (1976), on the function describing the income per capita. That, is the dependent variable is given by the logarithm of GDP per capita measured in US dollars at power purchasing parity. Its changes are the growth of an economy. Our goal is to follow the endogenous theory and examine the effect of the Index of Economic Complexity. The index represents a standardized variable (a value higher than 0 suggests that the economy is more complex than the average economy), calculated as the eigenvector (influence) centrality of an economy in the bipartite network connecting countries to the products they export, and should directly as well as indirectly promote the growth.

However, ECI, by itself is not enough to explain the changes in income in Southeastern and Central Europe, and therefore we include two additional variables that represent a rough approximation for other factors that can have an effect on economic growth and complexity through steady evolution over time. The first variable is the gross capital formation (GCF) as a percentage of GDP, which is a simplification of the investments in a country. The short run effect of the investments on growth is unclear, i.e. it can be positive, neutral or negative because it depends on the industry in which is invested. But it is clear that on the long run the gross capital formation indirectly promotes the technology and increases the wealth of an economy Levine and Renelt, (1992). The second variable is the export of goods and services as a percentage of GDP. With this variable we control for the openness of a country because even though ECI is calculated by using data from international trade, it does not exploit information on the magnitude of trade and at the same time, the willingness of a country to trade. Of course, both variables are expressed in their respective logarithmic transformations. A specific feature of our econometric model is that, unlike all the other papers, in which was assumed that economic complexity only has effect on the long run growth of a country, it assumes that there is both a long term and short term relationship between the economic complexity and the wealth of a nation.

We specify the long run relationship as:

$$\log\left(gdp_{it}\right) = \alpha_i + \beta_1 eci_{it} + \beta_2 \log\left(gcf_{it}\right) + \beta_3 \log\left(exp_{it}\right) + u_{it}, \tag{1}$$

where $i = 1, 2, \ldots, C$ and $t = 1, 2, \ldots, T$ are, respectively notations for the time and countries; eci_{it} is the economic complexity, while $\log\left(gcf_{it}\right)$ and $\log\left(exp_{it}\right)$ represent the logarithmic values of the gross capital formation as percentage of GDP and exports of goods and services as percentage of GDP. The dependent variable, the logarithm of real GDP at purchasing power parity per capita, is $log\left(gdp_{it}\right)$. The β coefficients quantify the long run effects of the independent variables over the dependent, whereas α_i is a specific cross sectional fixed effects intercept that helps to control for all omitted factors that are stable over time.

With the model presented in equation (1) we assume that there is a long run relationship between economic complexity, investment, openness and income. This assumption is valid only if the individual time series of all three variables are not stationary at their levels, integrated of the same order and form a cointegrated system Brooks, (2014) and Herzer, Strulik, and Vollmer, (2012). By definition, two or more variables are cointegrated if there is a linear combination of them that has a stationary random error, indicating that long-term cointegration relationship between variables is linear. Also, the stationarity of the error term means that there is no omitted variable bias; "Any omitted non-stationary variable that is part of the cointegrated system should enter the random error u_{it}, thereby producing nonstationary residuals and thus leading to a failure to detect cointegration" Herzer, Strulik, and Vollmer, (2012) and Herzer and Vollmer, (2012).

We use the cointegration relationship between the variables to model the short run relationship, and create an error correction model (ECM). Strictly speaking, the ECM represents a "bridge" between the short and long run as it allows direct quantification of the short run relationship and evaluation of the speed of convergence towards the long run equilibrium.

We specify the short run model of economic growth in Southeastern and Central Europe as follows:

$$\begin{aligned}\Delta \log (gdp_{it}) = a_i + b_1 \Delta eci_{it} + b_2 \Delta \log (gcf_{it}) + b_3 \Delta \log (exp_{it}) \\ + b_4 \Delta \log (gdp_{it-1}) - \theta u_{it-1} + e_{it}.\end{aligned} \tag{2}$$

This equation is a modification of equation (1) where now the b coefficients measure the short run effects of the independent variables and a_i are the short-term omitted factors that are stable over time. As was noted above, the error correction model is characterized with the introduction of the lag of the random error, u_{it-1}, of the long run model as an independent variable. Its parameter θ merges the long run and short run, and directly quantifies the speed of convergence towards the long-term equilibrium. In the case of economic growth, Arnold, Bassanini, and Scarpetta, (2011) theoretically derived the form of ECM for the Solow-Swan and the endogenous models (in which the authors assumed that there are constant returns to scale), and showed that it quantifies the rate of convergence to the steady state growth rate. Additionally, in equation (2) we also add the lagged first differences of the dependent variable (the logarithm of GDP per capita at purchasing power parity) in order to address the short run autocorrelation.

3 Data

To empirically test the validity of the relationship defined in (2.1) and (2.2) we use data on 16 countries from Southeastern and Central Europe: Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Estonia, Czech Republic, Hungary, Latvia, Macedonia, Moldova, Poland, Romania, Slovakia, Slovenia and Ukraine. We focus on the period from 1995, when most of the countries included in the sample started reporting the data, until 2013 which is the last period when data was reported. Thus we create a balanced panel with 304 observations ($C = 16$, $T = 19$). The data for the Economic Complexity Index are taken from the Observatory of Economic complexity (`http://atlas.media.mit.edu/en/`), while the data on GDP per capita in PPP terms, gross capital formation as a percent of GDP and export of goods and services as a percent of GDP are taken from World Bank's World Development Indicators database (`http://databank.worldbank.org/`).

Table 1 provides the summary statistics. On the one hand, the positive value of the average ECI implies that the complexity in Southeastern and Central Europe is higher than the world average (which is 0). Among the countries, highest average Economic Complexity Index has Czech Republic, followed by Slovenia and Slovakia; while Macedonia and Albania have the lowest ECI. On the other hand, according to average income per capita, the International Monetary Fond classifies the economies of Southeastern and Central Europe as developing or emerging countries. A characteristic of these two types of countries is that they experience high growth rates, but also possess higher volatility in it. Average income (GDP) per capita is the highest in Slovenia, followed by the Czech Republic and Hungary, and it is lowest in Moldova and Bosnia and Herzegovina. The formation of gross capital and the export of goods and services in terms of GDP display similar movements as those of ECI and GDP per capita - they are higher in countries with high income and ECI, and lower in countries with low average income and ECI. All in all, according to the summary statistics it seems that there is a positive relationship between the income, productive structure, investment and openness for countries in Southeastern and Central Europe.

4 Results

4.1 The Long Run Relationship

In econometrics the term of long run is a synonym to the technique of cointegration Engle and Granger, (1987), which was defined in Section 2. Although, initially it was made only for time series analysis, Pedroni, (2004, 1999) extended the technique to panel data, as the data is structured in this paper. The scheme for identifying and estimating a cointegrated system follows three steps: i) testing the stationarity and order integration; ii) testing the potential cointegration; and iii) estimating the cointegration (long run) relationship. In the following we describe the steps and present the results for the model specified in (1).

4.1.1 Stationarity Tests

The first step towards evaluating a cointegrated system is testing the properties of all panel (time series) data. Specifically, the variables presented in equation (1) should be non-stationary at their levels and integrated of the same order (standardly denoted as $I(d)$ where d is the order of integration).

The properties for the stationarity of a time series are determined by unit root tests. If the data has a unit root, then it is not stationary. In recent years, a multitude of unit root tests have been proposed that are especially suited for panel data. They can be divided into two groups: tests which assume that there is a common unit root among the cross sections; and tests that relax this assumption. In this paper we use the Breitung, (1999) and Maddala and Wu, (1999) tests. which according to Hlouskova and Wagner, (2006) and Maddala and Wu, (1999) outperform all other tests, when the model includes cross sectional fixed effects, as does ours.

The Breitung test belongs to the first group of unit root tests. Its statistic is calculated similarly as in the augmented Dickey-Fuller (ADF) regression:

$$\Delta y_{it} = \omega_i y_{it-1} + \sum_{j=1}^{L_i} \delta_{ij} \Delta y_{it-j} + \psi_i z_{it} + \upsilon_{it}, \tag{3}$$

where Δy_{it} denotes the first difference of the dependent variable, z_{it} is a vector of deterministic variables which help in explaining fixed effects or individual trends, and ψ_i is the corresponding vector of coefficients. Since the Breitung test belongs to the first group, it constraints the parameter ω_i, which can be written as $\rho_i - 1$ where rho_i is the autocorrelation across the i-th cross section, to be equal for each cross section, i.e. $\omega_i = \omega$. Under the null hypothesis it assumes that there is a unit root, whereas the alternative is that the variables are stationary. The statistic asymptotically follows a standard normal distribution.

The second test, developed by Maddala and Wu (MW) belongs to the second group of tests, i.e. it assumes that the ω_i coefficient in equation 3 is heterogeneous among the cross sectional units. The statistic for this test is calculated as the sum of the log of the p-values for each individual cross sectional augmented Dickey-Fuller test:

$$MW = -2\sum_{i=1}^{C} \log(p_i). \tag{4}$$

The MW test has the same hypotheses setup. However, differently from Braitung, it asymptotically follows the chi-squared distribution with $2C$ degrees of freedom.

The results for the two tests can be seen in Table 2. Both tests suggest that all variables have unit roots in their levels, and are integrated of order one ($I(1)$). We also conducted several other tests (which are not suited for our data structure), such as the tests of Hadri, (2000), Im, Pesaran, and Shin, (2003), and Levin, Lin, and Chu, (2002). Although slightly different, their statistics (available at request), do not notably change the results presented in Table 2. All these conclusions allow us to continue to the next step in evaluating the long run relationship.

4.1.2 Cointegration Tests

The second step in evaluating a cointegrated system is testing for cointegration between the variables. To test this feature we use two types of tests; (i) tests based on the Johansen, (1988) methodology; and, (ii) Engle-Granger alike tests.

The first group of tests examines the number of potential cointegration vectors between the variables[1]. In this paper we use the two standard tests developed by Johansen, (1988) - the Trace and Maximum Eigenvalue tests. Maddala and Wu, (1999) accommodated these tests to panel data using the methods suggested by Fisher et al., (1950). Basically, the statistic of both tests is related to the following error correction model:

$$\Delta y_{it} = \Pi_i y_{it-1} + \sum_{j=1}^{L} \Gamma_{ij} \Delta y_{it-j} + \psi_i z_{it} + \upsilon_{it}, \tag{5}$$

where y_{it} is a $k \times 1$ vector of endogenous variables[2]; k is the number of variables and Π_i a $k \times k$ matrix of long run relationships between the variables. If $1 < rank(\Pi_i) < k$,

[1]Recall that, if there are 4 variables, then there are 3 potential cointegration vectors.

[2]In our case $y_{it} =[\log(gdp_{it}), eci_{it}, \log(gcf_{it}), \log(exp_{it})])$.

the matrix can be decomposed as A_iB_i', where B_i is an $r \times k$ matrix whose rows are the cointegration vectors while A_i is a $k \times r$ matrix which approximates the contribution of each cointegration vector in the ECM. The statistics for the Fisher accommodated Johansen tests is calculated in a similar manner as in equation (4), with the exception that now it is summed over the p-values of the individual trace or maximum eigenvalue value statistic for each cross sectional unit. The difference between these two tests is in the formulation of the hypotheses. The Trace test is a one-sided test with an alternative hypothesis of more than r cointegration vectors, while with the Maximum Eigenvalue test a separate test is carried out on each eigenvalue of Π_i with the alternative hypothesis of exactly $r+1$ cointegration vectors. Both tests asymptotically follow the chi-squared distribution with $2C$ degrees of freedom. The advantage of these tests is that they do not require a specification of the long run relationship. Instead, they examine how many combinations of the variables produce stationary error terms. Therefore, even if we conclude that cointegration exists among the variables, we still have the problem of specifying its form.

In order to solve this problem, we further calculate four of the seven panel cointegration statistics proposed by Pedroni Pedroni, (2004, 1999): The panel ADF statistics, panel Phillips-Perron (PP) statistics, and the group ADF and PP statistics. These tests represent the panel modification of the basic method for assessing cointegration introduced by Engle and Granger, (1987), revolve around testing the stationarity of the random error u_{it} from equation 1, and asymptotically are normally distributed. Wagner and Hlouskova, (2009) show that the aforementioned four tests have superior performance than any other Engle-Granger based tests when the panel data has short time dimension.

The statistics for all tests are presented in Table 3. Clearly, the inability of both the Trace and Maximum Eigenvalue tests to reject the null hypothesis of less than two cointegration vectors and the rejection of the null hypothesis of zero cointegration vectors, indicates that there is a long run relationship between the income, economic complexity, investments and the openness of Southeastern and Central European countries. However, the tests have different conclusions about whether there are two cointegration vectors; the Trace test requires a significance level of 5% to reject it, whereas the Maximum Eigenvalue requires 10%. This suggests that potentially there is more than one long run relationship - due to the possible endogenous relationship between the variables. Nevertheless, because of the different levels of significance required by the tests we conclude that there is only one cointegration relationship[3]. All tests based on the Engle-Granger methodology reject the null hypothesis that the equation 1 is not a cointegration relationship between the four variables. This allows to conclude our model is a long run relationship between $\log(gdp_{it})$, eci_{it}, $\log(gcf_{it})$ and $\log(exp_{it})$.

4.1.3 Estimating the Long Run Relationship

Since the unit root and cointegration tests suggest that the variables are non-stationary at their levels, integrated of order 1 and cointegrated, as assumed in equation (1), we proceed with estimating the long run relationship. To quantify the relationship we implement the Panel Dynamic Ordinary Least Squares (DOLS) estimator, proposed by Kao and Chiang, (1999). We chose this estimator as it gives asymptotically unbiased and efficient estimates of the long run relationship, even when some of the regressors are endogenous Herzer, Strulik, and Vollmer, 2012. Moreover, Wagner and Hlouskova, (2009) show that the DOLS estimator has better performance than other available estimators, when the panel data has short time dimension (such as ours). The DOLS model given in equation 6 is a modification of equation 1:

$$\begin{aligned}\log(gdp_{it}) = {} & \alpha_i + \beta_1 eci_{it} + \beta_2 \log(gcf_{it}) + \beta_3 \log(exp_{it}) \\ & + \sum_{j=-L}^{Q} \phi_{1ij} \Delta eci_{it+j} + \sum_{j=-L}^{Q} \phi_{1ij} \Delta \log(gcf_{it+j}) + \sum_{j=-L}^{Q} \phi_{1ij} \Delta \log(exp_{it}) + \epsilon_{it}. \qquad (6)\end{aligned}$$

In the equation $\phi_{1ij}, \phi_{2ij}, \phi_{3ij}$ represent coefficients of the leading (Q) and lagging (L) differences of the explanatory variables that help generate unbiased estimates of $\beta_1, \beta_2, beta_3$ by eliminating the potential asymptotic endogeneity and serial correlation.

The estimates of the panel DOLS procedure are displayed in Table 4. All explanatory variables, eci_{it}, $\log(gcf_{it})$, $\log(exp_{it})$, are significant predictors of income at any level. More importantly, a positive change in any variable increases the growth in Southeastern and

[3]This is the only conclusion, if for example we take 10% level of significance and use the Maximum Eigenvalue test.

Central European countries. The coefficient of eci_{it} reports the long run semi-elasticity of income per capita in regard to economic complexity. It's magnitude, 0.45, means that a change in ECI of one standard deviation, on average increases the GDP per capita by 45%. On the other hand, the coefficients of gross capital formation as a percent of GDP and the exports of goods and services as a percent of GDP show the long run income elasticity in Southeastern and Central Europe with respect these variables. The size of the $\log(gcf_{it})$ coefficient implies that an increase of 1% in the gross capital formation to GDP ratio, increases income by 0.56%. Similarly, if the ratio of exports to GDP increases by 1%, the income in Southeastern and Central Europe on average increases by 0.71%.

To adequately compare the magnitude of the effect of eci_{it} with those of $\log(gcf_{it})$ and $\log(exp_{it})$, we standardize all coefficients by multiplying them with the ratio of the standard deviation of the respective independent variable and the standard deviation of the dependent variable. These results are reported in column 3 of Table 4. The standardized coefficient of ECI indicates that in the long run, an increase of one standard deviation in eci_{it} is associated with an increase in income per capita of 43.4% of its standard deviation, while an increment of $\log(gcf_{it})$ and $\log(exp_{it})$ of one standard deviation standardization contributes to the increase in income by, respectively, 30.1% and 43.9% of the standard deviation of $\log(gdp_{it})$.

The magnitude of the coefficient of economic complexity is notably bigger than the effect of gross capital formation. Also, interestingly, it is only a little lower than half of the combined effect of all other explanatory variables. This allows us to conclude that a increase in economic complexity has a huge economic impact in South East and Central Europe.

We tested the robustness of our results by: (i) estimating the long run model with alternative estimation methods (the group DOLS estimator, as well as the panel and group Fully Modified Ordinary Least Squares (FMOLS) estimator), and by adding the log of monetary stock M2 with respect to GDP, as a measure of financial development, to the model. The results (available at request) did not change significantly, thus allowing us to conclude that the econometric model presented in (1) is statistically and econometrically justified.

4.2 The Short Run Relationship

The final step of our analysis is using the results from the long run relationship to construct an extensive error correction model, equation (2), that captures the short run relationship between income controlled for investments and openness, and economic complexity. For the estimation of this relationship use the System Generalized Method of Moments (GMM) estimator, proposed by Blundell and Bond, (2000). We opt for this estimator, as it is structured to control for the potential unobserved heterogeneity and endogeneity across the variables, and cross sectional units (countries). To correct these biases, under the System GMM we assume that the other lags of the dependent variables, as well as their differences, are not correlated with the random error, and use them as instruments in estimating equation (2). Hence, it can be said that this estimator is a special type of Instrumental Variables estimation.

The estimates of the short run relationship are presented in Table 5. We can see that, in fact, changes in economic complexity have no effect on short run changes in the income in Southeastern and Central Europe. On the other hand, $\Delta \log(gcf_{it})$, $\Delta \log(exp_{it})$, $\Delta \log(gdp_{it-1})$, and error correction term, θ, are significant explanatory variables of short run income changes. From this, it follows that there is a negligible short run relationship between the productive structure and economic growth. This conclusion is not unexpected given the definition of ECI - it requires time and effort for an increase in complexity to be reflected in the wealth of a nation. In the last two column we show the statistics for the two most common tests used for assessing the validity of instruments in a System GMM. The first test, the Hansen J-test, Hansen, (1982), is an extension of the classic Sargan test, Sargan, (1958), for exogeneity of the instruments. Under the null hypothesis, this test assumes that the instruments are exogenous. In the case of the model specified in equation (2) the p-value is 0.17 which means that the used instruments are indeed reasonable. The second test, named after its creators Arellano and Bond, (1991), tests for possible second order autocorrelation. If this phenomena exists, then the random error is correlated with the instruments, i.e. they are invalid. The p-value of the Arellano-Bond test is 0.62, thus not rejecting the null of no autocorrelation.

The magnitude of the coefficients is also slightly different from the ones estimated for the long run relationship. In the dynamic model of the short run relationship the autorefressive coefficient, $\Delta \log (gdp_{it-1})$, is almost twice the size of the effect of changes in gross capital formation, and four times the effect of changes in exports. As previously noted, the θ parameter measures the speed of convergence towards the long run equilibrium. It ranges between 0 and 1, where a higher value indicates that deviations from the cointegration relationship have a greater contribution to the short run dynamics of the model. The negative is due to the fact that positive deviations from the long run relationship should have a negative effect on the short run in order to restore equilibrium. In our model of Southeastern and Central Europe the magnitude of the error correction term is around 0.05 which means that deviations from the long run equilibrium have relatively small effect on the short run income changes. Hence, we can conclude that the productive structure and economic growth exhibit an unstable relationship, since if there is a shock in the long run relationship, then the countries of Southeastern and Central Europe will slowly converge to the equlibrium [4]. To compare the size of error correction term with the other variables in the short run model, we do the same standardization as Section 4.1. The standardized coefficients are given in column 2 of Table 5. They differ in part from the raw, since in this case, gross capital formation has the highest marginal effect over income per capita. Nevertheless, the error correction term still has the lowest marginal effect (excluding ECI), thus implying, that indeed the countries have a very slow convergence rate.

5 Discussion

The results presented in this paper offer a new dimension for the countries in Southeastern and Central Europe to promote their economic development. Particularly, the fact that economic complexity has an enormous impact on the long run income changes, which is even bigger than the role of investments and openness, indicates that countries should focus on long run strategies for inventing (or producing) new goods and services. However, this does not mean that they should totally change their current industrial (production) policy.

[4]Specifically, they will be converging with a rate of around 5% per period.

Instead, in the spirit of the *product space*, the countries should base the strategies on the productive capabilities currently available in their economies. They should seek to produce new goods and services that are close to the complexity of the products present in their export basket, and, are stationed nearby in the space of products. When constructing these long run strategies, the countries should be aware of the neutral short run effect of economic complexity, which means that it takes a while until their effect can be felt. Also, they should be aware that it takes a lot of time for the system to return to its equilibrium when there is a shock in the long run relationship. Thus, it can be concluded that there is a slight trade-off when choosing between the long run and short run. Nevertheless, as our results imply, the best way for establishing long run prosperity and development is by leveraging the economic complexity.

In the future, it would be interesting to produce a more detailed analysis on the individual productive structure for each country in Southeastern and Central Europe. In that way we can offer more precise strategies and policies by specifically detecting the necessary and possible products which can be introduced in the economies. A perfect model would present an inductive approach, constructed of several sub-models that offer strategies to separate production sectors, which can be merged into a general cohesive framework.

References

[AB91] Manuel Arellano and Stephen Bond. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". In: *The review of economic studies* 58.2 (1991), pp. 277–297.

[ABS11] Jens Arnold, Andrea Bassanini, and Stefano Scarpetta. "Solow or Lucas? Testing speed of convergence on a panel of OECD countries". In: *Research in Economics* 65.2 (2011), pp. 110–123.

[BB00] Richard Blundell and Stephen Bond. "GMM estimation with persistent panel data: an application to production functions". In: *Econometric reviews* 19.3 (2000), pp. 321–340.

[Bre99] Jörg Breitung. "The local power of some unit root tests for panel data". In: (1999).

[Bro14] Chris Brooks. *Introductory econometrics for finance*. Cambridge university press, 2014.

[Cri+13] Matthieu Cristelli, Andrea Gabrielli, et al. "Measuring the intangibles: A metrics for the economic complexity of countries and products". In: *PloS one* 8.8 (2013), e70726.

[CTP15] Matthieu Cristelli, Andrea Tacchella, and Luciano Pietronero. "The heterogeneous dynamics of economic complexity". In: *PloS one* 10.2 (2015), e0117174.

[Dou76] Paul H Douglas. "The Cobb-Douglas production function once again: its history, its testing, and some new empirical values". In: *The Journal of Political Economy* (1976), pp. 903–915.

[EG87] Robert F Engle and Clive WJ Granger. "Co-integration and error correction: representation, estimation, and testing". In: *Econometrica: journal of the Econometric Society* (1987), pp. 251–276.

[Fis+50] Ronald Aylmer Fisher et al. "Statistical methods for research workers. Biological monographs and manuals. No. V." In: *Statistical methods for research workers. Biological mono-graphs and manuals. No. V.* 11th ed. (1950).

[Had00] Kaddour Hadri. "Testing for stationarity in heterogeneous panel data". In: *The Econometrics Journal* 3.2 (2000), pp. 148–161.

[Han82] Lars Peter Hansen. "Large sample properties of generalized method of moments estimators". In: *Econometrica: Journal of the Econometric Society* (1982), pp. 1029–1054.

[Har13] D Hartman. "A global comparative perspective on innovation and economic diversification in Turkey". In: *İstanbul: IPC* 1 (2013).

[Hau+14] Ricardo Hausmann et al. *The atlas of economic complexity: Mapping paths to prosperity*. Mit Press, 2014.

[HH09] César A Hidalgo and Ricardo Hausmann. "The building blocks of economic complexity". In: *proceedings of the national academy of sciences* 106.26 (2009), pp. 10570–10575.

[Hid+07] César A Hidalgo, Bailey Klinger, et al. "The product space conditions the development of nations". In: *Science* 317.5837 (2007), pp. 482–487.

[HSV12] Dierk Herzer, Holger Strulik, and Sebastian Vollmer. "The long-run determinants of fertility: one century of demographic change 1900–1999". In: *Journal of Economic Growth* 17.4 (2012), pp. 357–385.

[HV12] Dierk Herzer and Sebastian Vollmer. "Inequality and growth: evidence from panel cointegration". In: *The Journal of Economic Inequality* 10.4 (2012), pp. 489–503.

[HW06] Jaroslava Hlouskova and Martin Wagner. "The performance of panel unit root and stationarity tests: results from a large scale simulation study". In: *Econometric Reviews* 25.1 (2006), pp. 85–116.

[IPS03] Kyung So Im, M Hashem Pesaran, and Yongcheol Shin. "Testing for unit roots in heterogeneous panels". In: *Journal of econometrics* 115.1 (2003), pp. 53–74.

[Joh88] Søren Johansen. "Statistical analysis of cointegration vectors". In: *Journal of economic dynamics and control* 12.2 (1988), pp. 231–254.

[KC99] Chihwa Kao and Min-Hsien Chiang. "On the estimation and inference of a cointegrated regression in panel data". In: *Available at SSRN 1807931* (1999).

[Kre93] Michael Kremer. "The O-ring theory of economic development". In: *The Quarterly Journal of Economics* (1993), pp. 551–575.

[LLC02] Andrew Levin, Chien-Fu Lin, and Chia-Shang James Chu. "Unit root tests in panel data: asymptotic and finite-sample properties". In: *Journal of econometrics* 108.1 (2002), pp. 1–24.

[LR92] Ross Levine and David Renelt. "A sensitivity analysis of cross-country growth regressions". In: *The American economic review* (1992), pp. 942–963.

[MRW90] N Gregory Mankiw, David Romer, and David N Weil. *A contribution to the empirics of economic growth*. Tech. rep. National Bureau of Economic Research, 1990.

[MW99] Gangadharrao S Maddala and Shaowen Wu. "A comparative study of unit root tests with panel data and a new simple test". In: *Oxford Bulletin of Economics and statistics* 61.S1 (1999), pp. 631–652.

[Ped04] Peter Pedroni. "Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis". In: *Econometric theory* 20.03 (2004), pp. 597–625.

[Ped99] Peter Pedroni. "Critical values for cointegration tests in heterogeneous panels with multiple regressors". In: *Oxford Bulletin of Economics and statistics* 61.s 1 (1999), pp. 653–670.

[Ric17] David Ricardo. *Principles of political economy and taxation*. 1817.

[Rom94] Paul M Romer. "The origins of endogenous growth". In: *The journal of economic perspectives* 8.1 (1994), pp. 3–22.

[Sar58] John D Sargan. "The estimation of economic relationships using instrumental variables". In: *Econometrica: Journal of the Econometric Society* (1958), pp. 393–415.

[Sil+14] Monica Ioana Pop Silaghi et al. "Do business and public sector research and development expenditures contribute to economic growth in Central and Eastern European Countries? A dynamic panel estimation". In: *Economic Modelling* 36 (2014), pp. 108–119.

[Smi76] Adam Smith. *An Inquiry Into the Nature & Causes of the Wealth of Nations: Vol 1*. 1776.

[Tac+12] Andrea Tacchella et al. "A new metrics for countries' fitness and products' complexity". In: *Scientific reports* 2 (2012).

[WH09] Martin Wagner and Jaroslava Hlouskova. “The performance of panel cointegration methods: results from a large scale simulation study”. In: *Econometric Reviews* 29.2 (2009), pp. 182–223.

Table 1: Summary Statistics

Country	mean gdp_{it}	mean eci_{it}	mean gcf_{it}	mean exp_{it}
Albania	6813.13	−0.33	25.35	22.92
Belarus	10704.67	0.98	30.19	61.84
BiH	7146.34	0.41	24.26	28.40
Bulgaria	12024.12	0.45	22.03	48.46
Croatia	17888.11	0.76	23.85	36.55
Czech Republic	24306.76	1.67	29.89	55.79
Estonia	19486.90	0.67	30.36	63.73
Hungary	20012.43	1.25	24.38	66.44
Latvia	15166.60	0.48	26.66	42.73
Macedonia	9570.39	−0.19	22.65	36.71
Moldova	3178.02	−0.05	26.20	47.85
Poland	17079.10	0.99	21.50	33.34
Romania	13904.20	0.62	23.87	32.66
Slovakia	19348.73	1.41	28.09	67.87
Slovenia	24990.04	1.53	26.14	57.30
Ukraine	6568.95	0.57	21.81	49.91
Mean	14261.78	0.70	25.45	47.03
Std. Dev.	7204.24	0.60	5.78	16.67
Max	30822.97	1.80	42.00	92.95
Min	1793.17	−0.65	0.30	10.47

Table 2: Unit Root Tests

	Breitung		MW	
Variable	level	difference	level	difference
$\log(gdp_{it})$	2.811	-2.070^{**}	18.189	71.356^{***}
eci_{it}	-0.647	-4.883^{***}	34.112	137.740^{***}
$\log(gcf_{it})$	-0.828	-1.636^{*}	36.975	213.253^{***}
$\log(exp_{it})$	1.327	-5.351^{***}	32.034	157.40^{***}

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3: Cointegration Tests

Pedroni (Engle Granger) Tests	Panel	Group		
PP-statistic	-3.409^{***}	-2.878^{***}		
ADF-statistic	-3.261^{***}	-2.770^{***}		
Fischer (Johansen) Tests	$r \leq 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$
Trace statistic	152.20^{***}	51.05^{**}	24.45	13.10
Max. eigenvalue statistic	140.20^{***}	44.19^{*}	26.17	13.10

Notes: For Pedronis ADF and the Johansen tests the optimal number of lags was selected with the Bayesian (Schwarz) Information Criteria. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4: DOLS Long Run Estimates

Dependent variable $\log(gdp_{it})$

Variable	Raw	Standardized
eci_{it}	0.450^{***}	0.434
	(0.128)	
$\log(gcf_{it})$	0.556^{***}	0.301
	(0.089)	
$\log(exp_{it})$	0.714^{***}	0.439
	(0.088)	

Notes: The optimal number of lags and leads was selected with the Bayesian (Schwarz) Information Criteria. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 5: GMM Short Run Estimates
Dependent variable $\Delta \log (gdp_{it})$

Variable	Raw	Standardized
Δeci_{it}	-0.002	-0.007
	(0.017)	
$\Delta \log (gcf_{it})$	0.195^{***}	0.601
	(0.036)	
$\Delta \log (exp_{it})$	0.085^{***}	0.193
	(0.016)	
$\Delta \log (gdp_{it-1})$	0.365^{***}	0.404
	(0.059)	
u_{it-1}	-0.048^{**}	-0.113
	(0.019)	
Hansen J		
(p-value)	0.170	
Arellano- Bond		
(p-value)	0.620	

Notes: The lags of $\Delta \log (gdp_{it})$, their differences, $\Delta^2 \log (gdp_{it})$, and u_{it-1} were used as instruments. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$