The Impact of Corruption on the Effectiveness of Cohesion Policy in New EU Member States

Master's Thesis by

Oleg Izgorodin

and

He Wen

Supervised by: Niels Blomgren-Hansen, Økonomisk Institut

Number of pages: 105

Number of characters: 188,374

July 2015

Copenhagen Business School

MSc in Advanced Economics and Finance

EXECUTIVE SUMMARY

This thesis analyses how corruption affects the impact of European Union's Cohesion Policy payments on economic growth in 12 new EU member states. It also reviews theories of economic growth and the role of institutions in economic development, describes the corruption phenomenon and its theoretical effect on economic growth, provides main facts about the Cohesion Policy and models used for estimating its effect on economic development and provides recommendations for improving the regulations of Cohesion Policy based on the results of empirical research.

Empirical research is performed by running regressions on cross-sectional and panel datasets with different estimation techniques: random effects, fixed effects and Generalized Method of Moments (GMM). The main panel dataset covers 12 countries over the period 2007-2013. The countries are new European Union members which joined in 2004 (Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia) and 2007 (Bulgaria and Romania). Results of the regression analysis show that the effect of Cohesion Policy payments on GDP growth is negative in countries with high corruption levels and positive in countries with low corruption levels. We explain this finding by the fact that in high-corruption countries, Cohesion Policy funds are misappropriated and used for less efficient projects so that human capital is not used in the most welfare-improving way. Therefore, we recommend the European authorities to modify the current Cohesion Policy regulations in two main ways. Firstly, European authorities should have a more important role in monitoring and controlling the implementation of projects. This is currently done mostly on a local level and thus introduces the potential for corrupt activity. Secondly, Cohesion Policy payments should be allocated towards fighting corruption and improving local institutions in the first place: this would greatly increase the positive effect of Cohesion Policy projects on GDP growth.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Research question

Economic researchers have been concerned with explaining cross-country income differences for a long time. Apart from the usual explanations of different levels of physical and human capital accumulation, there also exists an institutional explanation. Institutional economics is concerned with how the level of institutions, such as the quality of democracy, level of bureaucracy, regulations, etc., affects economic performance of countries. Probably the most important contribution to institutional economics comes from Acemoglu, Johnson and Robinson (2001). The authors estimate the effect of institutions by analysing colonisation policies: their main hypothesis is that these policies affected the quality of institutions that the colonies received and, therefore, their subsequent economic development. Acemoglu et al. (2001) results confirm their hypothesis. Other researchers focus on the particular dimensions of institutions. Mauro (1995) was among the first to provide an in-depth analysis of the effects of corruption on economic development. The author found that lower levels of corruption are associated with higher economic growth levels. However, the effect of corruption on economic growth is not completely straightforward and other researchers have found evidence for both no effect and a positive effect of corruption in some cases.

In this thesis, we are not concerned with analysing the general institutional framework. Instead, we focus on one particular institutional quality – corruption – in a specific setting. This allows us to analyse the issue in more detail. Specifically, our research question is as follows: how does corruption affect the impact of European Union’s (EU) Cohesion Policy payments on economic growth in new EU member states?

EU Cohesion Policy, also called Regional Policy, is aimed at reducing economic differences among EU member states. It provides funding for growth-promoting projects in less developed EU countries. Ever since its inception, the policy has been attracting a large degree of speculation as the funds transferred to these countries are very significant, whereas the effectiveness of the policy is ambiguous. Since 10 new countries joined the EU in 2004 (Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia) and 2 more (Bulgaria and Romania) in 2007, the amounts transferred increased even more. In the 2007-2013 period, EU budget had 348 billion Euros allocated to Cohesion Policy, amounting to 36% of the total budget. Among the 12 new members, Lithuania received the most funds as a share of its GDP: the average annual amount received by Lithuania in 2007-2013 was around 3% of GDP, while the total amount received in the period equalled 18% of its 2013 GDP.

All of the above makes this a relevant topic to analyse in both an academic and a practical perspective. From the academic perspective, we narrow down the institutional approach to a specific context and expand the commonly used methodology of analysing the effects of Cohesion Policy on economic growth by including corruption into the analysis. We also focus on applying the methodology in a specific sample of countries – the 12 new EU member states, and use the latest Cohesion Policy data for the 2007-2013 period. To the best of our knowledge, this has not been analysed before, except for the effort by Grusevaja and Pusch (2011), but their analysis covers the 2000-2006 period, thereby not taking into account the latest data. From a practical perspective, based on the results of our analysis, we provide recommendations for EU policymakers on how to further improve Cohesion Policy so that its effect on economic growth of member states is maximised.

Based on the research question, the aim of this thesis is therefore to estimate the direction and significance of the effect of corruption on the impact of EU Cohesion Policy payments on economic growth in the 12 new EU member states. In order to reach the aim, we define the following objectives:

- Review the theories of economic growth and their application in institutional economics
- Describe the corruption phenomenon and the models used for estimating its effect on economic growth

- Provide an overview of EU Cohesion Policy
- Review the methods used for analysing the effect of Cohesion Policy on economic growth
- Perform empirical research on the effect of corruption on the impact of EU Cohesion Policy payments on economic growth in the 12 new EU member states
- Provide recommendations on improving EU Cohesion Policy

Empirical research is performed using regression analysis techniques. Regressions are run in *Stata* statistical software using cross-section and panel datasets and different estimation techniques, such as random effects, fixed effects and Generalized Method of Moments (GMM) based on Arellano-Bond and Blundell-Bover estimators.

1.2. Structure of the thesis

The first part of the thesis provides an overview of economic growth models, the efforts of explaining cross-country income differences and the effect of institutions on economic growth. It also defines corruption, explains the theory behind this phenomenon, its measurement and provides an overview of research focused on estimating the impact of corruption on economic growth. In the second part, we review the structure of EU Cohesion Policy and introduce the methods used for estimating its effect on economic performance. We also review regression estimation techniques, develop our regression specification, describe the dataset as well as provide results of the analysis and subsequent recommendations.

PART I: THEORY AND LITERATURE OVERVIEW

2. MODELS OF ECONOMIC GROWTH

2.1. Solow model

We start by providing an overview of economic growth models. Solow's model, developed in 1956, is one of the most influential economic models and is used as a basis by many modern economic models as a result. It is therefore important to provide a detailed overview of its main assumptions and results.

According to Solow (1956) and Solow (1957), there are four variables in the model: output (Y) is considered to be the only commodity, capital (K) and labour (L) are the two factors of production and the level of technology (A) is the final component. The production function at any time t is defined as follows: $Y(t) = F[K(t), A(t)L(t)]$. The product of technology and labour (AL) is defined as effective labour. Capital and effective labour are said to have constant returns to scale, that is: $F(cK, cAL) = cF(K, AL)$ for all nonnegative constant terms (c). This assumption makes it possible to rewrite the production function per effective labour: $\frac{1}{AL}F(K, AL) = F(\frac{K}{AL}, 1)$. Defining $\frac{K}{AL} = k$, $\frac{Y}{AL} = y$ and $f(k) = F(k, 1)$ allows us to write the production function in the following form: $y = f(k)$, where y is output per unit of effective labour and k is the stock of capital per unit of effective labour. Production function $f(k)$ is said to have the following properties: $f(0) = 0$, $f'(k) > 0$, $f''(k) < 0$ and it must satisfy Inada conditions.

Labour (L) and technology (A) are assumed to have constant growth rates of n and g, respectively, that is $\dot{L}(t) = nL(t)$ and $\dot{A}(t) = gA(t)$, where $\dot{L}(t)$ and $\dot{A}(t)$ are derivatives with respect to time, which is assumed to be continuous. It follows that labour and technology grow exponentially: $L(t) = L(0)e^{nt}$ and $A(t) = A(0)e^{gt}$.

The economy's output consists of investment and consumption. An exogenous and constant share s of output is assigned to investment, while the depreciation of current capital (δ) reduces capital stock, therefore: $\dot{K}(t) = sY(t) - \delta K(t)$.

Several mathematical manipulations yield the most important formula of Solow's model: $\dot{k}(t) = sf\big(k(t)\big) - (n + g + \delta)k(t)$. It shows that the stock of capital per unit of effective labour changes for two reasons: the first term is new investment per unit of effective labour while the second term is break-even investment, that is investment needed to keep the capital at its current level due to growing population, improving technology and the depreciation of capital. In the steady state, actual investment and break-even investment are equal and therefore the change of capital, $\dot{k}(t)$, is equal to 0. Properties of the production function lead to the fact that the level of capital always converges to the steady-state level of capital and stays there. In the steady state, capital per worker ($\frac{K}{L}$) and output per worker ($\frac{Y}{L}$) grow at rate g – the rate of technological progress. Therefore, the main conclusion of the Solow model is that in the long-term, economic growth per worker only happens due to technological progress.

Assumptions of the Solow model simplify the main characteristics of real economies in many ways, according to Romer (2012): the economy is said to only produce a single good with three production inputs; government's decisions and economic contribution do not enter the model; unemployment is not taken into account and main variables have constant and exogenous rates of growth. Nevertheless, the model serves as a useful starting point for more complicated economic analyses, is used extensively in modern research and has been found to provide a close approximation to long-run economic growth patterns of real economies.

2.2. Other models of economic growth

Several influential economic models expand on the Solow's model and its assumption of exogenous saving rate. In the Ramsey-Cass-Koopmans model, the saving rate is determined by households and companies who maximize their utilities and profits, but labour and technology are still assumed to grow at constant, exogenous rates. Companies work in perfectlycompetitive markets, earn zero profits and have the same production function as in the Solow model. Households supply companies with labour and capital and have to decide on the optimal combination of saving and consumption which maximizes their utility. Equilibrium in the model is described by the dynamics of consumption and capital. The central equation for the change of capital stock is slightly different from that in the Solow model: $\dot{k}(t) = f\big(k(t)\big) - c(t) - (n + g + \delta)k(t)$. Despite having an endogenous saving rate, the economy in the Ramsey-Cass-Koopmans model has the same behaviour as the Solow model in equilibrium: capital per worker and output per worker grow at rate g (Romer, 2012). Thus, once again, output per worker in the long term only grows due to technological progress.

The Diamond model is similar to the Ramsey-Cass-Koopmans model in many respects, with the major difference being that the Diamond model assumes that individuals which form households are born and die throughout time. Assumptions regarding firm behaviour are the same in these two models. In the special case of logarithmic utility and Cobb-Douglas production function, the balanced-growth path dynamics of the Diamond model's economy are the same as those in Solow and Ramsey-Cass-Koopmans models. Relaxing these assumptions can lead to cases with different behaviours of the economy: equilibrium can occur at multiple values of capital stock, capital stock may converge to zero or equilibrium may be undeterminable. However, as before, the general conclusion of the model is that technological progress is the main driver of growth of output per worker in the long term.

All the above mentioned models take technological growth as given – it is exogenous. In contrast, the endogenous growth theory tries to explain the reasons behind technological progress, and as its name implies, the rate of technological progress is endogenous in these models. Romer (1990) developed a widely used endogenous growth model which incorporates microeconomic principles. The economy is divided into 3 sectors: a research sector, an intermediate-goods sector and a final-goods sector. Rogoff and Obstfeld (1996) present a more streamlined version of the Romer model. The final goods production function is equal to $Y_t = L_{Y,t}^{1-\alpha} \sum_{j=1}^{A_t} K_{j,t}^{\alpha}$, where j stands for the type of capital goods used for production, A_t is the total number of types of capital goods that have been generated up to time t (a proxy for technological progress), L_Y is labour allocated to final goods production and K stands for capital goods. In the research sector, the production function of blueprints takes the form $A_{t+1} - A_t = \theta A_t L_{A,t}$, where L_A is labour allocated to blueprint production (so that total labour $L = L_Y + L_A$) and θ is a parameter of productivity. The intermediate goods sector is assumed to be composed of monopolistic firms which each buy a certain type of blueprint from the research sector, manufacture machines using these blueprints and sell them in the next period to firms in the final goods sector. On the demand side, consumers are said to live infinitely and have the following utility function: $U_t = \sum_{s=t}^{\infty} \beta^{s-t} \frac{C_s^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}$. In the special case of logarithmic utility, the model produces the following formula for the growth rate of technology: $\bar{g} = \frac{\alpha\beta\theta L-(1-\beta)}{1+\alpha\beta}$. Therefore, in this model, the size of population (amount of total labour L) has a positive effect on technological progress since a larger population can create more R&D products. Apart from the size of population, economic growth is affected by output elasticity parameter (α), a parameter for the rate of impatience of consumers (β) and productivity in the research sector (θ).

Overall, while Solow, Ramsey-Cass-Koopmans and Diamond models all lead to the conclusion that technological progress is the only reason for long-run growth of output per worker, Romer's endogenous growth model implies that the growth of technological progress (and economic growth as a result) is affected by other factors as well, such as the amount of labour supplied or research productivity.

3. ACCOUNTING FOR ECONOMIC DEVELOPMENT DIFFERENCES ACROSS COUNTRIES

For many years researchers have been trying to understand the large differences in average income per person around the world. The wealthiest countries in the world are about 30 to 40 times more affluent than the poorest ones (Cubas, Ravikumar, & Ventura, 2013). There are two main explanations for the cross-country income difference phenomenon. One of them claims that this income variation comes from the differences in accumulation of physical and human capital, while the other states that it comes from the differences in total factor productivity (TFP), that is, technological progress. Some researches try to estimate the contributions of each component of economic growth models to differences in levels or growth rates of income. Such accounting literature provides significant insights into the cross-country income difference puzzle. In the following part, before reviewing the accounting literature, we will first introduce an augmented Solow model that includes human capital.

3.1. Augmented Solow model

In the augmented Solow model, the output function is defined as $Y = F[A, K, H, L]$. Here, H is the stock of human capital, and other variables are the same as those in the original Solow model. The augmented Solow model was introduced by Mankiw, Romer and Weil (1992) in their influential paper *"A Contribution to the Empirics of Economic Growth"*. The reason for adding human capital into the model is that the empirical facts are not in accordance with the initial Solow model. For example, in Mankiw et al.'s (1992) paper, the initial Solow model is tested on factual data and the implied capital share α is found to be 0.59, which is much larger than the common sense of 1/3.

Human capital in the model refers to the quality of education that could change the productivity of a worker. Like physical capital, human capital accumulates from one year to the next year as a certain fraction of GDP is added to it.

Assuming a Cobb-Douglas production function, the augmented Solow model is described by the equation $Y(t) = K(t)^{\alpha_k} H(t)^{\alpha_h} [A(t)L(t)]^{1-\alpha_k-\alpha_h}$. Similarly to the original Solow model, we can get the dynamics of physical capital and human capital as:

$$\dot{k}(t) = s_k y(t) - (n + g + \delta)k(t),$$

$$\dot{h}(t) = s_h y(t) - (n + g + \delta)h(t),$$

where s_k and s_h are the fractions of income invested in physical and human capital, respectively. The steady state of the model is described by the following equations:

$$k^* = (\frac{{s_k}^{1-\alpha_h} {s_h}^{\alpha_h}}{n+g+\delta})^{1/(1-\alpha_k-\alpha_h)},$$

$$h^* = (\frac{{s_k}^{\alpha_k} {s_h}^{1-\alpha_k}}{n+g+\delta})^{1/(1-\alpha_k-\alpha_h)}.$$

The augmented Solow model performs much better in the empirical application than the original model. As Mankiw et al. (1992) show, in the cross-section regression of output per worker for 98 countries, the implied α_k is 0.31, which is close to 1/3. The adjusted R^2 is 0.78, which means that physical and human capital accumulation can explain nearly 80% of country variation in income. However, this conclusion is sensitive to the measurement of human capital. We will explore this subject in more detail in the following part.

3.2. Accounting-style income decomposition

Literature that focuses on decomposing income differences into components contains both levels accounting and growth accounting. In this section, we only focus on the levels accounting. The purpose of levels accounting is to decompose income differences into physical capital accumulation, human capital accumulation and other factors. As mentioned before, researchers do not reach the consensus in their conclusions about the reasons for cross-country income differences. For example, Mankiw et al. (1992) argue that differences in physical and human capital can explain a large part of disparity in income, while Hall and Jones (1999) and

Klenow and Rodriguez-Clare (1997) argue that the two factors account for much less of the variation. The disagreement mainly comes from the different measurements of human capital.

Many levels accounting researchers begin from assuming a Cobb-Douglas production function: $Y(t) = K(t)^{\alpha_k} H(t)^{\alpha_h} [A(t)L(t)]^{1-\alpha_k-\alpha_h}$, where $\alpha_k + \alpha_h < 1$ reflects the decreasing returns to all capital. By dividing by *L(t)* on both sides and rearranging, we can express income per capita as $\frac{Y(t)}{L(t)} = \mathrm{A}(\frac{K(t)}{Y(t)})^{\frac{\alpha_k}{1-\alpha_k-\alpha_h}}(\frac{H(t)}{Y(t)})^{\frac{\alpha_h}{1-\alpha_k-\alpha_h}}$. In order to estimate the coefficients, the main problem is the measurement of two capital intensities, $\frac{K}{Y}$ and $\frac{H}{Y}$.

The measurement of physical capital is similar across many papers. Researchers use the perpetual inventory method to estimate the capital stock. With data on investment and depreciation rate, the capital stock in year *t+1* is defined as: $K_{t+1} = (1-\delta)K_t + I_t$, where δ is the depreciation rate and I_t is investment in year *t*. We still need an initial capital stock K_0, but it can be calculated as $I_o/(g+\delta)$. If we suppose that 1960 is the first year of investment data available, then $K_{1960} = I_{1960}/(g+\delta)$, where g is the average geometric growth rate of investment between 1960 and the beginning of researching year, say, 1970. However, the measurement of human capital stock is quite different.

Mankiw et al. (1992) further rearrange the above formula by taking logarithms on both sides and substituting the steady state amount of k and h to get: $\ln(\frac{Y(t)}{L(t)}) = \ln\mathrm{A}(0) + \mathrm{g}t + \frac{\alpha_k}{1-\alpha_k-\alpha_h}\ln(s_k) + \frac{\alpha_h}{1-\alpha_k-\alpha_h}\ln(s_h) - \frac{\alpha_k+\alpha_h}{1-\alpha_k-\alpha_h}\ln(n+g+\delta)$. This leads to a result that $\frac{H}{Y} = \frac{s_h}{n+g+\delta}$. But the question is how to estimate the investment ratio in human capital, s_h. Mankiw et al. (1992) use a proxy to measure s_h, which is defined as:

$$s_h = \text{Secondary school enrollment rate} * \frac{\text{population aged } 15-19}{\text{population aged } 15-64}$$

However, they also argue that this measurement is not precise as it does not include the input of teachers and does not measure primary and higher education. With this measurement, Mankiw et al. (1992) use the average annual data from the period 1960-1985 to get s_k and s_h, and then use them in the regression of 1985 levels of $\frac{Y}{L}$. Their sample covers 98 countries, but excludes countries where oil production is the dominant industry. They obtain an adjusted R^2 of 0.78, and the estimated coefficients imply the physical capital elasticity $\alpha_k = 0.31$ and the human capital elasticity $\alpha_h = 0.28$, which are both reasonable. Finally, they conclude that the differences in accumulation of physical capital and human capital can explain most of cross-country differences in income.

Klenow and Rodriguez-Clare (1997) argue that Mankiw et al.'s (1992) measurement of human capital overestimates the variation across the world because it excludes primary school attainment, which varies much less across countries. In their test, by simply adding primary and tertiary school enrolment, the decomposition of income per capita moves from Mankiw et al.'s (1992) original of 78% capital accumulation and 22% other factors to 40% capital accumulation and 60% other factors. They suggest using Mincer regression to estimate human capital stocks, which is to run a regression of log wages on years of schooling and experience: $\ln(y) = \ln(y_0) + rS + \beta_1 X + \beta_2 X^2$, where y stands for wages, y_0 is the wage of a person without education and experience, *S* is the years of schooling and *X* is the years of experience. They construct the estimation of human capital as follows: $h_s = \left(\frac{K_H}{L_H}\right)^{\alpha_k} (h_T)^{\alpha_h} (Ae^{\left(\frac{\gamma_1 s + \gamma_2 exp + \gamma_3 exp^2}{1-\alpha_k-\alpha_h}\right)})^{1-\alpha_k-\alpha_h}$. By manipulating this equation, we can get the capital-output ratio as: $\frac{H}{Y} = (e^{\gamma_1 s} \sum_i \omega_i \, e^{\gamma_2 exp_i + \gamma_3 exp_i{}^2})^{\frac{1-\alpha_k}{\alpha_h}} \frac{AL}{Y}$, where *s* is the average years of schooling in the total population over 25 years old, exp_i is somebody's experience in age group i and $exp_i = (age_i - s - 6)$, ω_i is the proportion of the population in ith group, $age_i = \{27, 32, \ldots, 62\}$ for the age groups $\{25-29, 30-34, \ldots, 60-64\}$. The coefficients on schooling and experience γ_1, γ_2, γ_3 are set to be 0.095, 0.0495, 0.0007, respectively, based on the results of Mincer regression.

With this measurement, Klenow and Rodriguez-Clare (1997) find that the physical and human capital accumulation only explain 42% of GDP differences across countries, which is to say that the residual term has a dominant role for cross-country income differences. Therefore, they conclude that *"theorizing about international output differences should center at least as much on differences in productivity as on differences in physical or human capital intensity"* (Klenow and Rodriguez-Clare, 1997).

Hall and Jones (1999) use a similar calibration method as Klenow and Rodriguez-Clare (1997). Their production function is: $Y = AK^{\alpha}H^{1-\alpha}$. This function is a special case of what Mankiw et al. (1992) and Klenow and Rodriguez-Clare (1997) used with $Y = K^{\alpha_k}H^{\alpha_h}L^{1-\alpha_k-\alpha_h}$, where $\alpha_h = 1 - \alpha_k$.

In output per worker terms, the production function can be written as: $y = Ak^{\alpha}h^{1-\alpha}$. By manipulating it, we can get: $\ln(y) = \frac{1}{1-\alpha}\ln(A) + \frac{\alpha}{1-\alpha} ln\left(\frac{k}{y}\right) + \ln(h)$, where $\frac{k}{y}$ is physical capital-output ratio and h is human capital per worker.

Hall and Jones' (1999) measurement of human capital does not depend on experience. They construct the human capital function as: $h_i = e^{\phi(S_i)}$, where $\phi(S_i)$ is the average number of years of education of workers in country i. They also assume $\phi(S_i)$ is piecewise linear with a slope $\phi'(S_i) = 0.134$ if $S \leq 4$, 0.101 if $4 < S \leq 8$, 0.068 if $S > 8$. These rates of return on education are reported in Psacharopoulos (1994).

With this measurement, by comparing US and other 132 countries in the sample, they find large productivity disparities across countries. For example, Canada is 6.1% worse than US in output per worker terms, and this almost entirely comes from a 5% lower human capital per worker, thus the productivity in Canada is about the same as US. However, $\ln(y)$ in China is 2.815 lower than in the US, and of that 1.033 comes from physical capital, 0.319 comes from human capital, and this thus leads to a 1.462 difference in productivity. In general, OECD countries have productivity close to the US, while for developing countries differences in productivity are the crucial factor in explaining cross-country income differences. This conclusion is similar to that of Klenow and Rodriguez-Clare (1997), but different from Mankiw et al. (1992). Furthermore, Hall and Jones (1999) find that productivity is highly correlated with physical and human capital, while Mankiw et al. (1992) assume that they are uncorrelated.

Cubas, Ravikumar & Ventura (2013) provide another measurement of human capital. Their idea is similar to that of Kaarsen (2014), who argues that one year of schooling in USA corresponds to three or more years of schooling in less-developed countries. Cubas, Ravikumar & Ventura (2013) use PISA scores to measure the quality of labour and show that there are larger disparities in labour quality among countries than the Mincerian returns suggest. As a result, total factor productivity differences from their model are less significant than those from other authors. Kaarsen (2014) uses TIMSS scores to consider the quality of labour and finds that the variance of income explained by the augmented Solow model rises by 22%. However, the author also argues that TFP differences are still the dominant force in accounting for cross-country income differences.

Overall, the paper by Mankiw et al. (1992) can be seen as the revival of neoclassical growth theory, which argues that most of the variation in cross-country income differences is linked with variation in factor inputs. Their paper also triggered a great passion for researchers to figure out the cross-country income difference puzzle. Even though by using different measurements of capital stock, the explaining power of factor inputs becomes larger than what Hall and Jones (1999) believed it to be, it can be argued that TFP is at least as important as physical and human capital in explaining cross-country income differences (Caselli, 2005). The next question therefore is what determines the difference of TFP across countries? Inspired by Hall and Jones (1999), we consider that institutions play an important role in this subject. The following part will review the relevant literature and go deeper into the institutional theory.

4. INSTITUTIONS AND ECONOMIC DEVELOPMENT

Apart from decomposing the output per worker differences across countries into differences in inputs and productivity differences, Hall and Jones (1999) also come up with suggestions for the reasons behind these differences. Their main hypothesis is that differences in long-run levels of economic development between countries are due to social infrastructure. They define social infrastructure as *"the institutions and government policies that provide the incentives for individuals and firms in an economy"* (Hall and Jones, 1999). The authors claim that in the cases where thievery or corruption are widely spread, they can be thought of as taxes on output and therefore inputs do not bring their full returns to their owners. Moreover, the owners of inputs must invest a part of their funds into protecting their property from being stolen or misused. In terms of protection from such negative factors, the government is the most important agent as it can engage in protecting activities on a large scale and as it has the actual power to create rules and control their implementation.

Hall and Jones (1999) argue that perfect measures of social infrastructure do not exist, but that certain proxies can be used nevertheless. They use two measures of social infrastructure:

- GADP (government anti-diversion policies) index from a company named *Political Risk Services*, where the authors average the country scores across 5 categories of indicators: law and order; bureaucratic quality; corruption; risk of expropriation; government repudiation of contracts.
- Openness to international trade index compiled by Sachs and Warner in 1995, which rates countries according to the percentage of time that they were open to trade in the period from 1950 to 1994. Openness to trade is said to be important because barriers to trade create inefficiencies for companies operating in the market.

The main parameter of social infrastructure used by the authors is the average of these two indices. The structural model used to identify the impact of social infrastructure on economic development is as follows: $\log\frac{Y}{L} = \alpha + \beta * S + \epsilon$ and $S = \gamma + \delta * log\frac{Y}{L} + \theta * X + \eta$, where S

is the measure of social infrastructure and X is a combination of all other important variables. The main fact that can be observed from these equations is that social infrastructure is assumed to be endogenous: not only does it affect the level of output per worker, but it can be influenced by output per worker itself. In order for the equations to be identifiable, the authors assume that all the variables entering into X do not affect output per worker directly – that is, they can be used as instruments for the first equation. When it comes to the instruments, the authors use several variables which characterize the levels of Western European influence on different countries. The first of them is whether Western European languages are used as the first language, measured as the share of population speaking these languages as a mother tongue. The second instrument is distance from equator. These variables affect social infrastructure as the countries that were influenced by European countries consequently had better social infrastructure, and as Western European colonizers more often came to areas far from equator that had a similar climate. In order to be valid instruments, these variables should also have no direct effect on output per worker today, which, according to the authors, is true. They also use an additional instrument in several regressions – the share of trade in an economy.

Hall and Jones' (1999) main result from running these regressions is that an increase of 0.01 of social infrastructure (measured on a scale from 0 to 1) leads to an increase in output per worker of 5.14%. Hence, their explanation for large differences across countries of the productivity residual (and output per worker) is social infrastructure.

Institutions are also found to be important by Knack and Keefer (1995), but instead of using levels accounting, the authors use growth accounting. They claim that effective contract enforcement is the key to economic development. The authors test the importance of institutional variables for conditional convergence, which claims that countries with lower steady-state incomes will tend to move slower to this level from an original level of income. Moreover, countries with bad institutional characteristics will tend to have lower steady-state income levels.

Knack and Keefer (1995) analyse the following institutional variables from two organizations, *International Country Risk Guide (ICRG)* and *Business Environmental Risk Intelligence (BERI)*, which are then combined into two indices (ICRG and BERI):

- Expropriation risk
- Rule of law
- Repudiation of contracts by government
- Corruption in government
- Quality of bureaucracy
- Contract enforceability
- Infrastructure quality
- Nationalization potential
- Bureaucratic delays

Expropriation risk, Rule of law and Contract enforceability are measures of property rights; Repudiation of contracts by government and Nationalization potential – of government trustworthiness; Corruption in government, Quality of bureaucracy, Infrastructure quality and Bureaucratic delays – of government efficiency.

Knack and Keefer (1995) use the convergence equation developed by Barro and Sala-i-Martin (1992): $\ln\left(\frac{y_{iT}}{y_{i(0)}}\right) = a_i - (1 - e^{-\lambda})(\ln\left(y_{i(0)}\right) - g_i T) + \epsilon_i$, where $a_i = g_i + \left(1 - e^{-\lambda}\right)\ln(y_i^*)$. y_{iT} is the level of income of country i at time T, $y_{i(0)}$ is the initial level of income, y_i^* is the steady-state level of income, g_i is the rate of growth of technology (assumed to be constant) and λ is the convergence rate. The basic regression, against which the proposed institutional variables are tested, is a simple ordinary least squares regression of average GDP growth on the initial GDP level, primary and secondary school enrolment, government consumption as percentage of GDP, frequency of revolutions, frequency of assassinations and an investment deflator variable.

The main hypothesis is that the coefficient on the initial GDP level should be higher (more negative) when correct institutional variables are used: countries with low initial GDP levels will grow faster. The authors indeed find that the coefficient on the initial income level is more negative and more statistically significant when their ICRG and BERI indicators are added to the original equation. This, in their opinion, confirms the conditional convergence hypothesis and the fact that institutional variables are important explanatory factors in the model.

The most influential and discussed paper on the importance of institutions for economic development comes from Acemoglu, Johnson and Robinson (2001). The authors analyse the effects that different colonization policies of European settlers had on economic performance of the colonies. Their approach is similar to that of Hall and Jones (1999) as they use instrumental variables estimation as well.

The main idea behind Acemoglu, Johnson and Robinson's (2001) approach is that European settlers used different types of colonization in different countries: countries with dangerous diseases and high mortality rates received extractive colonization policies with bad property protection and authoritarian governments, while countries with favourable conditions received institutions similar to those of European countries as the colonizers planned to settle in these colonies. It is then likely that after the colonies regained independence, they built their new institutions based on their original institutions. Thus, the original institutional types have an effect on current economic performance of these countries.

The authors use mortality rates of early colonial settlers in the 17^{th}-19^{th} centuries as instrumental variables for econometric estimation. They argue that mortality rates were one of the most important factors that were taken into account by European colonizers when making decisions about where to settle. Mortality rates were even reported in the press in some countries. As an example, when the British officials had to decide where to send convicts, the original idea was to send them to the area around Gambia river in Africa. However, they decided that the disease environment in the region is too unfavourable and sent the convicts to Australia instead. Acemoglu, Johnson and Robinson (2001) also provide evidence for the persistence of colonial institutions after the colonies regained independence. For example, in Latin America, monopolies, forced labour and slavery were still common long after the countries became independent.

The authors use the following institutional variables in their analysis:

- Protection against expropriation (1985-1995 average), developed by an organization called *Political Risk Services*
- Constraints on executive in 1990, developed by Ted Robert Gurr and associates
- Constraints on executive in 1900, from Gurr data
- Constraints on executive in the first year of independence, from Gurr data
- Democracy index in 1900, from Gurr data

Acemoglu, Johnson and Robinson's (2001) main hypotheses are described by a system of four equations:

$$log\ y_i = \mu + \alpha * R_i + \gamma * X_i + \epsilon_i,$$

$$R_i = \lambda_R + \beta_R * C_i + \gamma_R * X_i + v_{Ri},$$

$$C_i = \lambda_C + \beta_C * S_i + \gamma_C * X_i + v_{Ci},$$

$$S_i = \lambda_S + \beta_S * logM_i + \gamma_S * X_i + v_{Si},$$

where y_i is per capita income in country i, R_i is the index of Protection against expropriation (measure of current institutions), X_i contains other important variables, C_i is the measure of early institutions, S_i measures the percentage of colony population with European descent, and M_i is the rate of mortality of early settlers. The authors therefore suppose that mortality rates had an effect on colonial settlement, colonial settlement had an effect on early institutions, early institutions had an effect on current institutions and that current institutions affect the current per capita incomes of countries.

The evidence for the effect of current institutions on current income levels is found to be statistically significant: institutions explain more than 50% of variation in incomes across countries. However, the direction of the effect is not clear: better institutions could lead to higher income per capita, or higher income per capita could lead to better institutions. For this reason, the authors use settler mortality as the instrumental variable, with the restriction that it does not affect current income per capita levels directly. The results suggest that current institutions do indeed affect the levels of per capita income.

Acemoglu and Robinson (2010) provide more details on the theory behind the effect of institutions on economic development of countries. They describe three types of institutional characteristics: economic institutions, political power and political institutions. Economic institutions affect economic development through the incentives that they provide for economic agents: good economic institutions foster investments into human and physical capital and technological advancement. Economic institutions are determined by the society itself, and therefore interest groups with the most political power have the largest effect on the institutional system. Political institutions (for example, government form) have *de jure* political power and affect incentives, too. There are also parts of society that have *de facto* political power because they are the owners of large amounts of resources and can affect the political institutions. Moreover, political institutions and the allocation of resources in a society are quite persistent. Therefore, reforming institutions is not easy. These ideas can be summarized in the following way: political institutions and the distribution of resources determine how *de jure* and *de facto* political power in a society is distributed, while the distribution of political power determines the dominating economic institutions, which, in turn, have an effect on the economic performance and distribution of resources in the following period.

Richter and Timmons (2012) expand on the empirical results of Acemoglu, Johnson and Robinson (2001) by analysing growth rates of income instead of levels of income. Firstly, they calculate the average annual growth rate of per capita income for the period 1820-1995, in order for the time period to correspond to that of Acemoglu, Johnson and Robinson (2001).

The authors also use the same institutional variables, namely, Protection against expropriation and Constraints on the executive, to have a similar econometric framework. In general, their econometric model is exactly the same as the one used by Acemoglu, Johnson and Robinson (2001), except for the fact that the main variable of interest is economic growth instead of the level of economic development: $g_i = \alpha + \beta * D_i + \varepsilon_i$, where g_i is the long-run average annual growth rate of income per capita (1820 to 1995) and D_i is the institutional variable. Similar to the original paper, Richter and Timmons (2012) find a statistically significant impact of institutions on long-run economic growth. However, the effect of institutions is rather small in absolute terms. The authors find that if the quality of institutions in a country improved by one standard deviation, the country's average annual growth rate of per capita income would increase by 0.42%. Over a very long term, this leads to significant differences in per capita income across countries due to the compounding effects of growth; however, such small gains are not substantial in the short term.

Nevertheless, not all researchers support the view of Acemoglu, Johnson and Robinson (2001). The main critique comes from Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004). They argue that the relationship between institutions and economic development works in the opposite way: instead of institutions causing economic growth, it is growth itself that causes improvements in institutions. According to the first view, political institutions that support protection of property rights will lead to more investment into physical and human capital and stimulate economic growth. The second view, on the other hand, specifies that investments into physical and human capital will lead to improvements in education and wealth, and this in turn will lead to improvements in political institutions.

Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) provide three main points of critique for Acemoglu, Johnson and Robinson's (2001) empirical results. First of all, they argue that the measurement of institutions is inaccurate. The authors claim that institutions have two main characteristics: they constrain individuals to behave for the benefit of the society and these constraints have to be permanent. The variables used by Knack and Keefer (1995), Hall and Jones (1999) and Acemoglu, Johnson and Robinson (2001) do not satisfy these criteria, according to them. Instead of being permanent, all the variables measure temporary outcomes and have a tendency of improving with economic development as well as being volatile. Moreover, none of the variables, apart from the ones measuring constraints on the executive, reflect actual constraints. Furthermore, these variables are not correlated with measures of constitutional constraints.

Secondly, the authors run simple ordinary least squares regressions and do indeed find that institutional variables are statistically significantly related to growth in per capita income, but only during the same time period. This means that causality could go in either way. On the other hand, Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) find that institutional variables in the current period do not predict future economic development, while the level of human capital does have this predictive power.

Finally, the authors argue that the instrumental variables approach used by Acemoglu, Johnson and Robinson (2001) is not a valid method of determining the relationship between institutions and economic growth. Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) claim that apart from bringing the set of institutions to the colonies, European settlers brought human capital as well. This means that the key assumption of the instrumental variables approach, that is the fact that these variables are uncorrelated with the error term, may not be true. If human capital is omitted in the regressions, instrumental variables (for example, settler mortality) will be correlated with the error term and regression results will be biased. Moreover, settler mortality is not correlated with measures of constitutional constraints, while the initial levels of human capital are significantly related to current levels of economic development. Finally, human capital variables also appear to be more statistically significant in the instrumental variables regressions than do institutional variables. Overall, Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) claim that human capital is more important for economic development of countries and that it is the level of human capital in the society that determines its institutions.

Acemoglu, Gallego and Robinson (2014) respond to this critique. They argue that the returns to human capital in usual regressions are biased upwards and that institutional characteristics are still a better predictor of economic development. According to the authors, developments of physical and human capital are not the causes of growth – they are the growth itself. Instead, what causes growth is the fact whether a country has favourable institutional conditions or not.

Firstly, Acemoglu, Gallego and Robinson (2014) provide historical evidence that European settlers did not always bring more human capital to the places where they settled. For example, South American colonizers were more educated than the colonizers of North America, even though North America is a much more developed region nowadays than South America is and has better institutions. As a proof of this, the authors provide statistics that show that around 80% of colonizers of South America could sign their name, while the same statistic for the colonizers of North America is only around 40%. There is therefore no historical connection between the level of human capital and economic development of colonies.

Secondly, Acemoglu, Gallego and Robinson (2014) run econometric regressions and find that human capital returns are lower than those reported by Glaeser, La Porta, Lopez-de-Silanes and Shleifer (2004) and that institutional variables are a more important predictor of economic development. The authors use the logarithm of GDP per capita in purchasing power parity terms in 2005 from the Penn World Tables as the main dependent variable and average years of schooling in 2005 as the educational indicator. Institutions are measured by the rule of law index in 2005, which comes from the World Bank, while logarithms of settler mortality and population density in 1500 are the two instrumental variables. Differences in human capital across the colonies are measured by the activity of protestant missionaries, that is the number of protestant missionaries per 10 thousand of population in 1920. The missionaries made a large contribution towards the creation of educational institutions in the colonies. Human capital is also measured by enrolment rates for primary education in 1900.

Basic ordinary least squares regressions of GDP per capita on human capital and institutional variables show that human capital returns are much larger than they are supposed to be based on microeconomic evidence. Returns to education are normally in the region of 6%-10%, while Acemoglu, Gallego and Robinson (2014) find them to be around 30-35% and conclude that there must be omitted variable bias. Semi-structural models that control for the determinants of institutions and years of schooling show that human capital does indeed have returns close to 6%-10%, while institutions are found to be more important. Structural models with instrumental variables also allow the authors to reach the same conclusions. Moreover, they run regressions of institutional measures on human capital measures to see if it is the human capital that determines institutional characteristics of a country. These regressions find human capital to be insignificant. The same results are evident when using cross-regional data, in contrast to cross-country data. All in all, Acemoglu, Gallego and Robinson (2014) conclude that institutions are a much more important channel of economic growth than human capital is.

As can be seen from the overview above, the relationship between institutions and economic development is well documented in academic research. Even though the methodology used to arrive at these conclusions and the conclusions themselves are subject to some critique, as is usually the case, most of the researchers agree that institutions are at least one of the explanations behind cross-country differences in economic development. The next chapter will focus on one of the institutional variables in particular – corruption, as it is at the center of attention of our research.

5. CORRUPTION

Corruption is a characteristic of the world that has been present throughout time and across countries. But only after 1996, when World Bank president James Wolfensohn raised it up at the joint annual meeting of World Bank and IMF, corruption began to attract much more of the world's attention. For many years, researchers have been spending large proportions of time to dig into this important topic. In some aspects their findings vary largely, while in others researchers seem to be reaching a consensus. In this section, we aim at providing a comprehensive introduction of corruption. Our logical clue will follow along these four topics: the definition of corruption, the theory of corruption, the measurement of corruption and the influence of corruption on economic development.

5.1. Definition of corruption

5.1.1. Definition

Corruption has been defined in many ways. It is hard for researchers to agree on a precise definition. Jain (2001) argued that *"corruption refers to acts in which the power of public office is used for personal gain in a manner that contravenes the rules of the game"*. In short, corruption in this sense can be defined as the abuse of public power for private gains (Svensson, 2005). Based on this definition, Aidt (2003) proceeds to characterize three conditions which define corrupt behaviour:

- *Discretionary power:* the public officials must have the power to create and oversee rules and laws in a discretionary manner
- *Economic rents:* the discretionary power can be used to extract rents
- *Weak institutions:* the institutions must leave room for officials to use their power to get or seek rents

However, corruption does not necessarily happen in the public sector. For example, in large private companies we can often observe corrupt behaviour in hiring, procurement or promotion processes.

5.1.2. Types of corruption

Jain (2001) identified three types of corruption based on where the corrupt behaviour happens and who abuses his authority. These are grand corruption, bureaucratic corruption, and legislative corruption.

Grand corruption arises when political elites design and implement policies to serve their own interests. This type of corruption has the worst influence on the nation since it comes from the top of the national hierarchy. The whole public sector may be designed as a rent-seeking machine, and the resources of the country may be transferred to private agents. Rose-Ackerman (2007) points out that this corruption type can destroy the economy and bring the country to the edge of failure. The opposite of grand corruption is petty corruption, which occurs at a lower hierarchical level than grand corruption and involves smaller payments. Jain (2001) calls petty corruption as bureaucratic corruption.

Bureaucratic corruption refers to the bribes that bureaucrats take to gain benefits that are not supposed to belong to them, as well as to the rent-seeking actions that bureaucrats perform. This kind of corruption is the most common type observed in the real world, but is not as damaging as grand corruption.

Finally, legislative corruption refers to the acts that influence the voting system. Large companies or interest groups may bribe legislators to enact specific legislation that can benefit them. This kind of corruption also includes "vote-buying" behaviour, where legislators pass regulations that are not based on public benefits in exchange for being re-elected (Jain, 2001).

5.2. Theory of corruption

From the definition above, it is not surprising that corruption is regarded as illegal behaviour in most countries because corruption distorts markets and creates unfair competition among market participants. However, we cannot simply assume that corruption only has negative effects. There are some countries which are characterized by both high levels of corruption and high economic growth. The most obvious examples are found in Asia, such as China, South Korea, and Thailand as well as in some developed countries, such as Italy. Researchers are attracted by the double-edged effects of corruption and have many theories that try to answer the question of how corruption is generated. Aidt (2003) provides four potential explanations, which are efficient corruption, corruption with a benevolent principal, corruption with a non-benevolent principal and self-reinforcing corruption. In the following sections, we will review these theories.

5.2.1. Efficient corruption

Efficient corruption, which is defined by Aidt (2003) as *"corruption which arises to facilitate beneficial trade between agents"*, is a theory that views corruption in a positive light. According to this point of view, corruption can be seen as a rational reaction to existing inefficiencies in public administration. Through corruption, both parties can benefit so that the social efficiency is improved. However, there are some limitations from the theoretical perspective which we will specify later.

Aidt (2003) makes a clear illustration of the mechanisms of efficient corruption. Given that the market has many inevitable inefficiencies created by government regulations and policies, the first-best solution is to eliminate the inefficiencies, and the second-best solution is to improve efficiency by letting market participants bypass such regulations and policies. Corruption is one method of the second-best solution, which can work through two channels: it can improve government efficiency and make the provision of government services more competitive.

Lui (1985) introduced the “queue model” to prove the allocative power of corruption. In this model, people are assumed to be waiting in a queue for a service. They can decide to bribe the service provider in order to buy better positions in the queue. Obviously, people with higher values of time are willing to pay more so that they can be placed ahead of those with lower values of time. Lui (1985) showed that in equilibrium, this “bribe rule” can minimize the average costs associated with spending time in the queue. Moreover, the service provider (bureaucrat) will not try to increase the time spent in a queue because this will cause a reduction of people joining in the queue and reduce his earnings from the bribe. As a result, resources are better allocated to the most willing people and social efficiency is improved.

Shleifer and Vishny (1994) analysed the interaction between private agents and politicians and showed that a bribe is an efficient method of allocating wealth between the two parties. They argued that without corruption, politicians would try to extract rents by other inefficient ways so that the resources would be allocated inefficiently. Therefore, corruption is a way to enable private agents to protect themselves from the politically imposed inefficiencies. From this point of view, corruption increases efficiency.

However, the theory of efficient corruption has theoretical limitations. Aidt (2003) argues that there are four main problems of the theory. Firstly, bureaucrats can often change the quality and quantity of their services and may be willing to deviate from the equilibrium outcome in order to maximize the bribe. Secondly, the cost of corruption cannot be neglected because such costs are used in the process of looking for whom to bribe and hiding the corrupt behaviour. Thirdly, corrupt contracts are not official and cannot be obligatory so their outcomes are very uncertain. Finally, efficient corruption theory assumes that the government inefficiency is not created by corruption itself and so is exogenous. In reality, it is reasonable to assume that politicians make improper policies because they want to benefit from them. Thus, the relationship between government inefficiency and corruption should be endogenous.

5.2.2. Corruption with a benevolent principal

This theory assumes that in a certain country, the government creates the institutional structure of the country, and hires agents to operate these bureaucracies for necessary activities, such as collection of taxes, provision of laws, etc. Therefore, the agents own some authority that is delegated to them by the government. It is reasonable to think that there exists the potential for these agents to extract rents from their position. That is how corruption is created. In this model, the government wants to design an optimal institutional framework that is corruption-free, but corruption still exists because of agents.

Aidt (2003) introduced a principal-agent model to illustrate this type of corruption. Suppose that in a tax collection activity, a firm is responsible to pay taxes t only when it has a positive profit, $\pi > 0$. The probability of $\pi > 0$ is h, thus with probability $1 - h$ the firm does not need to pay the tax. There is a tax collector whose responsibility is to report to the government which firm is liable to pay the tax and which is not. Obviously, there exists a possibility that the tax collector agrees not to report it if he is bribed. However, the government has a probability p to discover the illegal behavior. On this occasion, the tax collector and firm must pay a penalty f and g, respectively. The tax collector can earn a public wage of w and can earn w_0 if he works in the private sector. A fraction γ of tax collectors are assumed to be honest and all involved agents are risk neutral. There also exists a transaction cost to the bribe in the form of its fraction, k ∈ (0,1].

When a corrupt tax collector discovers a company which has to pay the tax, the company has a good reason to bribe him, thus its estimated gain is $\pi - pg$. Therefore, the bribe is $b = \max[k(\pi - pg), 0]$. The tax collector is willing to accept the bribe only when the expected gain, $(1 - p)(w + b) + p(w_0 - f)$, is larger than his public sector salary, w, that is $(1 - p)b + p(w_0 - w - f) > 0$. From the above equation, we can see that the occurrence of corruption depends on three variables: the wage w, the monitoring system p and the penalties f and g, which can be seen as indicators for a good design of public bodies.

Becker and Stigler (1974) argued that the government can set an "efficient wage" that can discourage the corrupt behavior as it raises the cost of illegal behavior and makes agents more hesitant to do so. If we set $f = 0$ in the above equation, which means that no one gets a fine for bribing, we can get the efficient wage $w^e = w_0 + \frac{1-p}{p} b$. From this equation, we can see that setting a high wage in the public sector is a useful tool of reducing corruption. Some countries have already adopted this policy and have achieved excellent results, for example, Singapore. As Rahman (1986) said, "*the government (in Singapore) believed that an efficient bureaucratic system is one in which the officers are well paid so the temptation to resort to bribes would be reduced*". However, some researchers argue that this policy will cause other problems. Firstly, even though the agents would be more hesitant to be corrupt when they face a high wage, the agents who remain corrupt can start requesting larger bribes to offset the increased threshold. Secondly, a high wage policy may be incompatible with other public policies. For example, Besley and McLaren (1993) showed that total government revenue will be lower when they use efficient wage policy, compared to "*capitulation wage*" policy – an extremely low wage that makes almost all agents corrupt and which only generates income for the government when corrupt agents are caught. Thirdly, the high wage in the public sector will attract more talents to it and reduce the number of talented employees in the private sector.

Another way through which corruption can be reduced is by enhancing the monitoring system (setting $p \approx 1$). However, it can also be argued that it is hard and costly to design an effective monitoring system. Laffont and N'Guessan (1999) raised an opposite point that the people performing the monitoring can be corrupt as well. In this case, a rise in p (hiring more monitors) may lead to a higher corruption level instead of reducing it.

Finally, what about increasing the penalties: f and g? Obviously, high penalties can decrease the expected gains from corruption. If we set $f = b$, the expected gain for agents would be $(1 - 2p)b + p(w_0 - w)$; then if $p > 1/2$, corruption can be eradicated if reservation wages are used. The problem is that things are more complicated in practice. In the real world, relatively few people are punished for acts of corruption, and there seems to be a wide gap between the penalties specified in the laws and the ones effectively imposed (Tanzi, 1998). In addition, the judicial system may also be corrupt or influenced by political reasons. All these factors decrease the effectiveness of the punishment policy.

5.2.3. Corruption with a non-benevolent principal

As showed above, the benevolent principal theory assumes that the principal aims at designing an optimal institutional framework, and that corruption is an unexpected but unavoidable part of the design. However, it is not surprising that in the real world many governments are corrupt themselves.

Aidt (2003) took a simple example to demonstrate the general point of non-benevolent principal corruption. Suppose that a kind of economic activity can only be launched by holding a license that is issued by a government officer. Let λ be the number of licenses already issued and $b(\lambda)$ be the price of an additional license. It is reasonable to assume that $b'(\lambda) = \frac{\partial b}{\partial \lambda} < 0$, which means that the more licenses there are in the market, the lower the price of obtaining a new one. When there is a free-competition market, $b(\lambda_H) = 0$, and λ_H represents the number of firms in the free-competition market. The government officer can ask for a bribe in exchange for issuing a license. He wants to maximize his bribe revenue $B(\lambda) = \lambda b(\lambda)$, so that he will issue $\lambda_L = -b(\lambda)/b'(\lambda)$ licenses. Notice that $\lambda_L < \lambda_H$, which means that when there exist market barriers (an imperfectly competitive market), corruption will be generated.

Based on the above mechanism, some researchers proposed to increase competition in the provision of public services to fight corrupt behavior. Aidt (2003) took another example to examine the argument. Assume that two different licenses are needed for two different economic activities. Let $b_i(\lambda_1, \lambda_2)$ be the price of a license of type $i = 1,2$, λ_1 and λ_2 are the licenses already issued. Similarly as above, assume that $\frac{\partial b_i}{\partial \lambda_i} < 0$. Now there are several conditions here: whether the two licenses are complements or substitutes, and whether the licenses are issued by the same officer.

When they are complements ($\frac{\partial b_i}{\partial \lambda_j} > 0$) and issued by the same officer, he wants to maximize his total revenue $\sum_i \lambda_i\, b_i(\lambda_1, \lambda_2)$. The first-order condition gives $b_i(\lambda_1, \lambda_2) + \sum_i \lambda_i\, b_i'(\lambda_1, \lambda_2) = 0$. We can get $\lambda_{i-optimal1} = \frac{\lambda_i b_i' + \mathrm{b}_1 + \mathrm{b}_2}{b_i'}$.

When the licenses are complements but issued by different officials, the results will be altered. Officer i wants to maximize his bribe revenue $\lambda_i b_i(\lambda_1, \lambda_2)$, which gives the first-order condition $\mathrm{b}_i(\lambda_1, \lambda_2) + \lambda_i b_i'(\lambda_1, \lambda_2) = 0$. We can get $\lambda_{i-optimal2} = -\frac{\mathrm{b}_i}{b_i'}$. Comparing the two results, we can find that $\lambda_{i-optimal1} > \lambda_{i-optimal2}$, which means that the total supply of licenses is reduced if they are issued by two officers. The reason behind this is that the two officers do not take each other's actions into account. The initial assumption that $b'(\lambda) < 0$ means that more licenses in the market will reduce corruption. Therefore, when government services are complementary, a monopolistic provider is preferred as more licenses will be issued and corruption levels will be lower.

When the licenses are substitutes, the conclusion is the opposite – competition in public services reduces corruption and leads to more licenses being generated.

5.2.4. Self-reinforcing corruption

Differently from the theories discussed above, which relate corruption levels to institutions, the self-reinforcing corruption theory emphasizes that history also plays an important role in generating corruption. That is to say, an individual's behavior depends on the number of other people in a society who are corrupt (Aidt, 2003). For example, it is common to see that if a public officer got caught for corruption, many of his colleagues might have also been involved in the case.

Aidt (2003) continues using the tax collection example (described in section 5.2.2) to illustrate the mechanism. Assume that if the tax collector is spotted by a supervisor when he performs the illegal action, he can avoid being fired by transferring his bribe revenue b to the supervisor who is also corrupt. For simplicity, assume the proportions of corrupt supervisors and corrupt tax collectors are the same, $1-\gamma$, set $f = g = w_0 = 0$, and assume that all officers have the potential to be corrupt but face different moral costs c, which are distributed according to a cumulative density function *F(.)*. Therefore, the expected gain of a tax collector is $(1-p)(w+b)+p(1-\gamma)w-c$. The condition of accepting the bribe is $(1-p)(w+b)+p(1-\gamma)w-c \geq w$. After transformation, we can get $c \leq (1-p)b-p\gamma w$. Therefore, the share of corrupt tax collectors is described by the equation $1-\gamma = F[(1-p)b-p\gamma w]$. It is easy to see that the more there are corrupt officers, say, γ is low, the larger is the expected gain from corruption and the lower the chances of losing the job. This, in turn, encourages more officers to be corrupt and is the way in which the self-reinforcing corruption works.

All in all, the four theories presented above answer the question of how corruption is generated from different perspectives. They also provide useful suggestions on fighting corruption. The positive view of corruption in the efficient corruption theory also gives us a new perspective, and to some extent answers the puzzle why corruption cannot be eliminated completely. But other questions come to mind: why is Denmark the most transparent country in the world, but not Somalia? How is corruption measured? We will try answering these questions in the next sections.

5.3. Measurement of corruption

There is no direct method of measuring corruption, of course. However, there are several indirect methods. In practice, researchers usually use two indices of corruption: one is the Corruption Perceptions Index (CPI), and the other comes from the Worldwide Governance Indicators (WGI).

5.3.1. Corruption Perceptions Index (CPI)

CPI, which is a composite index based on a number of perception-based sources, was created by *Transparency International* (TI). The data contains a variety of surveys and assessments, which are used to evaluate how different countries rank in terms of corruption. This research started in 1995, and the sources may vary from year to year. However, considering the annual reports of the last three years (from 2012 to 2014), there are no apparent fluctuations among the ranks and scores of most countries. For example, Denmark gained the first place in 2014, as it did in 2012 and 2013, with the scores of 90 (2012), 91 (2013) and 92 (2014).

According to *Transparency International* (2014), there are 12 data sources utilized to analyze corruption levels in the 2014 report. All the sources are from independent institutions specializing in governance and business climate analysis:

1) *African Development Bank Governance Ratings:* this rating covers 40 African countries and assesses a country's transparency, accountability and corruption in public sector.
2) *Bertelsmann Foundation Sustainable Governance Indicators:* the indicator provides scores for all 41 OECD and EU countries and judges whether officeholders abuse their positions for private interests.
3) *Bertelsmann Foundation Transformation Index:* it covers 129 countries and evaluates whether officeholders who engage in corruption can do so without fear of legal consequences and prosecution.
4) *Economist Intelligence Unit Country Risk Ratings:* it provides risk assessments for 120 countries.
5) *Freedom House Nations in Transit:* it measures democratisation in 29 nations throughout Central Europe and the Newly Independent States.
6) *Global Insight Country Risk Ratings:* it provides a six-factor analysis of the risk environment in 204 countries.
7) *IMD World Competitiveness Yearbook:* There are 60 countries under evaluation; the object is the national environment and its influence on corporate operations.
8) *Political and Economic Risk Consultancy:* It could be regarded as an interview for business people, in terms of corruption from different perspectives and in different institutions. It covers 15 Asian countries plus United States.

9) *Political Risk Services International Country Risk Guide:* it is an assessment of corruption within the political system in 140 countries.

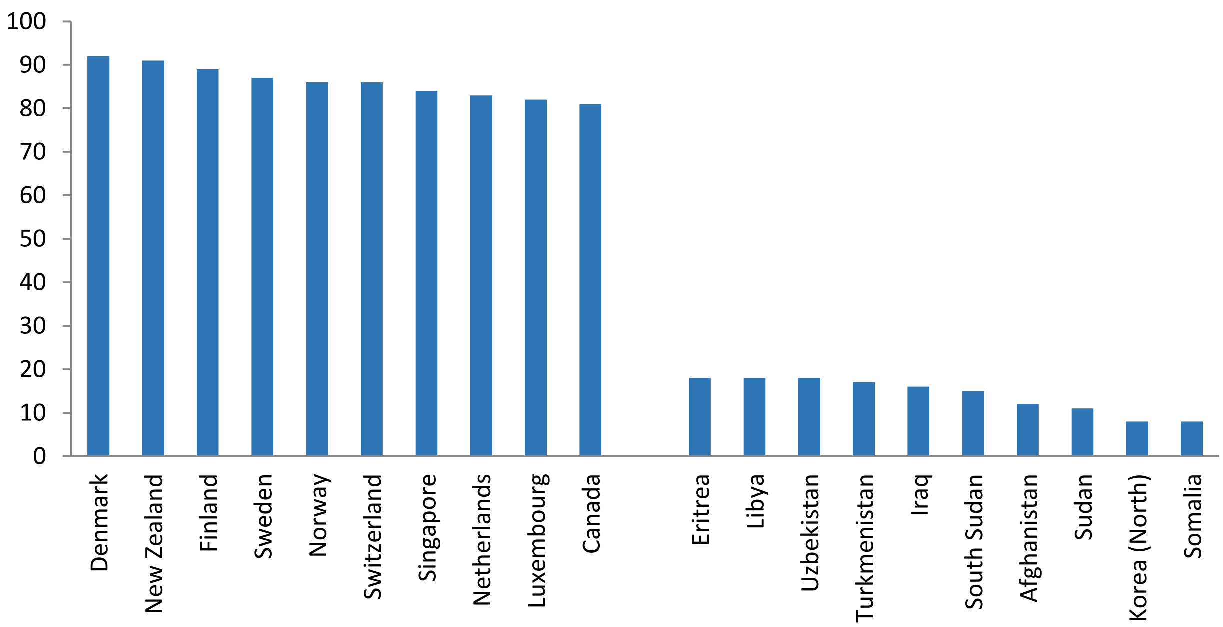
10) *World Bank Country Policy and Institutional Assessment:* the rating reflects a variety of information based on a country's corruption in public sector, and covers 81 countries.

11) *World Economic Forum Executive Opinion Survey:* 143 economies are assessed for the degree of firms giving bribes and diverting public funds to individuals.

12) *World Justice Project Rule of Law Index:* it assesses a nation's adherence to the rule of law in practice, and covers 99 countries.

The figure below shows the results of Corruption Perceptions Index 2014 for the 10 best and 10 worst countries. Denmark is the top scorer, while Somalia scores the worst. In general, countries with advanced economies score much higher than those with undeveloped or developing economies.

Figure 1. Top 10 and bottom 10 countries according to their scores in the Corruption Perceptions Index 2014 (higher scores mean lower corruption)



Source: Transparency International (2014)

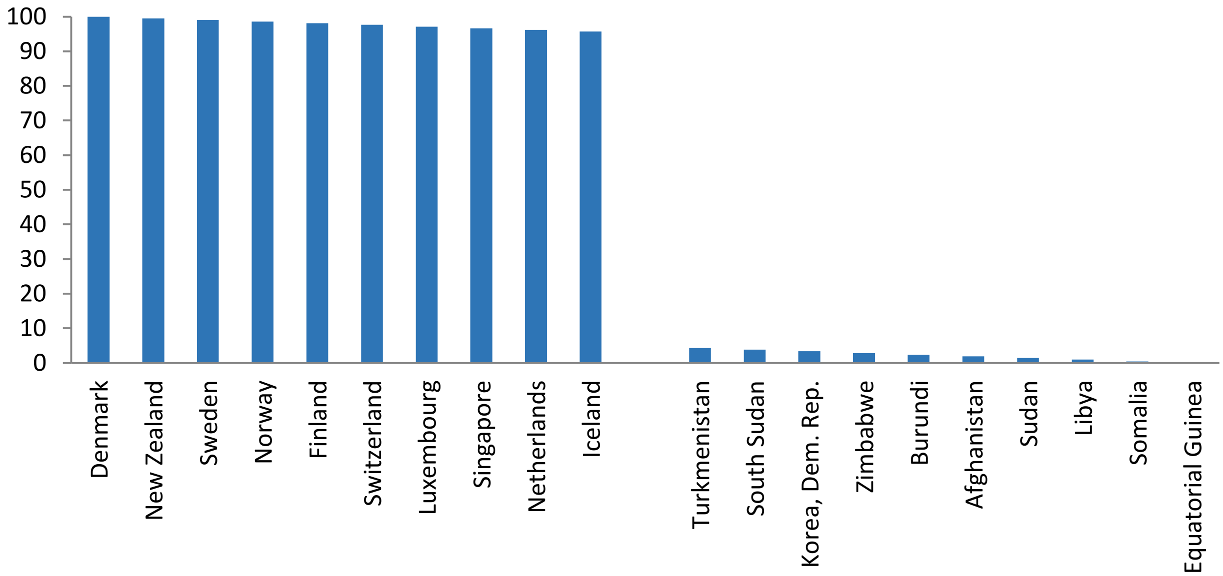
5.3.2. Worldwide Governance Indicators (WGI)

WGI was developed by a team led by Daniel Kaufmann at the World Bank. As Svensson (2005) notes, the World Bank team utilize "*a broader definition of corruption and include most cross-country indices reporting ranking of countries on some aspect of corruption*". According to The World Bank Group (2014), WGI consists of six dimensions of governance: voice and accountability (VA), political stability and absence of violence (PV), government effectiveness (GE), regulatory quality (RQ), rule of law (RL) and control of corruption (CC). WGI uses a different aggregation methodology than does CPI. It has three steps: assigning data from individual sources; preliminarily rescaling of the data; and constructing a weighted average score by using an Unobserved Components Model.

As our research focuses on corruption, we will only use the Control of Corruption (CC) score. CC assesses a country's corruption level and includes 22 different sources. Some of them overlap with CPI sources, for example, *Political Risk Services International Country Risk Guide* and *Bertelsmann Transformation Index*. Therefore, it is not surprising that the two ratings are highly correlated. For example, Denmark gained the first place in 2013 and 2014 in the WGI corruption measure, which is the same result as CPI. Treisman's (2007) research showed that the correlation of WGI and CPI was around 0.96 in 2002 and 0.98 in 2004.

The figure below shows the percentile ranks of countries according to the WGI 2014 Control of Corruption index. Nine out of top ten highest scoring countries are the same across the two different corruption measures, and seven out of ten lowest scoring countries are the same as well. Hence the two indices do indeed seem to be highly related to each other.

Figure 2. Percentile ranks of top 10 and bottom 10 countries according to their WGI 2014 Control of Corruption scores



Source: The World Bank Group (2014)

Having reviewed the theory behind corruption and its two main measurement methods, we next turn to the question of whether corruption has a positive or negative effect on economic development. The next section provides an overview of research performed on this topic.

5.4. The link between corruption and economic development

One of the first and most influential papers linking corruption and economic growth was written by Mauro (1995). The author provides two potential explanations for the relationship: on one hand, corruption may be beneficial for economic growth as it helps people avoid government bureaucracy and encourages government officials to work more productively to be able to require higher bribes; on the other hand, corruption may work as a form of tax and decrease the returns from capital investments, thus discouraging investments and lowering economic growth.

Mauro (1995) uses 9 institutional variables from *Business International*, a consulting company whose institutional assessments are based on analyst opinions and can be thought of as the opinions of foreign investors on risk factors in a particular country. The variables are: Political change, Political stability, Probability of opposition group takeover, Stability of labour, Relationship with neighbouring countries, Terrorism, Legal system, Bureaucracy and red tape, and Corruption. The author finds that all of these variables are highly correlated, so he combines them into separate sub-indices in order to avoid the problem of multicollinearity when running regressions.

Endogeneity is always a concern when it comes to institutional economics: corruption could affect economic growth, while growth could affect corruption as well. Moreover, analysts creating the rankings may be biased as a country's level of economic development could influence their subjective opinions. For this reason, Mauro (1995) uses the instrumental variables approach with the index of ethnolinguistic fractionalization acting as the instrument. The index is equal to $1 - \sum_{i=1}^{I}(\frac{n_i}{N})^2$, where N is total population and n_i is the number of people in a particular ethnic group. Therefore, the index of ethnolinguistic fractionalization measures the probability that two random people in a country will belong to different groups. Higher values of the index are associated with more fragmentation. In order for the index to be a valid instrument, it has to affect economic growth through institutional variables only. The author claims that the condition holds in this case. The index of ethnolinguistic fractionalization is negatively correlated with institutional variables. Mauro (1995) provides several explanations for this fact: first of all, ethnic differences may lead to conflicts and political instability; secondly, the level of corruption could also be higher because bureaucrats may offer protectionism to their ethnic group. The author also uses two additional instrumental variables in several regressions: a dummy variable for countries that have ever been a colony throughout their history and a dummy variable for whether the countries were still a colony in 1945.

Simple ordinary least squares regressions of the share of investment in GDP in 1980-1985 on the index of corruption for a sample of 67 countries show that there is a statistically significant negative relationship between the two variables. If the index of corruption in a country improves by one standard deviation, the share of investment in GDP grows by 2.9% of GDP. Moreover, the coefficients are similar in both low bureaucracy and high bureaucracy countries, so the hypothesis that corruption can be beneficial in the case of high bureaucracy is not approved. Mauro (1995) gets the same results when using instrumental variables and controlling for other factors as well, which provides evidence that corruption is one of the causes of low investment.

Mauro (1995) also finds a significant relationship between the index of corruption and economic growth in 1960-1985, even after controlling for other variables from the neoclassical growth model (population growth and education). If the index of corruption in a country improves by one standard deviation, the annual growth rate of GDP per capita increases by 0.8 percentage points. Using instrumental variables weakens these results, however. Nevertheless, the author concludes that corruption is an important factor that affects economic growth through investment rates.

Svensson (2005) updates Mauro's (1995) model by running the regressions on more recent data – over the period 1980 to 2000. The author tests several specifications and estimation methods, but in all of the regressions the coefficient on corruption is found to be statistically insignificant, even though it is negative. This shows that the results of Mauro (1995) only hold over the original time period, so the model specification may be wrong.

Mendez and Sepulveda (2006) analyse the relationship between economic growth and corruption in more detail. In particular, they look at the effect of political regimes and the size of the government on this relationship, as well as the fact whether a certain level of corruption can be beneficial to the society. For macroeconomic measurement, the authors use data from the *World Bank World Development Indicators* for the period 1960-2000. For measuring corruption, they use three different variables: index of corruption from the *International Country Risk Guide* published by *Political Risk Services Inc*., which reflects the view of foreign investors on corruption levels in a country; index of corruption from the *Institute for*

Management Development which reflects the view of local managers in a country; and the Corruption Perceptions Index from *Transparency International* which is based on many different surveys. In all the regressions that the authors run, they use the average levels of corruption in each country during the time periods when data for them is available. Finally, Mendez and Sepulveda (2006) measure political regimes on the basis whether the countries are free or not according to the index of freedom compiled by *Freedom House International*.

After running different cross-country ordinary least squares regressions of average GDP per capita growth in 1960-2000 on such variables as corruption, initial GDP per capita in 1960, rate of growth of population, secondary school enrolment ratio, the share of investment in GDP, government consumption expenditure as percentage of GDP, political stability and dummy variables for different geographical regions, Mendez and Sepulveda (2006) find that for most of the regression specifications, the coefficient on corruption is statistically significant and does not depend on the specific corruption variable used: lower levels of corruption are associated with higher GDP growth rates. The authors also test a non-linear model with the square of the corruption index added to the regressions on two different samples: free and not free countries. For the free countries, the authors find that there is a small but positive level of corruption that maximizes growth. In the case of not free countries, no statistically significant relationship can be established between economic growth and corruption levels.

Mendez and Sepulveda (2006) also address the issue of endogeneity, but in contrast to the usual approach of using instrumental variables, the authors use fixed-effects estimation by using five-year averages of variables. Results of the fixed-effects estimation for the sample of free countries once again show that a small positive level of corruption is growth-maximizing, while the results for the sample of not free countries are not statistically significant. The authors provide several potential explanations for the lack of any relationship between corruption and growth in not free countries: corruption indices in these countries could have high measurement errors or there could be fewer incentives to bribe officials in autocratic countries as the results of bribing are uncertain and its returns are low. Mendez and Sepulveda (2006) also run additional fixed-effects regressions on the sample of free countries and include government consumption expenditure as a percentage of GDP and the interaction of this variable and corruption as additional explanatory variables. This interaction variable is found to be statistically insignificant, which means that the size of the government does not have an effect on the optimal level of corruption. This fact therefore contradicts the theories which claim that corruption is beneficial in countries with large public sectors as it helps to overcome bureaucracy, or that corruption hurts economic growth more in countries with large governments as they become less efficient. Having this in mind, the finding that a small positive level of corruption is optimal for economic growth can be explained by the fact that the lower the corruption is, the more difficult it is to fight it and the more resources have to be spent for this, thus there exists a certain level of corruption where fighting it is no longer beneficial.

Another study, written by Mo (2001), analyses the channels through which corruption's effects are transmitted to economic growth. The author starts with a production function of the form $Y = Tf(K, L)$, where Y is total output, T is total factor productivity, K is the stock of capital and L is the stock of labour. This equation can be manipulated and transformed into a growth equation of the form $GR = F(\gamma, IY, dLL)$, where GR is the real GDP growth rate, γ is the growth rate of total factor productivity, IY is the share of investment in GDP and dLL is the rate of growth of the labour stock. γ itself is assumed to be a function of corruption, the initial level of per capita GDP and human capital.

Mo (2001) uses data for 46 countries for the period 1970-1985 to run ordinary least squares regressions where corruption data comes from *Transparency International*. The labour growth rate is measured by the growth rate of the total population, the human capital stock is measured by the average school years of people over 25 years old, political instability is measured by the number of assassinations per one million people and the number of revolutions, and other institutional characteristics are measured by the Gastil index of political rights. A regression of GDP growth rate on the corruption index, initial level of GDP per capita, the index of political rights and its squared value, as well as population growth shows that a rise of one unit in the corruption index is associated with a fall in GDP growth rate of 0.545 percentage points.

Expressed alternatively, the GDP growth rate is reduced by 0.72% per year with a one percentage rise in the corruption index. Running the same regressions with the share of investment in GDP, the level of human capital and political stability as dependent variables, Mo (2001) finds that corruption has a statistically significant negative effect on all of the above variables.

Moreover, the author decomposes the total effect of corruption on GDP growth into three components and finds that corruption’s effect on investment accounts for about 28% of the total effect, corruption’s effect on human capital accounts for about 9.7% and its effect on political stability accounts for about 64% of the total effect, when these components are analysed separately. When they are taken together, corruption’s effect on GDP growth can be decomposed into a 11.8% direct effect, 21.4% effect through investment, 14.8% effect through human capital and 53% effect through political stability. The results therefore show that corruption has a significant negative effect on economic growth and that the most important transmission channel of this effect is political stability. Mo (2001) argues that corruption increases inequality among different groups of a society and creates incentives for those hurt the most to fight back violently. This violence reduces political stability and leads to lower investment levels, and this reduces economic growth as a result.

An interesting and comprehensive approach to studying the relationship between corruption and economic growth is employed by Aidt, Dutta and Sena (2008). The difference in their approach comes from using a non-linear threshold model to estimate the relationship in countries with different governance regimes. The model assumes that the economy consists of a formal and an informal sector. People in the formal sector work either in the private or public sector, both of which have competitive labour markets. Public services are financed by a fee which has to be paid by everyone employed in the formal sector. There is no fee in the informal sector, but income levels here are lower than in the formal one. The larger the fee or the lower the amount of public services supplied, the more people switch to the informal sector. Provision of public services and collection of fees are performed by the ruler. In this context, the fee acts as a form of corruption: people pay a higher than necessary fee to work in the formal sector.

In the model of Aidt et al. (2008), political institutions limit the behaviour of a ruler through actions such as elections or revolutions that replace the ruler if the fee is too high. The authors specify two governance failures that prevent the ruler from re-election when the fee is acceptable to the citizens: a p-failure occurs when people can only re-elect the ruler with probability p, while a q-failure occurs when people can only replace the ruler with probability 1-q. Ideally, p should equal 1 and q should equal 0, but this is not possible in reality. Aidt et al. (2008) proceed to specify two governance regimes: regime G with good institutions and $p > q$ and regime B with bad institutions and $p \leq q$. In regime G, the ruler extracts a constant share of national income since he is threatened by re-election, while institutional improvements and economic growth lead to lower fees and corruption. In this regime, economic growth and corruption are interdependent: high growth leads to lower fees (lower corruption), which encourage more citizens to work in the formal sector and this therefore increases growth in the next period. In regime B, the ruler extracts the maximum amount from citizen fees since he is not threatened by re-election and citizens have to move to the informal sector. In this regime, institutional improvements and economic growth do not affect the size of the fee and the level of corruption, unless they are substantial and lead to a change of the regime.

Aidt et al. (2008) test their theoretical predictions empirically using a threshold growth model on a sample of up to 71 countries: they regress GDP per capita growth on measures of corruption, quality of institutions and a vector of other variables as well as a governance regime threshold parameter. GDP per capita growth is measured in the short term (1995 to 2000) and long term (1970 to 2000); corruption data comes from *Transparency International* (1996-2002 average) and *World Bank's World Governance Indicators* (1996-2002 average); the quality of institutions is measured by the Voice and accountability index from *World Bank's World Governance Indicators*; and other variables affecting GDP growth are the proportion of investment in GDP, growth of population, enrolment in primary education, initial level of GDP, and the country's law system (common or civil law). Moreover, the index of ethnolinguistic fractionalization from Mauro (1995) and the number of democratic rule years since 2000 are used as instrumental variables. The authors' empirical findings confirm their theoretical predictions: in countries with good institutions, corruption has a negative and statistically significant effect on economic growth, while in countries with bad institutions the effect is not statistically significant. These results hold for both the short term and the long term. In particular, the authors find that in countries with good governance, a decrease of one point in the corruption perception index leads to an improvement of short-run growth rates by about 0.5 percentage points and of long-run growth rates of about 0.4 percentage points. Overall, the regression results confirm the hypothesis that in countries with good institutions, high corruption leads more people to switch to the informal sector and this diminishes economic growth, while in countries with bad institutions small improvements in corruption levels do not make the citizens switch back to the formal sector.

The link between corruption, governance and economic growth is analysed by Meon and Sekkat (2005) as well. First of all, the authors provide an overview of two different views of the effect of corruption on economic growth in countries with bad governance: the "grease the wheels" and the "sand the wheels" hypotheses. According to the "grease the wheels" hypothesis, corruption can be beneficial when governance is of low quality. For example, it can decrease the time that citizens spend in queues waiting for decisions of government officials as a bribe will provide incentives for these officials to work faster. Corruption could also help avoid different inefficient regulations. Bribes can also serve as a form of additional wages for the officials and hence motivate them to improve the quality of their services. Corruption can also increase the efficiency of contract allocation in auctions since a company willing to pay the largest bribe is more likely to be the most efficient. Furthermore, corruption can lead to lower government revenue and hence decrease the level of inefficient public investment. On the other hand, the "sand the wheels" hypothesis states that corruption is detrimental to economic growth when governance is bad. In the case of speed of work of government officials, this hypothesis argues that corruption may lead the officials to slow down their work on purpose in order to get a bribe. When it comes to quality of public services, the "sand the wheels" point of view says that the quality will fall as corrupt officials will have a tendency to demand bribes whenever they can. Moreover, contract allocation may become more inefficient due to corruption as the company that pays the highest bribe may end up providing services of bad quality. Finally, there is evidence that corruption leads to increases in inefficient government investment, instead of reducing it.

Meon and Sekkat (2005) employ the following econometric model to test the relationship between corruption and growth in countries with bad governance: $\log(y_T) - \log(y_0) = \alpha_0 + \alpha_1 * \log(y_0) + \alpha_2 * \log(Sc_0) + \alpha_3 * [\log(pop_T) - \log(pop_0)] + \alpha_4 * \log(inv) + \alpha_5 * \log(open) + [\alpha_6 + \alpha_7 * \log(gov)] * log(cor) + \mu$, where the left-hand side is the average GDP per capita growth rate in 1970-1998, $\log(y_0)$ is initial GDP per capita, $\log(Sc_0)$ is the initial schooling level, $\log(pop_T) - \log(pop_0)$ is the average population growth over the period, $\log(inv)$ is the average share of investment in GDP, $\log(open)$ measures the economy's openness to trade, $\log(gov)$ measures the governance quality and $log(cor)$ measures corruption. They also test the effect of corruption on investment and run a similar regression where the share of investment in GDP is the dependent variable and population growth as well as investment do not appear on the right-hand side. The level of schooling is measured by enrolment in primary and secondary education, openness to trade – by the share of exports and imports in GDP, governance quality – by indices from the *World Bank's World Governance Indicators*, namely, Voice and accountability, Lack of political violence, Government effectiveness, Regulatory burden, and Rule of law. Corruption data comes from two sources: *Transparency International* and the *World Bank*, similarly to other research in this area. The data covers up to 71 developed and developing countries.

The most important parameters in this regression are α_6 and α_7. According to the "grease the wheels" hypothesis, α_6 should be greater than 0 and α_7 should be less than 0; according to the "sand the wheels" hypothesis, α_6 should be less than 0 and α_7 should be greater than 0. In a simple regression without the interaction term of governance and corruption, Meon and Sekkat (2005) find that Regulatory burden and Voice and accountability indices are not significant in the growth equations, thus they have no effect on the relationship of corruption and economic growth. The remaining three governance indicators are statistically significant. The coefficient on corruption is insignificant when *Transparency International* data is used, but becomes significant and negative when *World Bank* data is used. This serves as a proof that corruption slows down economic growth. Moreover, since the investment ratio is included in the regression, the results show that corruption has a direct effect on growth apart from its effect on investment. In a regression with the interaction term, the coefficient on corruption (α_6) is negative and significant for both of its measures as well as larger than before, while the interaction term (α_7) is positive and statistically significant. The results therefore support the "sand the wheels" hypothesis: corruption and bad governance slow down economic growth. The same results hold when the investment ratio is used instead of economic growth as the dependent variable: corruption has a larger negative effect on investment in countries with bad governance. The results of Meon and Sekkat (2005) contradict those of Aidt et al. (2008), who found that corruption does not affect growth in countries with bad governance. Nevertheless, both articles found a negative effect of corruption on economic development – only the groups of countries where this effect holds were found to be different.

All in all, while theoretical models describe both a positive and a negative effect of corruption on economic growth, the empirical evidence from different researchers points to the fact that corruption hinders economic growth. Even though transmission mechanisms and governance regimes where this relationship holds are found to be different across the academic papers, the general consensus in modern economic analysis supports the view of the negative effect of corruption on economic development. For this reason, more and more organizations and policies are created with the goal of fighting corruption.

PART II: RESEARCH

6. EU FUNDING

Before we proceed to analyse the effects of corruption on the efficiency of adoption of EU funding and the funding's effects on economic growth in new EU-member countries, it is necessary to provide an overview of how EU funding works. The following sections do exactly this.

6.1. Regional Policy of the European Union

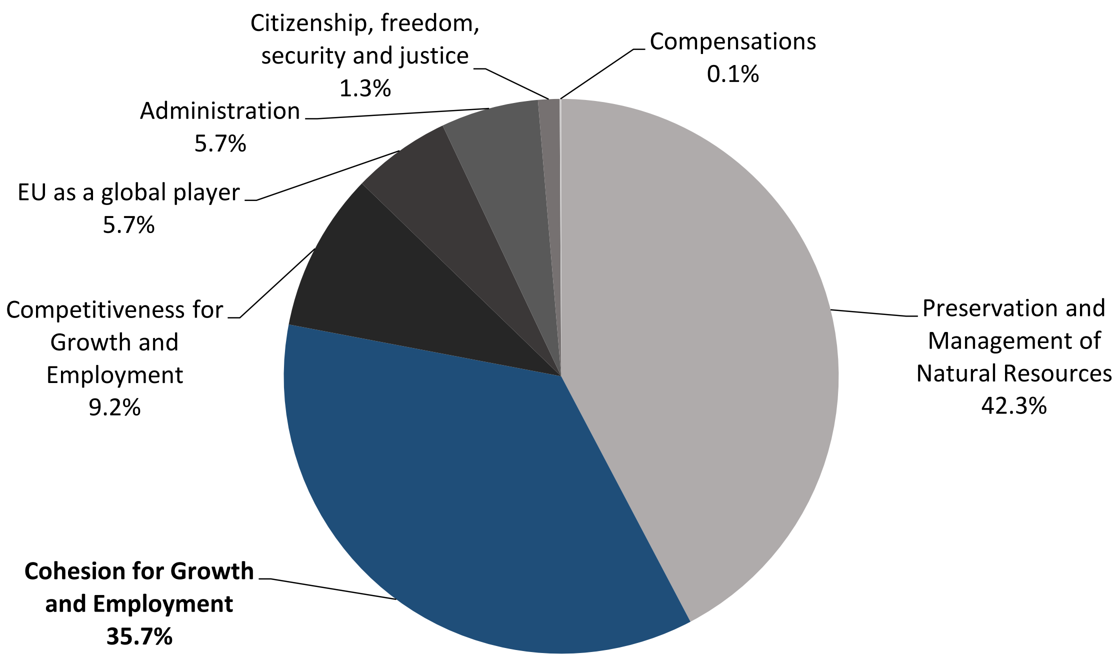
According to European Commission (2015), European Structural and Investment Policy is composed of five funds:

- European Agricultural Fund for Rural Development (EAFRD)
- European Maritime and Fisheries Fund (EMFF)
- European Regional Development Fund (ERDF)
- Cohesion Fund (CF)
- European Social Fund (ESF)

EAFRD and EMFF are the two investment funds, ERDF and ESF are the two structural funds, while ERDF, CF and ESF together make up the EU Regional Policy (also called Cohesion Policy), whose main goal is to reduce economic inequalities between EU members in the form of transfers from more developed to less developed countries. ERDF is focused on investments into employment-generating activities and infrastructure, ESF is focused on providing funding for the disabled and unemployed, while CF provides funding for transport and environment infrastructure projects for countries with Gross National Income per capita levels which are below 90% of the EU average (European Commission, 2015).

Expenditure on Regional Policy in the 2007-2013 EU budget amounted to 348 billion Euros. As can be seen from the figure below, it was the second largest expenditure category and represented almost 36% of the whole budget expenditure. Only Preservation and Management of Natural Resources category received more funding (42% of total expenditure). According to European Commission (2015), among the 348 billion Euros dedicated for Regional Policy, the two structural funds (ERDF and ESF) received 278 billion and the Cohesion Fund received 70 billion.

Figure 3. Structure of the EU budget expenditures by category for the period 2007-2013



Source: European Commission (2015)

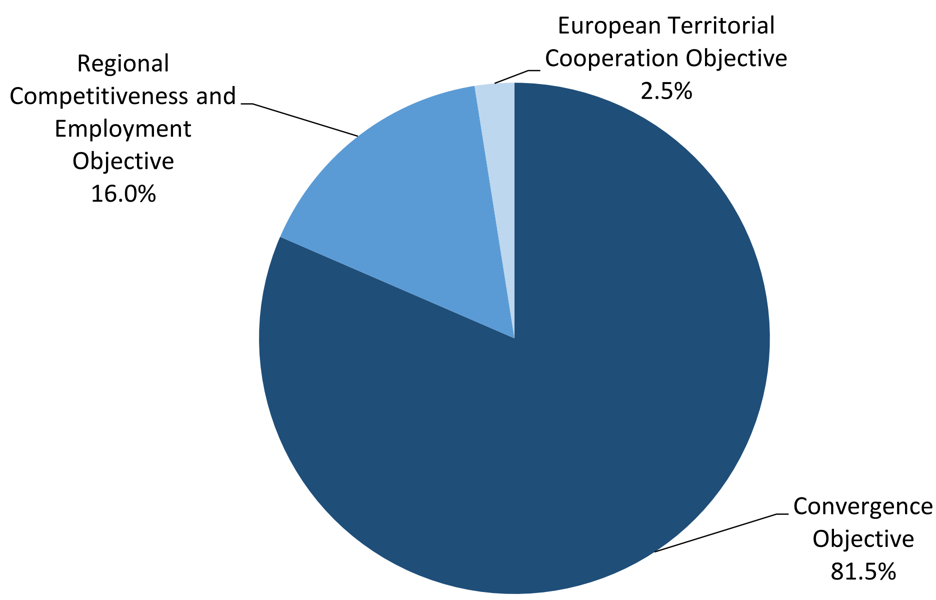
According to Varga and Veld (2011), EU Regional Policy has three objectives:

- "Convergence Objective" aims to stimulate growth-enhancing activities in less economically-developed EU countries so that they can catch up with more advanced EU economies;
- "Regional Competitiveness and Employment Objective" is focused on promoting innovation, employment, entrepreneurship and competitiveness in regions which do not receive funding under the "Convergence Objective";

- “European Territorial Cooperation Objective” is focused on strengthening cooperation between European regions and countries in terms of economic development as well as cooperation among SMEs.

As the figure below demonstrates, among these objectives the “Convergence Objective” was allocated by far the most funds in the 2007-2013 period – 81.5%, while the “Regional Competitiveness and Employment Objective” received 16% and the “European Territorial Cooperation Objective” received 2.5%. Convergence objective payments from the structural funds (ERDF and ESF) are allocated to regions with per capita GDP in purchasing power parity less than 75% of the EU average, while payments from the Cohesion Fund are allocated to countries with GNI per capita in purchasing power parity less than 90% of the EU average.

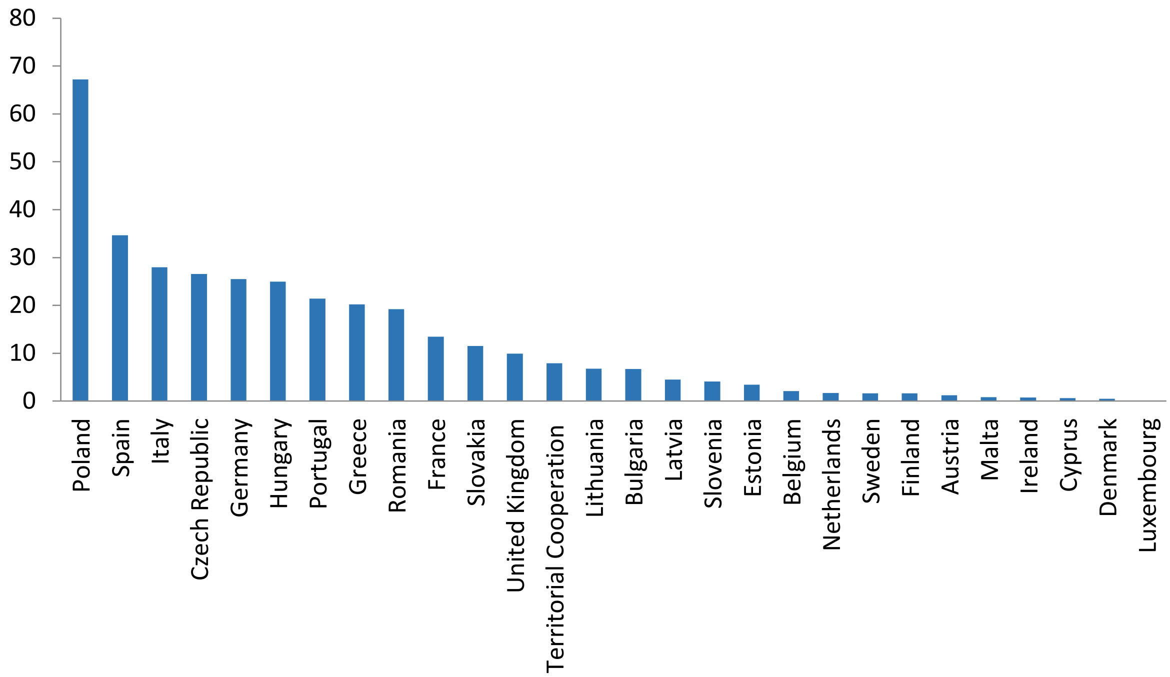
Figure 4. Distribution of expenditure on EU Regional Policy by objective for the period 2007-2013



Source: European Commission (2015)

The figure below shows that on a per country basis, Poland was the main beneficiary of Regional Policy and received 67 billion Euros in funding. The twelve new EU member countries which joined in 2004 and 2007 received 176 billion Euros, which is more than a half of the total funds allocated for the Cohesion Policy.

Figure 5. EU Regional Policy allocations per country for the period 2007-2013, billion Euros



Source: European Commission (2015)

6.2. How EU funds affect economic growth of receiving countries

European Commission (2001) states that *"cohesion policies are aimed at increasing investment to achieve higher growth and are not specifically concerned either with expanding consumption directly or with redistribution of income"*. Hence, investment is the main channel through which EU cohesion funds affect economic growth in less-developed countries. Apart from infrastructure investment, these funds also promote research & development and education, thus increasing technological progress. In the context of the neo-classical growth model, this means that successful EU Regional Policy leads to a rise in the steady-state amounts of physical and human capital stocks as well as improvements in total factor productivity.

However, Ederveen, Groot and Nahuis (2006) argue that it may not always be true: certain projects to which the funds are allocated can be focused on other issues than fostering growth (e.g., cultural beliefs); these funds can also lead to human capital being used on less beneficial projects than the ones that are planned outside of the policy; moreover, these projects are co-funded by receiving countries, so taxation used for this purpose can actually lead to an overall negative effect of the projects; finally, inefficient institutions and corruption can significantly decrease the positive returns from such projects.

In the case of actual Cohesion Policy allocations, European Commission (2015) provides the following distribution data for the 2007-2013 budget period by theme (% of total):

- *Innovation & RTD:* 15.2%
- *Environment:* 12.8%
- *Road:* 11.9%
- *Human capital:* 8.6%
- *Labour market:* 7.2%
- *Rail:* 6.9%
- *Other SME and business support:* 6.8%
- *Social infrastructure:* 5.1%
- *Other transport:* 4.7%
- *IT services and infrastructure:* 4.2%
- *TA & Capacity Building:* 3.7%
- *Energy:* 3.4%
- *Social inclusion:* 3.4%
- *Urban and territorial dimension:* 3.2%
- *Culture, heritage and tourism:* 2.8%

As can be seen, most of the funding is indeed allocated to innovation, infrastructure and human capital development.

7. METHODOLOGY

7.1. Review of regression estimation techniques

We start the methodology overview with introducing the general concept of regression analysis, its application to cross-sectional and panel data and, finally, we discuss the more advanced methods of model estimation, namely, Generalized Method of Moments (GMM).

7.1.1. Single-equation linear regression with cross-sectional data

Studenmund (2006) defines regression analysis as *"a statistical technique that attempts to explain movements in one variable, the dependent variable, as a function of movements in a set of other variables, called the independent (or explanatory) variables, through the quantification of a single equation*". Wooldrigde (2002) describes the basic single-equation linear regression model as having the following form: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$, where β's are the parameters to be estimated, y and x's are observable random scalars and u is the unobservable error term. In order for the parameters to be consistently estimated, the error term must have a zero mean and be uncorrelated with independent variables (x's), that is: $E(u) = 0$ and $Cov(x_j, u) = 0$. The mean of the error term is forced to be zero by including the constant term in the model - β_0. On the other hand, if at least one of the independent variables is correlated with the error term, the model suffers from the problem of endogeneity. Endogeneity can occur when one or several variables are omitted from the model due to unavailability of data or model misspecification, when a variable is measured imperfectly, and when the dependent and one or several independent variables are determined simultaneously, that is, are interdependent.

The standard and most commonly used method of estimating regression coefficients is Ordinary Least Squares (OLS). OLS estimation calculates the regression coefficients by minimizing the sum of squared residuals $\sum_{i=1}^{n} e_i^2$, where e_i are the differences between actual observations of the dependent variable (y) and estimates of the dependent variable obtained from the regression model for different observations (i = 1,2,…, n).

Studenmund (2006) lists the conditions that must be satisfied by the model in order for OLS to be the best estimator:

- The model must be linear in coefficients, correctly specified and have an additive error term.
- The error term must have a mean of zero.
- All independent variables must be uncorrelated with the error term.
- Different error term observations must not be correlated with each other; if they are, standard error estimates of the coefficients will not be reliable.
- The error term must have a constant variance, that is, the error term observations must come from identical distributions. If this assumption does not hold, the model will suffer from heteroscedasticity and standard errors of the coefficients will be unreliable once again.
- No independent variables can be perfect linear functions of other independent variables. If they are, the model suffers from multicollinearity.
- The error term must have a normal distribution in order for hypothesis testing to be reliably applied.

Given the first six assumptions, the Gauss-Markov Theorem states that OLS produces the estimator with the minimum variance of coefficients among all linear unbiased estimators. These coefficient estimates have the following properties, according to Studenmund (2006):

- They are unbiased: $E(\hat{\beta}) = \beta$. That is, the estimated coefficients have a mean equal to the real population parameters.
- The distribution of estimated coefficients around the real population parameters has a minimum variance.
- The estimates are consistent: as the sample size increases towards infinity, the estimates approach the real population parameters.

- The estimates are normally distributed if the error term has a normal distribution as well.

7.1.2. Panel data regressions

When observations for each cross-sectional unit are available across a set of time periods, the dataset becomes a panel dataset. Generally, more cross-section observations than time-series observations are required for the standard panel data estimation properties to hold. Before we introduce panel data estimation techniques, it is useful to rewrite the simple linear OLS model in matrix form as this representation is used in panel data analysis.

A model with N observations and K independent variables is described by the following matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{K1} \\ 1 & X_{12} & X_{22} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{1N} & X_{2N} & \cdots & X_{KN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

This equation can be written as $y = X\beta + u$, where y is a $N \times 1$ vector of dependent variable observations, X is a $N \times K$ matrix of independent variable observations (the first column is composed of 1's because we include a constant in the regression model), β is a $K \times 1$ vector of parameters to be estimated and u is a $N \times 1$ vector of error terms.

When describing panel data, Wooldridge (2002) assumes that independent, identically distributed cross-section observations are available: (X_i, y_i) for i = 1, 2, …, N, where X_i is a $T \times K$ matrix for T time periods and y_i is a $T \times 1$ vector. The model for one random draw from the whole population then takes the following form: $y_i = X_i\beta + u_i$, where β is once again a $K \times 1$ vector of parameters to be estimated and u_i is a $T \times 1$ vector of error terms. General panel data models are estimated by Pooled Ordinary Least Squares (POLS).

The first assumption necessary for consistent estimation of parameters with POLS is $E(X_i'u_i) = E(\sum_{t=1}^{T} x_{it}'u_{it}) = 0$. A sufficient condition for this to hold is $E(x_{it}'u_{it}) = 0$. Therefore, only observations of independent variables and error terms at the same time period are assumed to be uncorrelated. They can still be correlated when the time period for observations of x's is different from that of u's. The parameter vector β can be shown to equal $\beta = [E(X_i'X_i)]^{-1}E(X_i'y_i)$. In order for this equation to hold, $E(X_i'X_i)$ must be invertible and have a rank (number of linearly independent rows or columns) of K. This ensures that no independent variables are perfect linear functions of other independent variables. This is the second assumption necessary for identification of the consistent parameter vector β. The estimate of parameters by POLS, $\hat{\beta}$, is then found as $\hat{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}'x_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}'y_{it}\right)$. These two assumptions also guarantee that residuals are asymptotically normal: as the number of observations approaches infinity, the distribution of residuals approaches normal distribution. Adding homoskedasticity and no serial correlation assumptions leads to reliable standard error estimates: $E(u_t^2|x_t) = \sigma^2$ and $E(u_tu_s|x_t, x_s) = 0$.

The assumption of no correlation between independent variables and the error term in the same time period does not always hold. For example, there may exist an unobserved effect (c_i) that is constant over time and correlated with independent variables: $y_{it} = x_{it}\beta + c_i + u_{it}$. Fixed effect estimation takes this fact into account. Under the assumption of strict exogeneity ($E(u_{it}|x_i, c_i) = 0$), fixed effects estimation transforms the initial regression model by averaging the variables across time and thus eliminating the time-invariant unobserved effect (c_i). Subtracting these averaged values from the initial equation gives the following model: $\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it}$, where $\ddot{y}_{it} = y_{it} - \bar{y}_{it}$ and so on. Similarly as in previous models, the second assumption is that the rank of $E\left(\ddot{X}_i'\ddot{X}_i\right)$ is K. The usual assumptions of homoskedastic errors with respect to independent variables and the unobserved effect as well as no serial correlation in errors make the fixed effect estimator efficient. The disadvantage of fixed effects estimation is that the regression cannot contain any independent variables that are constant over time (e.g., gender). If they are included, fixed effects estimation will fail.

Another estimation technique is used when, contrary to the fixed effect case, the unobserved effect is assumed to be uncorrelated with independent variables. In this case, the standard errors from POLS will be unreliable, unless robust statistics are used. However, the usual technique used for such cases relies on random effects estimation. Wooldridge (2002) lists the following assumptions which are necessary for the random effects estimator: strict exogeneity – $E(u_{it}|x_i, c_i) = 0$; no correlation between the independent variables and the unobserved effect – $E(c_i|x_i) = E(c_i) = 0$; and the usual rank condition to prevent multicollinearity. The random effects method then uses Generalized Least Squares (GLS) to estimate the regression parameters.

7.1.3. Generalized Method of Moments (GMM) estimation

When a regression model contains endogenous variables, that is, the first assumption of POLS of no correlation between independent variables and error terms in the same period ($E(x_{it}'u_{it}) = 0$) does not hold, POLS will produce biased estimates. Furthermore, fixed effects estimation will also be inconsistent if there exists at least one independent variable which is correlated with the error term (after taking into account the unobserved effect c_i) and the assumption of strict exogeneity does not hold. A usual approach to solving the endogeneity problem in simple OLS regression is to use instrumental variables. An instrumental variable is a variable which is not included in the initial regression model and which is uncorrelated with the error term as well as partially correlated with the endogenous variable that it tries to correct once the remaining independent variables have been taken into effect. This approach is also used in panel data and is based on GMM estimation.

The first assumption necessary for GMM estimation is $E(Z_i'u_i) = 0$, where Z_i is a $T \times L$ matrix of instruments. The second assumption is that the rank of $E(Z_i'X_i)$ is K. Wooldridge (2002) states that under these two assumptions, the GMM estimator of β solves the following problem:

$$\min_b \left[\sum_{i=1}^{N} Z_i'(y_i - X_i b)\right]' \widehat{W} \left[\sum_{i=1}^{N} Z_i'(y_i - X_i b)\right],$$

where $\widehat{W}$ is a weighting matrix.

Continuing with the case of the unobserved effect model, $y_{it} = x_{it}\beta + c_i + u_{it}$, error terms are now allowed to be correlated with future values of independent variables: $E\left(u_{it}|x_{it}, x_{i,t-1}, x_{i,t-2}, \ldots, x_{i1}, c_i\right) = 0$. Estimation in this case must first remove the unobserved effect and then use instrumental variables. For example, the original regression can be transformed into a first-differenced regression and then lagged first differences of independent variables or simple lags of independent variables can be used as instruments in GMM estimation.

Roodman (2009) reviews the two commonly used GMM estimators in dynamic panels: Arellano-Bond and Arellano-Bover/Blundell-Bond. According to the author, these estimators are used when:

- panel datasets have many cross-sectional observations and few time periods,
- the model is described by a linear relationship,
- the dependent variable depends on its values in previous periods,
- independent variables do not satisfy strict exogeneity conditions,
- fixed effects are used,
- observations within individuals are autocorrelated and heteroskedastic,
- only instruments constructed from the model itself can be used – they are based on lags of existing variables.

The model is described as follows: $y_{it} = \alpha y_{i,t-1} + x_{it}'\beta + \epsilon_{it}$, $\epsilon_{it} = \mu_i + v_{it}$, $E(\mu_i) = E(v_{it}) = E(\mu_i v_{it}) = 0$. The main regression equation can also be written as $\Delta y_{it} = (\alpha - 1)y_{i,t-1} + x_{it}'\beta + \epsilon_{it}$. The first step in estimation is to eliminate the fixed unobserved effect. Both the Arellano-Bond estimator (also called the difference GMM estimator) and the

Arellano-Bover/Blundell-Bond estimator (also called the system GMM estimator) transform the original equation into the first-difference equation: $\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta x'_{it}\beta + \Delta v_{it}$. This transformation solves the fixed effects problem, but it does not solve the endogeneity problem since the error terms in the transformed equation can still be correlated with independent variables. Therefore, the second estimation step uses instrumental variables to eliminate endogeneity.

The Arellano-Bond estimator uses lags of original variables as instruments. Lags from order 2 to the last available ones are used for endogenous variables; for predetermined variables (the variables which are uncorrelated with past and current values of errors, but are correlated with future error values) lags from order 1 and up are used; similarly, for $y_{i,t-1}$, which is also assumed to be predetermined, the instruments start from $y_{i,t-2}$ and up. Roodman (2009) highlights that this estimator assumes that errors are not correlated across individuals, but are correlated within individuals.

The Arellano-Bover/Blundell-Bond approach augments the Arellano-Bond estimator by adding another step which uses differenced instruments in a regression with usual, non-differenced variables. It requires an assumption of no correlation between the fixed unobserved effects and the differenced instruments across all time periods as well as the absolute value of α being less than 1. The important feature of the Arellano-Bover/Blundell-Bond estimation technique is that it allows the inclusion of variables which are constant over time (e.g., gender) whereas the usual fixed effects or the differenced GMM estimator eliminate the effect of such variables. It also improves on the Arellano-Bond technique as the latter becomes biased when there are not enough time series observations available or the autocorrelation of y is high (large α).

7.2. Commonly used methods to study the effects of EU structural funds on growth

Hagen and Mohl (2011) provide a useful overview of econometric research about the effects of EU Cohesion Policy on convergence and economic growth. The authors note that most of the research on this topic has been based on the neoclassical growth theory, and while econometric evaluation was performed on cross-sectional data at first, more recently researchers usually use panel data. These econometric models are developed by starting with the basic neoclassical regression and then augmenting it with additional relevant variables.

According to Hagen and Mohl (2011), the standard neoclassical growth regression is based on the concept of β-convergence. The main assumption behind it is that countries with similar economies converge to the same balanced growth path and steady-state income levels. In this context, β-convergence means that less economically developed countries will have larger growth levels than countries with high initial income. The following regression is used to test this assumption: $\ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha + \beta * \ln(y_{i,t-1}) + u_{i,t}$, where y is the level of income and u is the error term. A negative β in this regression acts as a proof of β-convergence and this relationship has been found to hold empirically.

Based on the above regression, researchers usually use a model of the following form to analyse the effects of cohesion policy on economic growth: $\ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha + \beta_1 * \ln(y_{i,t-1}) + \beta_2 * \ln(sf_{i,t-1}) + \beta_3 * \ln(sav_{i,t-1}) + \beta_4 * (n_{i,t-1} + g + \delta) + \beta_5 * \ln(educ_{i,t-1}) + \mu_i + \lambda_t + u_{i,t}$, where sf are the structural fund payments (corresponding to investments) and are measured as a percentage of nominal GDP, sav is the saving rate, n is population growth, g is the rate of technological progress, δ is the time discount factor, $educ$ is the level of education in the country, μ_i and λ_t are fixed country (or region) and time effects, respectively. As in Mankiw et al. (1992), g and δ are assumed to be constant and sum to 5%. A positive and statistically significant coefficient on the structural fund variable means that these funds increase the rate of growth with which the economy moves towards its steady state.

When institutional characteristics are added to the regression, it usually takes the following form: $\ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha + \beta_1 * \ln(y_{i,t-1}) + \beta_2 * \ln(sf_{i,t-1}) + \beta_3 * inst_{i,t} + \beta_4 *$

$inst_{i,t} * \ln\left(sf_{i,t-1}\right) + \beta_5 * \ln\left(sav_{i,t-1}\right) + \beta_6 * \left(n_{i,t-1} + g + \delta\right) + \beta_7 * \ln\left(educ_{i,t-1}\right) + \mu_i + \lambda_t + u_{i,t}$, where $inst$ is the variable measuring the quality of institutions.

Hagen and Mohl (2011) highlight several methodological issues that are important when analysing the effects of structural funds on economic growth. First of all, endogeneity is a problem. The usual cases of omitted and unobserved variables could be one of the reasons behind endogeneity issues. For example, there may be spillover effects from one neighbouring country to the other arising from the structural fund support – this would lead to an omitted variable and biased estimates. Moreover, endogeneity could occur due to the fact that structural funds themselves are assigned based on economic performance of the region (GDP per capita of the region compared to the EU average) – this means that the relationship between economic growth and structural funds could go both ways. Secondly, even though the neoclassical model is used in most of the influential papers on this topic, this does not mean that the relationship is characterized by this model in reality. The functional form and control variables used may also introduce biased results since they are not known for sure before running the regressions.

In order to avoid these problems, Hagen and Mohl (2011) recommend running the regressions on panel data. Including fixed country or region effects in panel data regressions helps eliminate omitted variables which are constant over time. Similarly, unobserved time effects which are present in all countries (regions), for example, common macroeconomic conditions, can be eliminated by using time effects. Endogeneity problems can be fixed by including instrumental variables in the regression, but Hagen and Mohl (2011) note that no good instrumental variables have been found so far. Using Generalized Method of Moments (GMM) estimation instead of the usual Maximum Likelihood Estimation (MLE) could be another solution to the endogeneity problem. However, GMM requires a large number of observations to be consistent, so country-level data could produce inconsistent estimates. Running the regressions on regional data is not possible in our case since corruption data is only available on a country level, hence we have to treat GMM results with caution. Hagen and Mohl (2011) also overview the results of the most influential papers on the topic. They conclude that in the case of country-level analysis, cohesion policy is found to be effective only in countries with a good quality of institutions. Other cross-country results are inconclusive. Research with data on regional level is also inconclusive: evidence for both a significant effect of structural funds on economic growth and no effect can be found depending on the sample, estimation technique and regression specification. We provide more details on the most notable studies below.

Ederveen et al. (2006) use panel data analysis on a cross-country sample to study the effectiveness of Cohesion Policy. They argue that cross-country data is more suitable for this purpose as it is less likely to have spillover effects than regional data, allows including variables that are not available on a regional level and does not have as large an endogeneity problem as regional data has since structural fund support is assigned based on regional economic performance. Overall, the authors use the following model: $g_{it} = \beta_0 + \beta_1 * \ln(y_{it}) + \beta_2 * \ln(s_{k,it}) + \beta_3 * \ln(s_{h,it}) + \beta_4 * \ln(n_{it} + g_A + \delta) + \beta_5 * SF_{it} + \beta_6 * COND_{it} * SF_{it} + \varepsilon_{it}$, where g_{it} is average annual rate of growth of real GDP per capita, y_{it} is initial GDP per capita, $s_{k,it}$ is the average gross domestic savings rate, $s_{h,it}$ is the human capital accumulation rate, n_{it} is the growth rate of population, g_A is the technological progress rate, δ is the depreciation rate, SF_{it} is the natural logarithm of one plus received ERDF funds as a share of GDP, and $COND_{it}$ measures the quality of institutions. For all countries and points in time, g_A and δ are assumed to sum to 5%. Ederveen et al. (2006) use the data for 13 EU countries from 1960 to 1995, in five year periods, for a total of 7 observations over time.

When the ICRG institutional quality index developed by Sachs and Warner is used as the institutional variable, the coefficient on the Structural Funds variable is negative and statistically significant, while the coefficient on the interaction term is positive and statistically significant. Ederveen et al. (2006) conclude that in countries with good institutions, structural funds foster growth, while in countries with bad institutions they hamper growth. The authors also do robustness checks by running different regression specifications: including trade openness (natural logarithm of the sum of exports and imports as a percentage of GDP), corruption from *Transparency International* and WGI indicators (Political Stability, Government Effectiveness, Rule of Law) as institutional variables confirms the main results. Further robustness checks by including lags of the Structural Fund variable, using country and time fixed effects and using GMM estimation also confirm the original results.

A slightly different approach is employed by Beugelsdijk and Eijffinger (2005). The authors estimate a panel GMM model by using the data of 15 EU countries for the period 1995-2001. While Ederveen et al. (2006) used data over 7 periods, Beugelsdijk and Eijffinger (2005) use annual data for a single 7-year period. Their final model takes the following form: $GDPgrowth_{i,t} = \beta_0 + \beta_1 * GDPgrowth_{i,t-1} + \beta_2 * GDPgrowth_{i,t-2} + \beta_3 * \ln(y_{i,t-1}) + \beta_4 * SF_{i,t-3} + \beta_5 * Corruption\, SFRate_{i,t-3} + \beta_6 * CorruptionIndex_{i,t} + u_{i,t}$, where y is GDP per capita, SF is the amount of structural funds received as a share of GDP and $CorruptionIndex$ is the Corruption Perceptions Index (higher values mean lower corruption). The interaction term of corruption and structural fund variables is added in order to assess the different effect of structural funds for different levels of corruption. Beugelsdijk and Eijffinger (2005) expect corruption to weaken the relationship: in countries with higher corruption, structural funds should have a weaker effect on GDP growth. This means that both β_5 and β_6 should be positive.

In a basic model without corruption and the interaction term, Beugelsdijk and Eijffinger (2005) find that structural funds have a statistically significant and positive effect on GDP growth three years ahead (a three-period lag is used for the structural fund variable): an increase of structural funds as a percentage of GDP by one percentage point is associated with an increase in annual GDP growth by 0.32 percentage points. However, the final regression specification has statistically insignificant coefficients on the corruption variable and the interaction term. Therefore, contrary to the results of Ederveen et al. (2006), Beugelsdijk and Eijffinger (2005) find that there is no reason to believe that corruption negatively affects the relationship between structural funds and economic growth.

The effects of structural funds on economic growth on a regional level are studied by Mohl and Hagen (2010). Their panel data covers 124 EU regions annually during the period 1995-2005.

The authors also differentiate between different objectives (Objective 1, 2 and 3) and run regressions on the total amounts of support and on each of the objectives separately. These objectives, however, are no longer the same in the 2007-2013 period and further on. The final model used by Mohl and Hagen (2010) is: $\ln(y_{i,t}) - \ln(y_{i,t-1}) = \beta_0 + \beta_1 * \ln(y_{i,t-1}) + \beta_2 * \ln(inv_{i,t-1}) + \beta_3 * (n_{i,t-1} + g + \delta) + \beta_4 * \ln(innov_{i,t-1}) + \beta_5 * \ln(sf_{i,t-1}) + \mu_i + \lambda_t + u_{i,t}$, where y is real GDP per capita, inv is gross fixed capital formation (investment) as a percentage of nominal GDP, and $innov$ is a proxy for the level of education on a regional level as this data is not available and is measured as the number of patents per million people. All the other variables are the same as in previously presented regressions, and sf is the amount of structural funds as a percentage of nominal GDP. The authors also test a specification with time lags of the structural fund variable, ranging from one and up to five lags, in case the funds have a lagged effect on economic growth. Across different specifications of the model and different estimation techniques, Mohl and Hagen (2010) find that the total sum of structural funds has mostly positive effects on economic growth; however, the results are not always statistically significant. In contrast, the effect of Objective 1 payments is found to be consistently positive and statistically significant: a one percentage increase in Objective 1 funds received is associated with an increase of GDP growth of 0.5% annually. Time lags are also found to have an effect on results: across the specifications, lags of up to four years are statistically significant.

Another regional study is performed by Pinho, Varum and Antunes (2014). The panel data in their dataset covers 12 EU countries with annual data from 1995 to 2009. The basic model they use is:

$$gy_{i,t} = \alpha_i + c_1 * \ln(y_{i,t-1}) + c_2 * \ln(gpop_{i,t-1}) + c_3 * \ln(s_{i,t-1}) + c_4 * \ln(pat_{i,t-1}) + c_5 * \ln(hc_{i,t-1}) + c_6 * \ln(sf_{i,t-1}) + u_{i,t},$$

where gy is the growth rate of real income per capita, y is real income per capita, $gpop$ is the annual growth rate of population, s is the share of investment in GDP, pat is the number of patents per million people (a proxy for innovation), hc is the proxy for human capital and is measured as the percentage of population aged 25-64 with tertiary education, and sf is the structural funds variable, measured as a percentage of GDP or per capita, depending on the specification. Pinho et al. (2014) also include interaction terms of the structural fund variable with real income per capita (y), patents (pat) and human capital (hc). Overall, the authors find that the strongest positive effect of structural funds on economic growth occurs in countries with higher real income per capita, more patents per million population and better human capital.

Grusevaja and Pusch (2011) perform panel data analysis on a regional level with a particular focus on Central and Eastern European countries. Their data covers a period from 1999 to 2007 and is divided into two sub-periods of four years each, thus the time dimension amounts to two observations. The regression model is as follows: $g_t = c + Y_{t-1} + I_t + HC_t + n_t + SF_t + SF_t * Inst_t + Inst_t$, where g is the difference in logarithms of GDP per capita in purchasing power standards in the beginning and end of a sub-period, Y is the logarithm of GDP per capita in purchasing power standards, I is the logarithm of investment as percentage of GDP, HC is the human capital growth, n is the logarithm of the sum of population growth and a depreciation rate of 5%, SF is the amount of ERDF funds received as a percentage of GDP, and $Inst$ is an institutional variable. One of the several institutional variables used by Grusevaja and Pusch (2011) is the Corruption Perceptions Index from *Transparency International*, calculated as an average for 2003-2006. Among institutional variables, only corruption is found to be statistically significant: countries with lower corruption experience a larger positive growth effect from the structural funds.

7.3. Proposed regression model

Based on the discussion above, we specify our regression function which is intended to analyse the effect of corruption on the effectiveness of EU Regional Policy payments' impact on GDP growth as follows: $\ln(y_{it}) - \ln(y_{it-1}) = \alpha + \beta_1 \ln(y_{it-1}) + \beta_2 \ln\left(s_{k,it-1}\right) + \beta_3 \ln\left(s_{h,it-1}\right) + \beta_4 ln(n_{it-1} + g + \delta) + \beta_5 Corruption_{it-1} + \beta_6 \ln(CP_{it-1}) + \beta_7 Corruption_{it-1} * \ln(CP_{it-1}) + \varepsilon_{it}$.

The dependent variable in our regression is the growth rate of real GDP per capita in purchasing power standards (PPS). The independent variables include:

- Initial real GDP per capita in purchasing power standards (y_{it-1})
- Rate of physical capital accumulation ($s_{k,it-1}$) or investment, which is measured by gross fixed capital formation as a percentage of GDP.
- Level of human capital ($s_{h,it-1}$), which is measured either by human capital index per person or the average years of education of people over 25 years old.
- Population growth rate (n_{it-1}), rate of technological progress (g) and the time discount factor (δ), where $g + \delta$ are set to be constant at 5% following Mankiw et al. (1992).
- Corruption term ($\mathrm{Corruption}_{it-1}$), which is measured either by Corruption Perceptions Index (CPI) or Control of Corruption index from WGI.
- EU Cohesion Policy payments received as a percentage of GDP (CP_{it-1}), with values for the total payment amount and its three components: ERDF, CF and ESF.
- Interaction term ($\mathrm{Corruption}_{it-1} * \ln(\mathrm{CP}_{it-1})$), which measures the conditional effect of Cohesion Policy payments with regard to corruption levels.

Our regression function is consistent with existing research on the topic. Note that:

- If we set $\beta_5 = \beta_6 = \beta_7 = 0$, the function is the standard neoclassical growth model (augmented Solow model) introduced by Mankiw et al. (1992).
- If we set $\beta_5 = 0$, the function is similar to that of Ederveen (2006). We add the corruption term because we want to examine the direct effect of corruption on GDP growth in accordance with institutional theory research.
- If we set $\beta_5 = \beta_7 = 0$, the function is similar to Mohl and Hagen (2010). We add the interaction term because we also want to examine the effect of corruption on GDP growth through the channel of Cohesion Policy payments.

The expected signs of each independent variable are specified below:

- The coefficient on initial GDP is expected to be negative. The rationale behind it is the existence of β-convergence, which means that less economically developed countries will have larger growth levels than countries with high initial income.
- The coefficients on physical and human capital are expected to be positive because of their positive effect on the steady-state level of output and economic growth in economic theory.
- The coefficient on $n + g + \delta$ is considered to be negative in accordance with empirical research and the fact that population growth decreases the amount of capital available for each person.
- The coefficient on corruption is expected to be positive (high values of corruption indices mean low levels of corruption) as we expect that lower levels of corruption will contribute to a higher GDP growth. The sign is not straightforward, however, as relatively high corruption can also have a positive effect on GDP according to the efficient corruption theory. Furthermore, as we explain below, corruption enters the interaction term as well so its coefficient has to be interpreted carefully.
- The coefficient on Cohesion Policy payments is expected to be positive as higher amounts of structural fund support should stimulate economic growth. However, the sign is not straightforward once again as Ederveen (2006), for example, finds it to be negative, and since the variable enters the interaction term as well.
- The expectation for the sign of the coefficient on the interaction term is not clear initially. It depends heavily on the signs of coefficients on corruption and Cohesion Policy payments, as explained below.

Summarized expectations are provided in the table below:

Table 1. Expected coefficient signs

Independent Variable	Expected sign
$\ln(y_{it-1})$	$\beta_1 < 0$
$\ln(s_{k,it-1})$	$\beta_2 > 0$
$\ln(s_{h,it-1})$	$\beta_3 > 0$
$\ln(n_{it-1} + g + \delta)$	$\beta_4 < 0$
$Corruption_{it-1}$	$\beta_5 > 0$ (not necessarily)
$\ln(CP_{it-1})$	$\beta_6 > 0$ (not necessarily)
$Corruption_{it-1} * \ln(CP_{it-1})$	Not clear

As noted above, the inclusion of the interaction term makes the regression results more difficult to interpret. In a simple linear regression such as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, the coefficient on x_2 shows that an increase in the x_2 by one unit will cause an increase of β_2 units of y, holding other independent variables constant: $\frac{\partial y}{\partial x_2} = \beta_2$. In a semi-log regression with the dependent variable measured in levels and the independent variables measured in logs, such as $y = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \epsilon$, the effect of x_2 is a semi-elasticity: $\frac{\partial y}{\% \partial x_2} = \beta_2/100$. That is, when x_2 increases by 1%, y increases by $\beta_2/100$ units.

Moving on to the interaction terms, in a model such as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$, the partial effect of x_2 on y (holding all other independent variables fixed) is: $\frac{\partial y}{\partial x_2} = \beta_2 + \beta_3 x_1$. For example, if $\beta_3 > 0$, the effect of x_2 on y is larger for observations with a larger x_1. In this case, β_2 is the effect of x_2 on y when x_1 is zero.

In the case of our regression, $\ln(y_{it}) - \ln(y_{it-1}) = \alpha + \beta_1 \ln(y_{it-1}) + \beta_2 \ln(s_{k,it-1}) + \beta_3 \ln(s_{h,it-1}) + \beta_4 ln(n_{it-1} + g + \delta) + \beta_5 Corruption_{it-1} + \beta_6 \ln(CP_{it-1}) + \beta_7 Corruption_{it-1} * \ln(CP_{it-1}) + \varepsilon_{it}$, the dependent variable is GDP growth in levels, so the partial effect of corruption on growth is $\frac{\partial GDPG_{it}}{\partial Corruption_{it-1}} = \beta_5 + \beta_7 * \ln(CP_{it-1})$, where $GDPG_{it}$ is the growth of GDP ($\ln(y_{it}) - \ln(y_{it-1})$). So β_5 is the effect of a one unit increase in corruption on growth when $\ln(CP_{it-1}) = 0$, which happens when $CP_{it-1} = 1$, that is, Cohesion Policy payments amount to 1% of GDP. The partial effect of Cohesion Policy payments on GDP growth is $\frac{\partial GDPG_{it}}{\%\partial CP_{it-1}} = (\beta_6 + \beta_7 * Corruption_{it-1})/100$. So $\beta_6/100$ is the effect of a 1% increase in Cohesion Policy payments on GDP growth when $Corruption_{it-1} = 0$. This cannot practically happen when corruption is measured by CPI since the smallest available value is 0.8 for Somalia in 2014. In the case of Control of Corruption from WGI, this can theoretically happen as the index ranges from approximately -2.5 to 2.5 and values very close to 0 are observed in reality, for example, Saudi Arabia had a score of -0.01 in 2013 and Belize had a score of 0.02 in 2013.

These facts lead to different combinations of the coefficient signs on the corruption and Cohesion Policy variables and their interaction term. For example, in the case of corruption, when $\beta_5 > 0$ and $\beta_7 > 0$, the partial effect of lower corruption (higher corruption index values) on economic growth is positive and the effect is stronger for countries with a higher amount of Cohesion Policy payments received. When $\beta_5 > 0$ and $\beta_7 < 0$, the partial effect of lower corruption (higher corruption index values) on economic growth is positive initially and the effect is weaker and can become negative for countries with a higher amount of Cohesion Policy payments received. When $\beta_5 < 0$ and $\beta_7 > 0$, the partial effect of lower corruption (higher corruption index values) on economic growth is negative initially and the effect is weaker and can become positive for countries with a higher amount of Cohesion Policy payments received. Finally, when $\beta_5 < 0$ and $\beta_7 < 0$, the partial effect of lower corruption (higher corruption index values) on economic growth is negative initially and the effect is stronger for countries with a higher amount of Cohesion Policy payments received.

In the case of Cohesion Policy payments, when $\beta_6 > 0$ and $\beta_7 > 0$, the partial effect of more Cohesion Policy payments on economic growth is positive and the effect is stronger for countries with lower corruption (higher values of corruption indices). When $\beta_6 > 0$ and

$\beta_7 < 0$, the partial effect of more Cohesion Policy payments on economic growth is positive initially and the effect is weaker and can become negative for countries with lower corruption (higher values of corruption indices). When $\beta_6 < 0$ and $\beta_7 > 0$, the partial effect of more Cohesion Policy payments on economic growth is negative initially and the effect is weaker and can become positive for countries with lower corruption (higher values of corruption indices). Finally, when $\beta_6 < 0$ and $\beta_7 < 0$, the partial effect of more Cohesion Policy payments on economic growth is negative initially and the effect is stronger for countries with lower corruption (higher values of corruption indices).

8. DATA DESCRIPTION

Our main dataset used for regression analysis consists of panel data for 12 countries over the period 2007-2013. We also have GDP per capita values for 2014, but they are only used for estimating GDP growth for the last year and not used in any other way, hence we don't report the data for 2014 for any of the other variables further on. Therefore, the final panel dataset covers 12 countries over 7 time periods, for a total of 84 observations. The 12 countries of interest are the new EU members which joined the union in the two latest enlargements: Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia (all of which joined EU in 2004), as well as Bulgaria and Romania which both joined EU in 2007.

The first variable in the dataset is the Real GDP per capita in Purchasing Power Standards (PPS). The variable is acquired from the Eurostat database and is available for all countries and time periods. It is used to calculate real GDP per capita growth further in the analysis. The second variable is Gross fixed capital formation as a percentage of GDP, which is acquired from the Eurostat database as well. It measures investment as a percentage of GDP and is used by Mohl and Hagen (2010) in their structural fund regressions. Population growth rate is acquired from the World Bank's WDI database. It has one missing value for Bulgaria in 2007 and we use data from Eurostat to substitute for it.

As is common in the literature, we follow Mankiw et al. (1992) and assume that the sum of the technological progress and discount rate ($g + \delta$) is 5%. This value is stored in a separate variable. Our corruption data comes from two sources. The first variable uses the Corruption Perceptions Index (CPI) from *Transparency International*. Until 2012, CPI was reported on a scale from 0 to 10, but starting with 2012 it is now reported on a scale from 0 to 100, with higher scores assigned to countries with lower corruption. In order for the data to be comparable across years, we divide the 2012 and 2013 CPI values for all countries by ten to transform them to the 0 to 10 scale. The second corruption variable comes from the World

Bank's World Governance Indicators (WGI) and is called Control of Corruption. It ranges from -2.5 to 2.5, with higher values once again attributed to countries with lower corruption.

We also employ two different measures of human capital. The first variable comes from Penn World Tables version 8.1 developed by Feenstra et al. (2015) and reports the Human capital index per person. The data is based on Barro and Lee's (2012) years of schooling and Psacharopoulos' (1994) returns to education estimations. To be specific, it sets the human capital of country i at time t as a function of the average years of schooling, s: $hc_{it} = e^{\emptyset(s_{it})}$. The average years of schooling data used in this calculation comes from Barro and Lee (2012), and covers ages of 15 and older, which is different from Hall and Jones (1999), who consider the age of 25 and older. The rationale behind this is that the age from which people can start to work is 15 in many countries, thus excluding the ages of 15-25 may underestimate the amount of human capital. The function $\emptyset(\mathrm{s})$ is based on Psacharopoulos (1994) to reflect the fact that early years of education have a relatively higher return:

$$\emptyset(s) = \begin{cases} 0.134 * s & s \leq 4 \\ 0.134 * 4 + 0.101(s-4) & 4 < s \leq 8 \\ 0.134 * 4 + 0.101 * 4 + 0.068(s-8) & s > 8 \end{cases}$$

Human capital data from the Penn World Tables is not available for 2012 and 2013. However, as the 2011 data for all countries is the same as 2010, we assume that it has stayed the same in other years as well and assign the 2011 values to 2012 and 2013 data for all countries to avoid having missing observations.

The second measure of human capital comes from the United Nations Development Programme and measures the average years of education of people over 25 years old. It is based on the Barro and Lee's (2012) methodology, and observations for all countries and years are available. Barro and Lee (2012) create a relatively accurate measurement of educational attainment which is widely used as a proxy for human capital stock. Their data comes from national census and surveys and provides the distribution of educational attainment from the age of 15 and over by 5-year age groups. For missing values, they use the forward and backward extrapolation method, adjusted by mortality rate. In the newest Barro and Lee dataset, the world population aged 15 and above is estimated to have an average of 7.9 years of schooling in 2010, whereas advanced countries have 11.3 years of schooling and developing countries have 7.2 years of schooling on average.

Finally, our EU funds data comes from the European Commission's research on Regional Policy dataset. It contains actual payments to EU members by programming periods, years and funds. We use the separate payments from ERDF, CF and ESF funds as well as a total sum of all these funds. Following Mohl and Hagen (2010) and other research on this topic, we express these payments as a percentage of Nominal GDP of each country in every period. The Nominal GDP data comes from Eurostat. The table below summarizes all the information about the initial variables in the dataset.

Table 2. Description of initial variables in the panel dataset

Variable name	Description	Source
country	Country IDs (1 to 12)	---
year	Years for the data (2007-2013)	---
y	Real GDP per capita in purchasing power standards (PPS), euro	Eurostat
inv	Gross fixed capital formation, % of GDP	Eurostat
n	Population growth, %	World Bank WDI, Eurostat
deltag	Sum of technological progress and discount rate, %	Based on Mankiw et al. (1992)
cpi	Corruption Perceptions Index (CPI)	Transparency International
CC	Control of Corruption Index	World Bank WGI
hc	Human capital index per person	Feenstra et al. (2015): Penn World Tables 8.1
educ	Average years of education of people over 25 years old	United Nations Development Programme
erdf	Payments from the ERDF, % of GDP	European Commission and authors' calculations
cf	Payments from the CF, % of GDP	European Commission and authors' calculations
esf	Payments from the ESF, % of GDP	European Commission and authors' calculations
CP	Total EU Regional Policy payments (sum of ERDF, CF and ESF), % of GDP	European Commission and authors' calculations

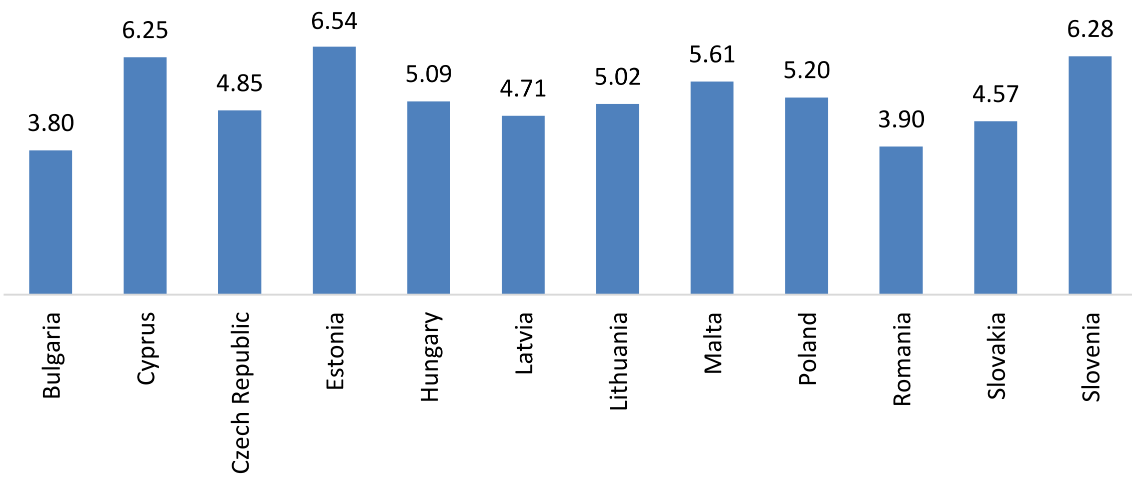
Below, we also provide the summary statistics for these variables (except for *country* and *year*).

Table 3. Summary statistics of initial variables in the panel dataset

	Number of observations	Mean	Standard deviation	Minimum	Maximum
y	84	17,807.14	4,092.15	10,400.00	27,200.00
inv	84	23.67	4.99	13.40	38.40
n	84	-0.17	0.82	-2.26	1.40
deltag	84	5.00	0.00	5.00	5.00
cpi	84	5.16	0.92	3.30	6.80
CC	84	0.42	0.44	-0.30	1.24
hc	84	3.10	0.17	2.85	3.47
educ	84	11.35	0.83	8.90	12.80
erdf	84	0.84	0.57	0.07	2.77
cf	84	0.56	0.35	0.00	1.74
esf	84	0.26	0.18	0.03	0.87
CP	84	1.66	1.00	0.20	4.34

The main variables of interest in our research are the indicators of corruption and EU funds. Therefore, we provide graphs with more information about these variables below. As can be seen, the country with the lowest corruption (highest CPI score) among the 12 countries in the dataset was Estonia, followed closely by Cyprus and Slovenia. The countries with the worst levels of corruption were Bulgaria and Romania.

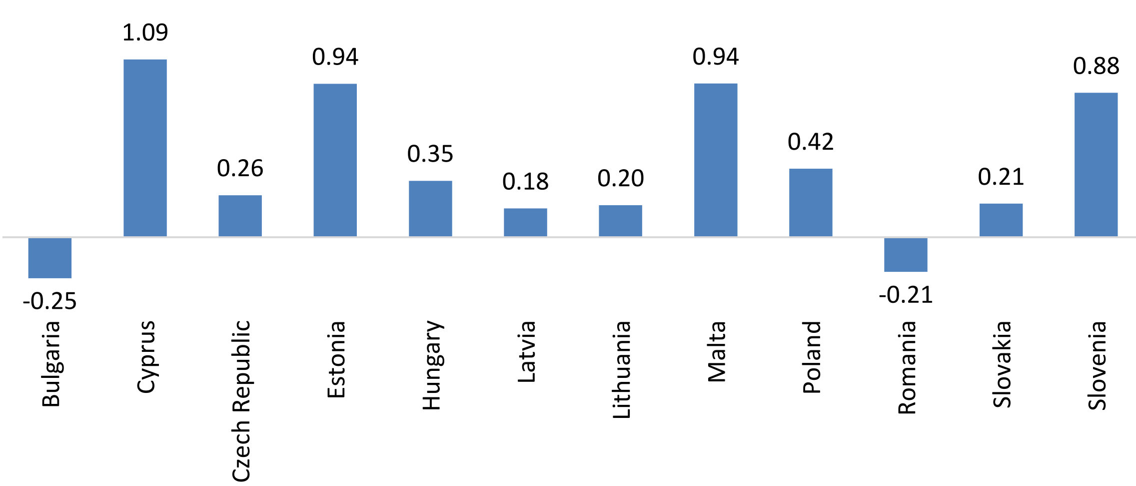
Figure 6. Average Corruption Perceptions Index (CPI) for the period of 2007-2013



Source: Transparency International (2015)

The results are only slightly different when looking at the Control of Corruption index from the World Bank WGI, as the graph below demonstrates. Cyprus has the lowest corruption (highest score), followed by Estonia and Malta. The highest corruption is observed in Bulgaria and Romania once again. Overall, we find the correlation between the CPI and Control of Corruption indices in our dataset to be 0.9, indicating a high degree of similarity.

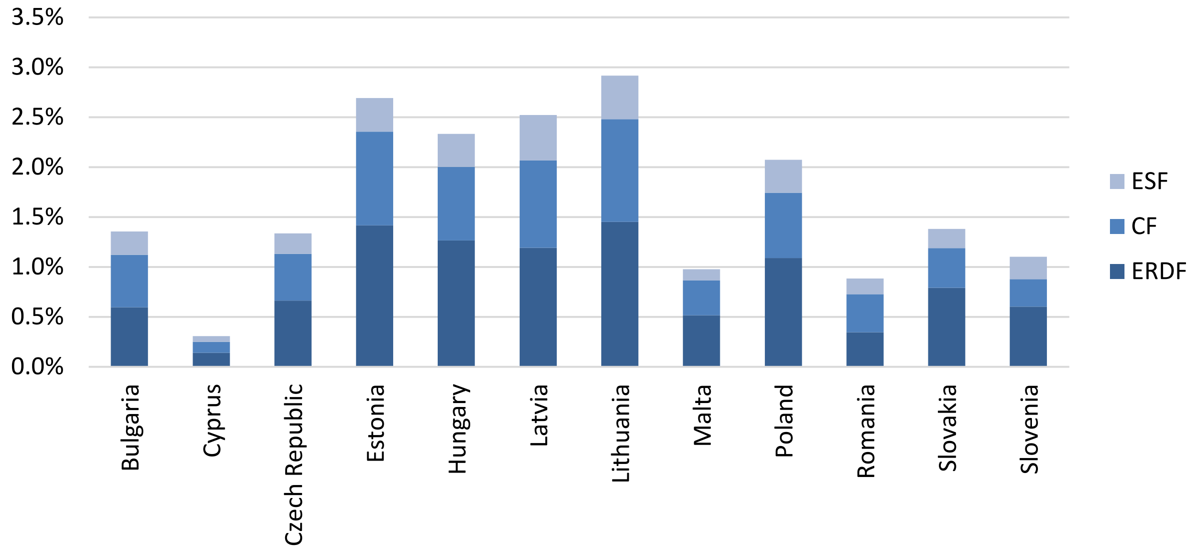
Figure 7. Average Control of Corruption index (CC) for the period of 2007-2013



Source: World Bank (2015)

We also analyse the distribution of EU funds. The graph below provides an overview of the total Cohesion Policy payments received by each country as an average percentage of GDP in 2007-2013. As can be seen, Lithuania received funds amounting to an average of almost 3% of GDP every year, while Estonia and Latvia received slightly less. The countries which received the smallest support are Cyprus and Malta (less than 1% of GDP). ERDF payments were larger than the remaining two funds sources in all countries.

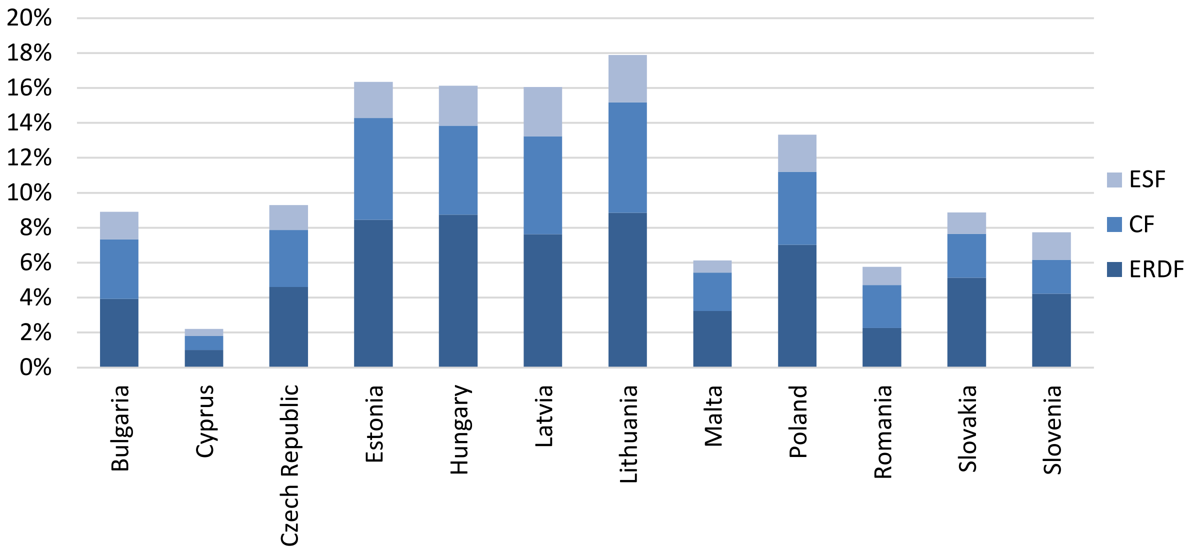
Figure 8. Average Cohesion Policy payments as a percentage of GDP for the period 2007-2013



Source: European Commission (2015)

When the total Cohesion Policy payments are expressed as a percentage of the 2013 GDP of countries, the results are largely the same. Lithuania received payments amounting to around 18% of its 2013 GDP in the 2007-2013 period. Estonia, Hungary and Latvia received about 16% of their GDPs. Cyprus, on the other hand, received only 2.2% of its GDP. Once again, ERDF was the largest source of support.

Figure 9. Total Cohesion Policy payments as a percentage of 2013 GDP for the period 2007-2013



Source: European Commission (2015)

9. REGRESSION ANALYSIS

In the following sections, we present the results of our regression estimations. We start with simple cross-sectional regressions and then move to more advanced panel data estimation techniques. We run all our regressions in the statistical software called *Stata* and provide the full code and output in the Appendix.

9.1. Cross-sectional regression

Our basic regression in the panel data case is $\ln(y_{it}) - \ln(y_{it-1}) = \alpha + \beta_1 \ln(y_{it-1}) + \beta_2 \ln\left(s_{k,it-1}\right) + \beta_3 \ln\left(s_{h,it-1}\right) + \beta_4 ln(n_{it-1} + g + \delta) + \beta_5 Corruption_{it-1} + \beta_6 \ln(CP_{it-1}) + \beta_7 Corruption_{it-1} * \ln(CP_{it-1}) + \varepsilon_{it}$. It is not possible to run this regression on cross-sectional data since the time dimension cannot be used in this case. Therefore, we use average and initial values of variables in the period. Specifically, our dependent variable is the average annual real GDP growth rate over the period 2007-2013. The first independent variable (initial GDP), $\ln(y_{it-1})$, is measured as the natural logarithm of GDP per capita in PPS in 2006. $\ln\left(s_{k,it-1}\right)$, a measure of investment, is the natural logarithm of the average gross fixed capital formation as a percentage of GDP in 2007-2013. Similarly, the human capital variable $\ln\left(s_{h,it-1}\right)$, measured as either the natural logarithm of the human capital index or average years of schooling, is an average over 2007-2013. The same holds for population growth, technological progress and the discount rate. Corruption, measured as either CPI or Control of Corruption index, is an average over 2007-2013 as well. Finally, cohesion fund variables are natural logarithms of average cohesion fund payments over 2007-2013. Summary statistics of these variables are presented in a table below. All the variable names are the same as described previously, with an additional variable "*gdpg*" which measures GDP growth.

Table 4. Summary statistics of cross-sectional data

	Number of observations	Mean	Standard deviation	Minimum	Maximum
gdpg	12	3.34	2.10	-0.26	6.76
y	12	15675.00	4624.47	9100.00	24200.00
inv	12	23.67	3.26	19.36	29.24
n	12	-0.17	0.80	-1.43	1.21
deltag	12	5.00	0.00	5.00	5.00
cpi	12	5.15	0.89	3.80	6.54
CC	12	0.42	0.45	-0.25	1.09
hc	12	3.10	0.18	2.89	3.41
educ	12	11.35	0.84	9.43	12.44
erdf	12	0.84	0.43	0.14	1.45
cf	12	0.56	0.29	0.11	1.03
esf	12	0.26	0.12	0.06	0.45
CP	12	1.66	0.83	0.31	2.92

The table below presents results of different cross-sectional regressions. We start with the basic model described above, whose results are shown in the first column. As the results indicate, none of the variables in the regression are significant. F-tests for joint significance of the coefficients on corruption variables (β_5 and β_7) and Cohesion Policy variables (β_6 and β_7) also show that none of them are jointly significant. Furthermore, the signs of coefficients are not always as expected. For example, the coefficient on the investment variable (gross fixed capital formation) is negative, indicating a negative effect of investment on growth after controlling for effects of other variables. The coefficient on the corruption variable, *CPI*, is negative as well. As lower CPI scores are associated with higher corruption rates, a negative coefficient means that improvements in corruption are associated with a fall in GDP growth initially. This is contrary to our initial assumptions. Combined with a positive interaction term, this means that the partial effect of corruption is negative initially and the effect is weaker and can become positive for countries with a higher amount of Cohesion Policy payments received. The coefficient on total Regional Policy payments is negative as well, indicating a negative effect on growth initially. The interaction term is positive, however, so this can be interpreted as evidence for the fact that the effect of Regional Policy payments in countries with high levels of corruption is negative, but it becomes weaker and moves towards becoming positive in countries with better levels of corruption. This is the same finding as in Ederveen (2006). The R^2 of the basic regression is 0.888, which means that the estimated regression model is able to explain about 89% of the variation of the dependent variable around its mean. R^2 adjusted for degrees of freedom is also high (69%), indicating a good fit. Overall, we do not want to pay too much attention to the results of the cross-section regression as it only has 12 observations and uses values averaged over time. Therefore, its estimates can be expected to be unreliable. Furthermore, structural funds can have a lagged effect on economic growth, but it is not possible to include lagged variables in the cross-sectional regression.

We also perform robustness tests of the basic regression. The second column reports results of regression using the Control of Corruption index from WGI instead of CPI. The results are very similar to the basic regression: none of the variables are significant and the signs of all coefficients are the same as before. The third column reports results when the index of human capital from the Penn World Tables is used instead of the average years of schooling indicator. Coefficient on initial GDP becomes positive against expectations of a negative sign, and the coefficient on the human capital index itself is now negative. This goes against economic theory. All other coefficient signs remain the same. Finally, based on Ederveen (2006), we measure Regional Policy payments by only including ERDF payments. In this specification, the coefficient on CPI becomes positive, indicating a positive effect of low corruption on economic growth initially. The coefficient on the payments variable remains negative and the interaction term is positive, providing support for the view that countries with lower corruption experience a better effect of Regional Policy payments. R^2 across the regressions remain very similar.

Table 5. Results of cross-sectional regressions

	Specification names			
	Basic	*Control of Corruption*	*Human capital*	*ERDF*
Log of initial Real GDP per capita	-3.576	-4.632	0.112	-3.305
	(2.69)	(3.08)	(4.87)	(3.09)
Log of gross fixed capital formation	-9.043	-8.766	-3.744	-8.827
	(4.64)	(5.33)	(4.88)	(5.40)
Log of average years of education	7.897	7.783		7.776
	(6.43)	(7.08)		(7.20)
Log of $n + g + \delta$	-6.097	-6.932	-4.773	-4.416
	(4.26)	(5.06)	(4.53)	(4.14)
CPI	-1.159		-1.369	0.502
	(0.95)		(1.18)	(0.76)
Log of total Regional Policy payments	-16.283	-4.123	-14.062	
	(8.57)	(2.84)	(8.72)	
CPI * log of Regional Policy payments	2.759		2.638	
	(1.43)		(1.57)	
Control of Corruption		-0.888		
		(1.99)		
Control of Corr. * Regional p. payments		4.775		
		(3.02)		
Log of human capital index			-14.460	
			(16.66)	
Log of ERDF payments				-12.630
				(8.22)
CPI * log of ERDF payments				2.198
				(1.40)
Constant	63.546**	69.010**	45.288	48.043
	(21.76)	(24.61)	(33.61)	(25.56)
R^2	0.888	0.870	0.870	0.866
Adjusted R^2	0.691	0.642	0.642	0.630
Joint significance of corruption variables (p-value of F test)	0.2643	0.3559	0.3416	0.3797
Joint significance of Cohesion Policy variables (p-value of F-test)	0.2637	0.3758	0.3356	0.3780
Observations	12	12	12	12

Notes: * means significance at 10%, ** - at 5%, and *** - at 1%. Standard errors are reported in parentheses.

We also tried using robust standard errors which estimate standard errors correctly in the presence of heteroskedasticity and correlated error terms. None of the coefficients became significant, but the F-test of joint significance of all independent variables now allowed to not accept the hypothesis that all the independent variables in the model are zero, indicating an acceptable fit of the regression. Regressions with non-robust standard errors indicated bad joint significance of all independent variables, with only the basic model having an F-statistic significant at the 10% level. Overall, however, we conclude that most of the signs of coefficients were in accordance with theoretical predictions.

In the next section we perform regression analysis on panel data, which is the main focus of our empirical analysis.

9.2. Panel regression

We start our panel data analysis with generating a lagged variable of real GDP per capita, taking natural logarithms of this variable and the original real GDP per capita and calculating the difference between them: $\ln(y_{it}) - \ln(y_{it-1})$. This is our dependent variable which measures annual real GDP per capita growth. In a similar fashion, we also take natural logarithms of the investment variable, the human capital variable, the term measuring population growth, technological progress and discount rate, as well as the Regional Policy payment variables. Finally, we create interaction terms between the corruption variable and the Regional Policy variables. We then proceed to running different regression specifications.

As stated previously, our basic model takes the following form: $\ln(y_{it}) - \ln(y_{it-1}) = \alpha + \beta_1 \ln(y_{it-1}) + \beta_2 \ln(s_{k,it-1}) + \beta_3 \ln(s_{h,it-1}) + \beta_4 ln(n_{it-1} + g + \delta) + \beta_5 Corruption_{it-1} + \beta_6 \ln(CP_{it-1}) + \beta_7 Corruption_{it-1} * \ln(CP_{it-1}) + \varepsilon_{it}$. Results of this model and models with different specifications for robustness tests are presented below. By default, *Stata* uses random effects estimation on panel data and we begin our analysis with random effects models. In the basic model, the coefficient on the logarithm of initial Real GDP per capita is negative, but not statistically significant. The sign of the coefficient is in accordance with theoretical expectations. Contrary to expectations, we find the coefficient on the investment variable to be negative and statistically significant at 1% level. This could happen due to the specificity of the period being analyzed (2007-2013): the global financial crisis occurred during this time and GDP of all countries contracted. If, for example, the governments of certain countries decided to stimulate economic growth with increased public investment, this could lead to the inverse relationship between GDP growth and investment levels during the period. The coefficient on the human capital variable (years of education) is positive and in line with expectations, but not statistically significant. The coefficient on the variable combining population growth, technological progress and the discount rate is negative and statistically significant at the 10% level, as expected. We find the coefficient on corruption (measured by CPI) to be negative and statistically significant at 1%, the coefficient on Regional Policy funds to be negative and statistically significant at 1% and the coefficient on the interaction of corruption and Regional Policy funds to be positive and statistically significant. F-tests for joint significance of the coefficients on corruption variables (β_5 and β_7) and Cohesion Policy variables (β_6 and β_7) also show that both of them are jointly significant at a 1% level.

In our opinion, these effects can be explained as follows: first of all, the effect of high corruption levels on economic growth when $\ln(CP_{it-1}) = 0$, which happens when $CP_{it-1} = 1$, that is, Cohesion Policy payments amount to 1% of GDP, is positive, but as the amount of funds received increases, high corruption levels start having a negative effect on growth. This can be explained by the theory of efficient corruption: countries in our sample all have corruption levels which are higher than in the most developed EU countries and also have lower institutional quality. Hence, corruption in this case can serve as a natural response to existing inefficiencies and can help bypass them. But as the amount of EU funds received increases, the country gets supervised more closely by the EU officials, and corruption can no longer help bypass the inefficiencies. Hence, it starts having a negative effect on growth. Secondly, the coefficients on the Regional Policy and interaction variables mean that in countries with high corruption, the effect of Regional Policy funds on economic growth is negative, while in countries with low corruption it becomes weaker and can become positive.

This can be explained by the fact that in countries with high corruption levels the funds are not used effectively and are allocated to projects which do not improve economic growth, hence human capital is allocated inefficiently and projects which could potentially improve economic performance are understaffed. So improvements in corruption levels, as measured by CPI, would lead to improvements in the effectiveness of Regional Policy payments.

Table 6. Results of random effects panel regressions

	Specification names			
	Basic	*Control of Corruption*	*Human capital*	*ERDF*
Log of initial Real GDP per capita	-2.598	-6.122*	-3.026	-1.691
	(3.29)	(3.53)	(4.94)	(3.29)
Log of gross fixed capital formation	-12.826***	-12.683***	-13.048***	-11.532***
	(2.83)	(3.99)	(3.81)	(2.88)
Log of average years of education	4.291	3.388		3.703
	(2.95)	(4.15)		(2.84)
Log of $n + g + \delta$	-3.069*	-3.281*	-3.449*	-1.807
	(1.68)	(1.96)	(1.98)	(1.46)
CPI	-1.546***		-1.460**	0.115
	(0.59)		(0.64)	(0.55)
Log of total Regional Policy payments	-15.083***	-2.520	-14.322***	
	(3.76)	(1.66)	(3.89)	
CPI * log of Regional Policy payments	2.795***		2.647***	
	(0.64)		(0.70)	
Control of Corruption		-0.465		
		(1.33)		
Control of Corr. * Regional p. payments		4.323***		
		(1.62)		
Log of human capital index			6.031	
			(13.62)	
Log of ERDF payments				-11.086***
				(2.51)
CPI * log of ERDF payments				2.161***
				(0.49)
Constant	70.477**	99.366**	79.086*	48.126
	(32.04)	(38.99)	(42.30)	(32.27)
R^2	0.305	0.266	0.305	0.304
Joint significance of corruption variables (p-value of χ^2 test)	0.0000***	0.0270**	0.0005***	0.0001***
Joint significance of Cohesion Policy variables (p-value of χ^2-test)	0.0000***	0.0001***	0.0008***	0.0001***
Observations	84	84	84	84

Notes: * means significance at 10%, ** - at 5%, and *** - at 1%. Standard errors are robust (allow for intragroup correlation and heteroskedasticity) and reported in parentheses.

Our second model uses Control of Corruption instead of CPI. In this case, the coefficient on the initial GDP becomes statistically significant at 10% and keeps its negative sign, the coefficients on Regional Policy and corruption become insignificant and their interaction remains positive and statistically significant. Corruption coefficients are jointly significant at the 5% level and Cohesion Policy coefficients are still jointly significant at 1% level, however. We conclude that CPI is a slightly better measure of corruption in our sample as the individual coefficients of corruption and Regional Policy payments lose their significance when Control of Corruption is used. The third model uses the index of human capital from Penn World Tables instead of the average years of schooling. The only difference with the basic model in this case is that CPI becomes less statistically significant, but is still significant at the 5% level. Corruption coefficients and Cohesion Policy coefficients are still jointly significant at 1% level. Finally, the model with ERDF payments instead of total Regional Policy payments has an insignificant coefficient on the variable combining population growth, technological progress and the discount rate, a positive, but insignificant coefficient on corruption, while other results remain the same as in the basic model. Therefore, it seems that the basic model provides the best fit for our regression specification.

Following Mohl and Hagen (2010), we also include fixed country and time effects in the model by estimating a fixed effects panel regression with time dummies. The model specification in this case becomes $\ln(y_{it}) - \ln(y_{it-1}) = \alpha + \beta_1 \ln(y_{it-1}) + \beta_2 \ln\left(s_{k,it-1}\right) + \beta_3 \ln\left(s_{h,it-1}\right) + \beta_4 ln(n_{it-1} + g + \delta) + \beta_5 Corruption_{it-1} + \beta_6 \ln(CP_{it-1}) + \beta_7 Corruption_{it-1} * \ln(CP_{it-1}) + \mu_i + \lambda_t + \varepsilon_{it}$, where μ_i and λ_t are fixed country and time effects, respectively. We report results for regressions with only country fixed effects, only time fixed effects and both country and time fixed effects. All the regressions are based on our original model. Firstly, the regression with country fixed effects produces a more significant estimate on the initial GDP variable and less significant estimates on the investment, Regional Policy and interaction terms. Coefficient on the variable combining population growth, technological progress and the discount rate becomes insignificant, as does the coefficient on CPI. Corruption coefficients become jointly insignificant and Cohesion Policy coefficients remain jointly significant, but only at a 10% level. All the coefficient signs remain the same, except for education, which becomes negative. The reason for the insignificant coefficient on CPI is that fixed effects estimation eliminates the effect of variables which stay constant over time and CPI might not have changed much during our research period so its effect has mostly been eliminated as well. Ederveen (2006) cautions against using fixed effects estimation in such regressions exactly for this reason. Furthermore, we perform a Hausman test for the choice between fixed and random effects estimations: the p-value of the test is 0.1933. The null hypothesis in the Hausman test is that the random effects model is preferred. Based on the p-value, we cannot reject the null hypothesis. Therefore, the random effects specification seems to be preferable in our case.

The regression with time dummies produces exactly the same results as the original equation, with the only difference being a more significant estimate on the variable combining population growth, technological progress and the discount rate. Finally, the specification which includes both country and time fixed effects produces an insignificant coefficient on CPI once again and has fewer significant coefficients in general. The joint significance hypotheses are only valid at 10% levels. Therefore, we conclude once again that our basic model provides a good estimate of the true population model.

Table 7. Results of fixed effects (FE) panel regressions

	Specification names			
	Basic	*Country FE*	*Time FE*	*Country and time FE*
Log of initial Real GDP per capita	-2.598	-18.692**	-0.177	-13.504
	(3.29)	(7.11)	(2.78)	(19.18)
Log of gross fixed capital formation	-12.826***	-12.880**	-8.381***	-5.046
	(2.83)	(4.32)	(2.32)	(4.30)
Log of average years of education	4.291	-38.930	1.129	-31.171
	(2.95)	(34.61)	(3.66)	(28.91)
Log of $n + g + \delta$	-3.069*	-9.564	-5.454**	-13.370**
	(1.68)	(6.23)	(2.52)	(5.53)
CPI	-1.546***	-0.944	-1.288***	-0.062
	(0.59)	(1.94)	(0.35)	(1.89)
Log of total Regional Policy payments	-15.083***	-12.622*	-12.792***	-10.544**
	(3.76)	(5.99)	(2.88)	(4.01)
CPI * log of Regional Policy payments	2.795***	2.910**	2.349***	2.084**
	(0.64)	(1.25)	(0.48)	(0.86)
Constant	70.477**	338.711***	46.255*	248.610
	(32.04)	(96.41)	(23.98)	(181.55)
year=2008			-11.508***	-10.867***
			(2.22)	(2.36)
year=2009			-0.420	-0.048
			(1.03)	(1.16)
year=2010			-1.805	-0.653
			(1.41)	(2.88)
year=2011			-2.707**	-0.579
			(1.14)	(3.76)
year=2012			-3.668***	-1.389
			(1.01)	(3.77)
year=2013			-3.293***	-0.515
			(1.17)	(4.20)
R^2	0.305	0.183	0.700	0.448
Joint significance of corruption variables (p-value of χ^2 or F-test)	0.0000***	0.1089	0.0000***	0.0894*
Joint significance of Cohesion Policy variables (p-value of χ^2 or F-test)	0.0000***	0.0989*	0.0000***	0.0686*
Observations	84	84	84	84

Notes: * means significance at 10%, ** - at 5%, and *** - at 1%. Standard errors are robust (allow for intragroup correlation and heteroskedasticity) and reported in parentheses.

We also control for a possible lagged effect of Regional Policy funds by running regressions with this variable lagged by 1, 2 and 3 periods. In all the specifications, the coefficient on the lagged Regional Policy variable becomes insignificant, but remains negative. In two of the specifications the interaction term becomes insignificant as well. Corruption coefficients and Cohesion Policy coefficients remain jointly statistically significant in the specification with one lag, but become jointly insignificant in other models. We do not consider it as evidence against the lagged effect of regional payments since our sample only includes 7 time periods and we cannot produce lags for many periods without losing many observations. For example, the model with a 3-period lag only has 48 observations remaining out of the original 84. Therefore, the lagged effect can exist in reality, but is not evident in our sample.

Table 8. Results of random effects panel regressions with a lagged Regional Policy payments variable

	Specification names			
	Basic	*1-period lag*	*2-period lag*	*3-period lag*
Log of initial Real GDP per capita	-2.598	-3.399	-4.042	-3.997
	(3.29)	(3.15)	(3.92)	(4.99)
Log of gross fixed capital formation	-12.826***	-17.769***	-0.992	0.085
	(2.83)	(6.32)	(3.14)	(1.14)
Log of average years of education	4.291	11.414	3.954	6.235
	(2.95)	(7.73)	(8.20)	(8.12)
Log of $n + g + \delta$	-3.069*	0.872	-5.916*	-7.340**
	(1.68)	(2.86)	(3.03)	(3.10)
CPI	-1.546***	-1.691**	0.164	0.157
	(0.59)	(0.76)	(0.30)	(0.55)
Log of total Regional Policy payments	-15.083***	-10.578	-2.801	-0.784
	(3.76)	(7.08)	(3.84)	(4.12)
CPI * log of Regional Policy payments	2.795***	2.103*	0.631	0.243
	(0.64)	(1.16)	(0.62)	(0.77)
Constant	70.477**	69.913*	44.850	37.475
	(32.04)	(38.94)	(34.04)	(37.60)
R^2	0.305	0.411	0.492	0.568
Joint significance of corruption variables (p-value of χ^2 test)	0.0000***	0.0318**	0.4196	0.7235
Joint significance of Cohesion Policy variables (p-value of χ^2-test)	0.0000***	0.0046***	0.1272	0.7149
Observations	84	72	60	48

Notes: * means significance at 10%, ** - at 5%, and *** - at 1%. Standard errors are robust (allow for intragroup correlation and heteroskedasticity) and reported in parentheses.

Finally, we run GMM regressions in order to correct for potential endogeneity. All the regressions are based on our original model, but include the lag of Real GDP growth as an additional independent variable. We also restrict the number of lags used as instruments in order for the number of instruments not to exceed the number of countries by a very large margin. If the number of instruments is very large, the errors are no longer asymptotically normal. The first column in the table below reports results for the Arellano-Bond specification when all variables are assumed to be exogenous. The coefficient on the lag of Real GDP growth is negative and statistically significant at 10%, the coefficients on initial Real GDP and investment are negative and statistically significant at 1%, the coefficient on CPI is negative and statistically significant at 10%, the coefficient on the Regional Policy variable is negative and statistically significant at 5% and the interaction term is positive and statistically significant at 1%. Average years of education and the variable combining population growth, technological progress and the discount rate are insignificant, but have expected signs. Overall, all the coefficient signs are the same as in our basic regression. Corruption coefficients and Cohesion Policy coefficients also remain jointly statistically significant at 1%. We also run the Arellano-Bond tests for zero correlation in first-differenced errors with the null hypothesis of no autocorrelation. If autocorrelation is present, some of the lags used as instruments become invalid. We see evidence for autocorrelation of order 1 at the 5% significance level. Therefore, the model's instruments can be invalid.

The second column reports the results of Arellano-Bond estimation, where we assume, following Mohl and Hagen (2010), that initial GDP, investment, Regional Policy payments and the interaction term are endogenous. The signs of coefficients remain the same as before, with the only exception being the sign on education which becomes negative. Many of the parameters also lose their significance and corruption coefficients become jointly insignificant. On the other hand, we do not find evidence for autocorrelation in the first-differenced error terms.

The third and fourth columns report results of the Blundell-Bover two-step estimation, where once again the first model assumes exogeneity of all independent variables and the second model uses the same endogeneity assumptions as the Arellano-Bond estimation above. All the coefficients in these regressions become insignificant, except for investment in the first model. Corruption and Cohesion Policy coefficients also become jointly insignificant. The signs of some coefficients are also not supported by economic theory. On the other hand, the instruments appear valid as there is no evidence for autocorrelation once again.

Table 9. Results of panel regressions with GMM estimation

	Specification names			
	A-B (exogenous)	*A-B (endo-genous)*	*B-B (exogenous)*	*B-B (endo-genous)*
Lag of Real GDP growth	-0.155*	-0.077	0.004	-0.114
	(0.09)	(0.13)	(0.28)	(0.20)
Log of initial Real GDP per capita	-28.412***	-11.492	-30.889	-17.856
	(8.62)	(13.19)	(22.43)	(25.30)
Log of gross fixed capital formation	-22.087***	-25.372***	-23.468*	-24.782
	(5.71)	(6.73)	(12.02)	(24.25)
Log of average years of education	40.010	-99.023	12.289	19.483
	(59.81)	(62.89)	(111.54)	(117.54)
Log of $n + g + \delta$	-8.607	-13.808*	12.457	1.839
	(5.58)	(7.53)	(28.69)	(22.69)
CPI	-2.747*	-1.879	-3.687	-1.847
	(1.41)	(1.66)	(3.16)	(4.22)
Log of total Regional Policy payments	-8.375**	-0.840	-7.320	0.367
	(4.19)	(12.57)	(42.31)	(60.43)
CPI * log of Regional Policy payments	2.019***	1.625	1.945	0.099
	(0.73)	(2.17)	(7.61)	(9.12)
Constant	278.035*	462.132***	346.376	212.779
	(148.46)	(134.85)	(320.82)	(136.19)
Number of observations	60	60	72	72
Number of countries	12	12	12	12
Number of instruments	13	29	18	54
Joint significance of corruption variables (p-value of χ^2 test)	0.0042***	0.3362	0.4916	0.8444
Joint significance of Cohesion Policy variables (p-value of χ^2-test)	0.0030***	0.0000***	0.4671	0.9947
AR(1) p-value	0.0389	0.2632	0.6931	0.7804
AR(2) p-value	0.8056	0.9474	0.7885	0.9715

Notes: * means significance at 10%, ** - at 5%, and *** - at 1%. Standard errors are robust and reported in parentheses.

Overall, GMM estimation seems to perform poorer than our original model. One of the reasons for this underperformance is the limited number of cross-country observations that we have. Arellano-Bond and Blundell-Bover estimators need many cross-sectional observations to produce reliable results. Furthermore, even though we restricted the number of lags to be used as instruments, the number of instruments still exceeded the number of cross-sectional observations in all the specifications. Therefore, hypothesis testing could have been unreliable as well.

10. INTERPRETATION OF RESULTS AND RECOMMENDATIONS

Before we proceed to provide policy recommendations, we provide estimates of the effects of corruption and Cohesion Policy payments based on our estimated regression coefficients. Our main regression model estimation gives the following results: $\ln(y_{it}) - \ln(y_{it-1}) = 70.477 - 2.598 * \ln(y_{it-1}) - 12.826 * \ln\left(s_{k,it-1}\right) + 4.291 * \ln\left(s_{h,it-1}\right) - 3.069 * ln(n_{it-1} + g + \delta) - 1.546 * Corruption_{it-1} - 15.083 * \ln(CP_{it-1}) + 2.795 * Corruption_{it-1} * \ln(CP_{it-1})$. Based on the formulas for partial effects of corruption and Regional Policy payments, below we present the estimated size of effects of corruption and Regional Policy payments on economic growth in our sample, holding other independent variables constant.

The partial effect of corruption is $\frac{\partial GDPG_{it}}{\partial Corruption_{it-1}} = \beta_5 + \beta_7 * \ln(CP_{it-1})$. We use the average value of the logarithm of total Cohesion Policy payments in 2007-2013 for each country as the value for $\ln(CP_{it-1})$. This allows us to estimate the partial effect of corruption on annual real GDP growth per capita at the mean value of Cohesion Policy payments. The results presented below show the effect (in percentage points) of a 1 unit increase in the CPI score on annual real GDP growth per capita. Therefore, e.g., if the value reported in the table is 0.82, this means that a 1 unit increase in the CPI score would increase the annual real GDP growth per capita by 0.82 percentage points at the mean value of Cohesion Policy payments. We see that in half of the countries, an improvement in CPI scores would be associated with a positive impact on GDP growth, while in the other half the effect would be negative. This means that at the average level of Regional Policy payments, the effect of lower corruption levels on GDP is still negative in half of the countries, so it would take a higher amount of Regional Policy payments than the historical average to make low corruption have a positive effect on GDP growth. For this reason, we also report the effect on growth of a 1 unit increase in the CPI score, assuming that average Regional Policy payments as a percentage of GDP are 1 percentage point higher than they were in reality in 2007-2013. In this case, the effect becomes positive in all countries except Cyprus. As was mentioned before, our potential explanation of this effect is that as the amount of Regional Policy payments received by a country increases, so does the amount of monitoring and control from European authorities. Therefore, this leads to a better performance of the country's supervising authorities and efficient corruption cannot exist in its purest form any longer. So the effect of lower corruption on GDP growth becomes positive in this case.

Table 10. Effect of a 1 unit improvement in CPI score on annual real GDP per capita growth by country, percentage points

Country	Effect of a 1 unit improvement in CPI score	The same effect under the assumption that historical average Regional Policy payments are 1 p.p. higher
Bulgaria	-0.70	0.85
Cyprus	-4.86	-0.80
Czech Republic	-0.74	0.83
Estonia	1.22	2.11
Hungary	0.82	1.82
Latvia	1.04	1.97
Lithuania	1.45	2.27
Malta	-1.61	0.36
Poland	0.49	1.59
Romania	-1.89	0.22
Slovakia	-0.65	0.88
Slovenia	-1.28	0.53

We do the same analysis for the effect of Regional Policy payments. The partial effect of Regional Policy payments is $\frac{\partial GDPG_{it}}{\%\partial CP_{it-1}} = (\beta_6 + \beta_7 * Corruption_{it-1})/100$. We use the average value of the CPI score in 2007-2013 for each country as the value for $Corruption_{it-1}$. This allows us to estimate the partial effect of Regional Policy payments on annual real GDP growth per capita at the mean value of CPI. The results presented below show the effect (in percentage points) of a 1% increase in Regional Policy payments as a percentage of GDP on annual real GDP growth per capita. Therefore, e.g., if the value reported in the table is 0.024, this means that a 1% increase in the Regional Policy payments as a percentage of GDP would increase the annual real GDP growth per capita by 0.024 percentage points at the mean value of CPI. We see that at the mean value of CPI, the effect of more Regional Policy payments on GDP growth is positive only in countries with lower corruption levels (Cyprus, Estonia, Malta and Slovenia). As was previously noted, this can be explained by the fact that in countries with high corruption levels, the funds are not used effectively and are allocated to projects which do not improve economic growth, hence human capital is allocated inefficiently and projects which could potentially improve economic performance are understaffed. So improvements in corruption levels, as measured by CPI, would lead to improvements in the effectiveness of Regional Policy payments. For the sake of comparison, we also include the effect on growth of a 1% increase in Regional Policy payments, assuming that the average CPI score of each country is 1 unit higher than it was in 2007-2013. A higher score is associated with lower corruption, and we see that in this case the effect of Regional Policy payments on growth becomes positive in all countries but Bulgaria and Romania, since these two countries have the worst corruption levels among the twelve countries analyzed.

Table 11. Effect of a 1% increase in Regional Policy payments on annual real GDP per capita growth by country, percentage points

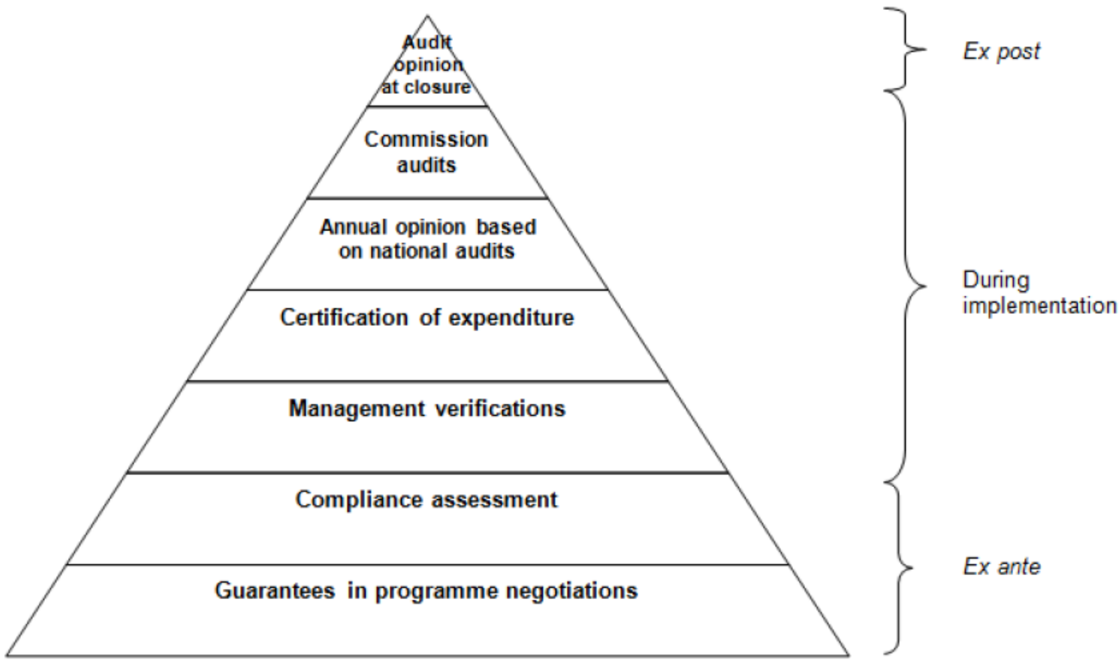
Country	**Effect of a 1% increase in Regional Policy payments**	**The same effect under the assumption that historical average CPI score is 1 unit higher**
Bulgaria	-0.045	-0.017
Cyprus	0.024	0.052
Czech Republic	-0.015	0.013
Estonia	0.032	0.060
Hungary	-0.008	0.020
Latvia	-0.019	0.009
Lithuania	-0.010	0.017
Malta	0.006	0.034
Poland	-0.006	0.022
Romania	-0.042	-0.014
Slovakia	-0.023	0.005
Slovenia	0.025	0.053

We view these results with a certain degree of caution, however, due to the traditional limitations of regression analysis. First of all, regression results do not prove the existence of a causal relationship between the dependent and independent variables. The causal relationship must be driven by underlying theory, while regression results can only provide evidence of a statistically significant relationship between the variables and the direction (positive or negative) of this relationship. Secondly, even though we use many measures to overcome methodological issues and perform various robustness analyses, the issues may still be present. Our main regression model is run on 84 observations, hence the estimator may not necessarily be asymptotically normal. Omitted and unobserved variables can also impact the results, even though we attempted to eliminate their impact by using fixed effects and GMM estimation. Moreover, the functional form of our regression and the selection of variables is based on the neoclassical economic theory, but it does not necessarily hold in reality, even though it is used universally in research on this topic. Despite the abovementioned facts, our results align well with the findings of other researchers on this topic and are robust and significant, at least in statistical terms.

Before providing recommendations on how to improve EU Cohesion Policy regulations based on our results, we briefly overview the current regulations aimed at controlling the use of provided funds. First of all, as specified by the Council of the European Union (2005), the countries receiving EU funds have to co-finance the projects, contributing about 15%-20% of the total amount. This limits the degree to which the funds can be appropriated, but still leaves many opportunities for corrupt behavior. European Commission (2011) states that currently, management and control systems of EU Cohesion Policy payments are mainly implemented by the member states themselves, and the Commission found that the error rate (which apart from unintentional misreporting, also includes fraud) for the Cohesion payments is much higher than for other components of the budget. This shows that the payments are susceptible to corruption. For example, the Commission reports that about 40% of all errors involve public procurement errors, such as unlawful selection criteria or award of contracts without a tender process. As the amounts of money distributed are large, it pays off for the companies to invest into pressuring the authorities to make favorable decisions, often involving bribes, especially in countries with high corruption levels.

Overall, the European Commission (2011) summarizes the current control system in the following figure:

Figure 10. Current control system of EU Cohesion Policy payments



Source: European Commission (2011)

The main step before the project implementation stage is compliance assessment when the Commission reviews the regulatory setup in a certain country. At the stage of implementation, control takes place at national level: local authorities perform control of management, expenditure certification and audits. European Commission also performs certain audits itself. Finally, at the close of the projects, the Commission reviews the final report submitted by the local authorities.

Our recommendations are based on the regression results: we find that in countries with high corruption, the effect of Regional Policy funds on economic growth is negative, while in countries with low corruption it becomes positive. The standard recommendations could be strengthening the monitoring and control mechanisms, increasing the punishment for corrupt behavior, making certain activities illegal (such as the selection of projects based on political priorities) or finding a completely new way of fostering cohesion which would be different from the current system of EU subsidies. Apart from these usual recommendations, several ideas come to our mind.

Firstly, instead of leaving most of the management and control systems in the hands of national authorities, the European authorities could take on more responsibility in these areas. There could be a central Cohesion Policy control office, independent from local authorities and interest groups, which would perform a more thorough evaluation of project implementation from the EU side. Of course, this would also increase the administration costs of Cohesion Policy, so the amount of benefits from centralization must be carefully weighed against its costs.

Secondly, we propose to make a share of the funds allocated to each country go towards improving its institutional quality and fighting corruption instead of just being allocated to different projects. Since the success of Cohesion Policy depends critically on the level of corruption, the policy should be focused on improving bureaucratic processes in the first place. Currently, it is the regions with GDP and income levels below those of the EU average that receive the largest amount of funds. Instead, the countries with highest corruption levels should get a larger share of funds allocated to fighting corruption. The whole process of fighting corruption should be implemented and supervised by EU authorities or international organizations, such as *Transparency International*. Otherwise, the funds allocated to fighting corruption could become misappropriated as well.

11. CONCLUSIONS

This thesis aimed to investigate how corruption affects the impact of European Union's Cohesion Policy payments on economic growth in new EU member states. Regression analysis results indicate that in countries with high corruption levels, the effect of Cohesion Policy payments on GDP growth is negative, but it is positive in countries with low corruption levels. We believe that this can be explained by the fact that the allocated funds are not used effectively in countries with high corruption: the funds are allocated to projects which are less welfare-improving and provide private gains to public authorities or private agents implementing the projects. Therefore, the effect of Cohesion Policy payments is only positive when corruption is low and the funds are not misappropriated. Based on the results of our regression analysis, we recommend the European authorities to restructure the current process of granting Cohesion Policy funds. First of all, the monitoring and control of project implementation could be centralized by the EU authorities. Secondly, the funds could be allocated towards improving institutional quality and fighting corruption in the first place.

Apart from the empirical analysis, this thesis also provided a theoretical overview of the issue being analyzed. We found that the neo-classical economic model is most often used in this kind of research. The basis for our analysis comes from institutional economics. Research in this field provides evidence for the impact of institutional quality on development differences across countries. In particular, corruption is one of the institutional characteristics that has been widely analyzed. Overall, the largest share of research points to the fact that corruption hinders economic growth, although positive and no effect of corruption on growth have also been documented. We also overviewed the different corruption models and found that although most of them view corruption as a negative phenomenon, the theory of efficient corruption states that corruption can have a positive impact on growth when it helps to overcome the inefficiencies in public administration.

The results of our regression analysis are in line with those of other researchers. For example, Ederveen (2006) found the same effect of corruption on Cohesion Policy effectiveness.

Nevertheless, the results are limited by the usual shortcomings of regression analysis. In particular, endogeneity is the most important issue to be considered. We employed fixed effects and GMM estimation to overcome this issue. GMM estimators of Arellano-Bond and Blundell-Bover use lags of variables already included in regression as instrumental variables since no reliable external instruments have been documented in research so far. Therefore, future research on the topic could focus on testing different instruments based on the original model. Our methodology could also be tested on regional data for GMM estimation to be more consistent. This would require the use of regional corruption measures as well as a proxy for human capital (for example, the number of patents) instead of the years of schooling.

BIBLIOGRAPHY

1. Acemoglu, D., Gallego, F., & Robinson, J. A. (2014). Institutions, Human Capital and Development. *Annual Reviews of Economics, 6*, 875-912.

2. Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review, 91*(5), 1369-1401.

3. Acemoglu, D., & Robinson, J. (2010). The Role of Institutions in Growth and Development. *Review of Economics and Institutions, 1*(2).

4. Aidt, T., Dutta, J., & Sena, V. (2008). Governance regimes, corruption and growth: Theory and evidence. *Journal of Comparative Economics, 36*(2), 195-220.

5. Aidt, T. S. (2003). Economic analysis of corruption: A survey. *Economic Journal, 113*(491), F632-F652.

6. Barro, R., & Lee, J-W. (2012). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, *104*, 184-198.

7. Barro, R., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy, 100*(2), 223-251.

8. Becker, G. S., & Stigler, G. J. (1974). Law Enforcement, Malfeasance, and Compensation of Enforcers. *Journal of Legal Studies (0047-2530), 3*(1), 1.

9. Besley, T., & McLaren, J. (1993). Taxes and Bribery: The Role of Wage Incentives. *Economic Journal*, *103*(416), 119-141.

10. Beugelsdijk, M., & Eijffinger, S. C. W. (2005). The Effectiveness of Structural Policy in the European Union: An Empirical Analysis for the EU-15 in 1995–2001. *JCMS: Journal of Common Market Studies*, *43*(1), 37-51.

11. Caselli, F. (2005). *Accounting for cross-country income differences*. Centre for Economic Performance, London School of Economics and Political Science.

12. Council of the European Union (2005). *Financial Perspective 2007-2013*. Brussels. Retrieved July 1, 2015, on World Wide Web: http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/misc/87677.pdf

13. Cubas, G., Ravikumar, B., & Ventura, G. (2013). Talent, Labor Quality, and Economic Development. *Federal Reserve Bank of St. Louis: Working Papers, 2013-027*.

14. Ederveen, S., Groot, H. L. F., & Nahuis, R. (2006). Fertile soil for structural funds? A panel data analysis of the conditional effectiveness of European cohesion policy. *Kyklos, 59*(1), 17-42.

15. European Commission (2011). *Analysis of errors in Cohesion Policy for the years 2006-2009. Actions taken by the Commission and the way forward*. Brussels

16. European Commission (2015). *Cohesion Policy Data*. Retrieved April 29, 2015, on World Wide Web: https://cohesiondata.ec.europa.eu/

17. European Commission (2015). *Financial framework 2007 – 2013*. Retrieved April 27, 2015, on World Wide Web: http://ec.europa.eu/budget/figures/fin_fwk0713/fwk0713_en.cfm

18. European Commission (2015). *Regional Policy: Data for research*. Retrieved June 6, 2015, on World Wide Web: http://ec.europa.eu/regional_policy/en/policy/evaluations/data-for-research/

19. European Commission (2015). *Summaries of EU legislation. Glossary*. Retrieved April 26, 2015, on World Wide Web: http://europa.eu/legislation_summaries/glossary/structural_cohesion_fund_en.htm

20. European Commission (2015). *The EU's main investment policy*. Retrieved April 26, 2015, on World Wide Web: http://ec.europa.eu/regional_policy/en/policy/what/investment-policy/

21. European Commission (2001). *Unity, solidarity, diversity for Europe, its people and its territory. Second report on economic and social cohesion*. Brussels.

22. Eurostat (2015). *Eurostat database*. Retrieved June 6, 2015, on World Wide Web: http://ec.europa.eu/eurostat/data/database

23. Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The Next Generation of the Penn World Table. Forthcoming *American Economic Review*, available for download at www.ggdc.net/pwt

24. Glaeser, E. L., LaPorta, R., López-de-Silanes, F., & Shleifer, A. (2004). Do Institutions Cause Growth? *Journal of Economic Growth, 9*(3), 271-303.

25. Grusevaja, M., & Pusch, T. (2011). How Does Institutional Setting Affect the Impact of EU Structural Funds on Economic Cohesion? New Evidence from Central and Eastern Europe: Halle Institute for Economic Research.

26. Hagen, T., & Mohl, P. (2011). Econometric Evaluation of EU Cohesion Policy: A Survey. *International Handbook on the Economics of Integration, Volume III*: Edward Elgar.

27. Hall, R. E., & Jones, C. I. (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics*, *114*(1), 83-116.

28. Jain, A. K. (2001). Corruption: A review. *Journal of Economic Surveys*, *15*(1), 71-116.

29. Kaarsen, N. (2014). Cross-country differences in the quality of schooling. *Journal of Development Economics, 107*(0), 215-224.

30. Klenow, P. J., & Rodríguez-Clare, A. (1997). The Neoclassical Revival in Growth Economics: Has It Gone Too Far?*: Vol. 12. NBER Macroeconomics Annual* (pp. 62-103).

31. Knack, S., & Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics, 7*(3), 207-227.

32. Laffont, J.-J., & N'Guessan, T. (1999). Competition and corruption in an agency relationship. *Journal of Development Economics, 60*(2), 271.

33. Lui, F. T. (1985). An Equilibrium Queuing Model of Bribery. *Journal of Political Economy, 93*(4), 760-781.

34. Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics, 107*(2), 407-437.

35. Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics, 110*(3), 681-712.

36. Méndez, F., & Sepúlveda, F. (2006). Corruption, growth and political regimes: Cross country evidence. *European Journal of Political Economy, 22*(1), 82-98.

37. Méon, P. G., & Sekkat, K. (2005). Does corruption grease or sand the wheels of growth? *Public Choice, 122*(1-2), 69-97.

38. Mo, P. H. (2001). Corruption and Economic Growth. *Journal of Comparative Economics, 29*(1), 66-79.

39. Mohl, P., & Hagen, T. (2010). Do EU structural funds promote regional growth? New evidence from various panel data approaches. *Regional Science and Urban Economics, 40*(5), 353-365.

40. Pinho, C., Varum, C., & Antunes, M. (2014). Structural Funds and European Regional Growth: Comparison of Effects among Different Programming Periods. *European Planning Studies, 23*(7), 1302-1326.

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

42. Rahman, A. T. R. (1986). Legal and administrative measures against bureaucratic corruption in Asia. In L. V. Carino (Ed.), *Bureaucratic corruption in Asia: Causes, consequences, and controls* (pp. 109-162). Quezon City: JMC Press and College of Public Administration, University of the Philippines.

43. Richter, B. K., & Timmons, J. F. (2012). Why Not Adopt Better Institutions? *Oxford Development Studies*, *40*(2), 272-281.

44. Rogoff, K., & Obstfeld, M. (1996). *Foundations of International Macroeconomics*. Cambridge, Massachusetts: MIT Press.

45. Romer, D. (2012). *Advanced Macroeconomics* (4th ed.). New York, NY: McGraw-Hill.

46. Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy, 98*(5), S71-S102.

47. Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal, 9(1), 86-136*.

48. Rose-Ackerman, S. (2007). *International Handbook on the Economics of Corruption*: Edward Elgar.

49. Shleifer, A., & Vishny, R. W. (1994). The Politics of Market Socialism. *Journal of Economic Perspectives, 8*(2), 165-176.

50. Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, *70*, 65-94.

51. Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics*, *39*, 312-320.

52. Studenmund, A.H. (2006). *Using econometrics: a practical guide* (5th ed.). Boston: Pearson/Addison Wesley.

53. Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives, 19*(3), 19-42.

54. Tanzi, V. (1998). Corruption around the world. *IMF Staff Papers, 45*(4), 559.

55. The World Bank Group (2014). *WGI 2014 Dataset.* Retrieved April 20, 2015, on World Wide Web: http://info.worldbank.org/governance/wgi/wgidataset.xlsx

56. The World Bank Group (2014). *WGI 2014 Documentation.* Retrieved April 20, 2015, on World Wide Web: http://info.worldbank.org/governance/wgi/index.aspx#doc

57. The World Bank Group (2014). *World Development Indicators database*. Retrieved June 6, 2015, on World Wide Web: http://data.worldbank.org/products/wdi

58. Transparency International (2014). *Corruption Perceptions Index 2014: Full Source Description.* Retrieved April 20, 2015, on World Wide Web: http://files.transparency.org/content/download/1842/12378/file/2014_CPISources_EN.pdf

59. Transparency International (2014). *Corruption Perceptions Index 2014: Results.* Retrieved April 20, 2015, on World Wide Web: http://www.transparency.org/cpi2014/results

60. Treisman, D. (2007). What have we learned about the causes of corruption from ten years of cross-national empirical research? In N. W. Polsby (Series Ed.), *Annual Review of Political Science: Vol. 10* (pp. 211-244).

61. United Nations Development Programme (2015). *Human Development Reports data: Mean years of schooling (of adults) (years).* Retrieved June 6, 2015, on World Wide Web: http://hdr.undp.org/en/content/mean-years-schooling-adults-years

62. Varga, J., & In'T Veld, J. (2011). Cohesion policy spending in the new member states of the EU in an endogenous growth model. *Eastern European Economics*, *49*(5), 29-54.

63. Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

APPENDICES

LIST OF APPENDICES

APPENDIX 1. CROSS-COUNTRY DATASET

country	gdpg	y	inv	n	delta g	cpi	CC	hc	educ	erdf	cf	esf	CP
1	4.1	9100	25.3	-0.6	5	3.8	-0.2	2.9	10.5	0.6	0.5	0.2	1.4
2	-0.3	24200	20.8	1.2	5	6.3	1.1	2.9	11.3	0.1	0.1	0.1	0.3
3	1.5	19800	27.2	0.4	5	4.9	0.3	3.4	12.4	0.7	0.5	0.2	1.3
4	3.5	15700	27.4	-0.3	5	6.5	0.9	3.3	12.0	1.4	0.9	0.3	2.7
5	2.3	15100	21.3	-0.3	5	5.1	0.3	3.3	11.3	1.3	0.7	0.3	2.3
6	3.8	13500	25.8	-1.4	5	4.7	0.2	3.0	11.3	1.2	0.9	0.5	2.5
7	5.6	13600	20.5	-1.4	5	5.0	0.2	3.1	12.3	1.4	1.0	0.4	2.9
8	2.5	19200	19.4	0.6	5	5.6	0.9	3.0	9.4	0.5	0.4	0.1	1.0
9	5.6	12300	20.6	0.1	5	5.2	0.4	2.9	11.6	1.1	0.7	0.3	2.1
10	6.8	9300	29.2	-0.8	5	3.9	-0.2	3.0	10.6	0.3	0.4	0.2	0.9
11	4.1	15300	23.2	0.1	5	4.6	0.2	3.2	11.6	0.8	0.4	0.2	1.4
12	0.7	21000	23.3	0.4	5	6.3	0.9	3.3	11.8	0.6	0.3	0.2	1.1

APPENDIX 2. PANEL DATASET

country	year	inv	n	deltag	cpi	CC	hc	educ	erdf	cf	esf	CP	y
1	2007	28.6	-0.7432	5	4.1	-0.2335	2.8811	10.2	0.2011	0.3979	0.0744	0.6733	10400
1	2008	33.5	-0.7019	5	3.6	-0.3037	2.8874	10.4	0.2638	0.3972	0.0976	0.7585	11200
1	2009	28.7	-0.6443	5	3.8	-0.2471	2.8937	10.5	0.3554	0.6563	0.1314	1.1431	10600
1	2010	22.9	-0.6583	5	3.6	-0.207	2.9	10.6	0.9498	0.3374	0.1167	1.4039	11000
1	2011	20.8	-0.6412	5	3.3	-0.225	2.9	10.6	0.4549	0.7501	0.2178	1.4228	11500
1	2012	21.5	-0.5792	5	4.1	-0.237	2.9	10.6	0.952	0.8182	0.3324	2.1026	11900
1	2013	21.3	-0.5596	5	4.1	-0.2926	2.9	10.6	0.9894	0.3129	0.672	1.9743	11900
1	2014												12300
2	2007	25.6	1.4001	5	5.3	1.0756	2.8527	11	0.0672	0.0983	0.0321	0.1976	25800
2	2008	27.3	1.3078	5	6.4	1.2413	2.8838	11.1	0.0981	0.0931	0.0594	0.2506	27200
2	2009	23.5	1.2423	5	6.6	0.9329	2.9153	11.1	0.2225	0.1227	0.0564	0.4016	25600
2	2010	21.8	1.197	5	6.3	1.0048	2.9471	11.3	0.1169	0.1661	0.0681	0.3511	25900
2	2011	19.1	1.1556	5	6.3	0.8872	2.9471	11.5	0.1175	0.2462	0.0738	0.4375	24900
2	2012	15.2	1.1117	5	6.6	1.2419	2.9471	11.6	0.0987	0.0469	0.073	0.2186	24700
2	2013	13.4	1.0724	5	6.3	1.2355	2.9471	11.6	0.2516	0.0005	0.0303	0.2824	23600
2	2014												23300
3	2007	29.6	0.5835	5	5.2	0.2342	3.4749	12.8	0.3711	0.1793	0.1176	0.668	21600
3	2008	29	0.8294	5	5.2	0.27	3.4449	12.6	0.4552	0.3917	0.1894	1.0363	21100
3	2009	27.1	0.5697	5	4.9	0.3282	3.4152	12.5	0.686	0.5088	0.1347	1.3296	20200
3	2010	27	0.2914	5	4.6	0.2591	3.3858	12.3	0.8623	0.3899	0.1486	1.4008	20600
3	2011	26.6	0.2067	5	4.4	0.2967	3.3858	12.3	0.8671	0.0249	0.1601	1.0521	21600
3	2012	26.1	0.1399	5	4.9	0.229	3.3858	12.3	0.8897	0.9137	0.1861	1.9895	21800
3	2013	24.9	0.0332	5	4.8	0.1903	3.3858	12.3	0.5004	0.8792	0.4957	1.8754	21900
3	2014												23000
4	2007	36.6	-0.4562	5	6.5	0.9137	3.2814	11.9	0.6817	0.5314	0.1544	1.3675	17700
4	2008	31.2	-0.2681	5	6.6	0.8696	3.2901	12	0.5347	0.7048	0.2094	1.449	17600
4	2009	22.7	-0.1928	5	6.6	0.9124	3.2987	12	2.3524	0.9097	0.3892	3.6513	15200
4	2010	21.2	-0.2281	5	6.5	0.861	3.3074	12	2.3044	1.0034	0.5391	3.8468	16100
4	2011	25.7	-0.3036	5	6.4	0.9292	3.3074	12	0.995	0.4439	0.1726	1.6114	17800
4	2012	27	-0.3579	5	6.4	0.9809	3.3074	12	1.475	1.7363	0.6128	3.8241	18800
4	2013	27.3	-0.3559	5	6.8	1.1091	3.3074	12	1.5576	1.2404	0.2994	3.0973	19500
4	2014												19900
5	2007	23.7	-0.1549	5	5.3	0.5588	3.2544	11.2	0.7214	0.4183	0.171	1.3106	15600
5	2008	23.3	-0.1751	5	5.1	0.3842	3.2589	11.2	0.4638	0.3989	0.2269	1.0895	16200
5	2009	22.9	-0.1549	5	5.1	0.3412	3.2635	11.3	1.1025	0.9253	0.2792	2.307	15600
5	2010	20.4	-0.226	5	4.7	0.2524	3.2681	11.3	1.2097	0.6739	0.2386	2.1222	16400
5	2011	19.8	-0.2834	5	4.6	0.3192	3.2681	11.3	2.1784	0.8514	0.5655	3.5953	17000
5	2012	19.1	-0.5164	5	5.5	0.2792	3.2681	11.3	1.4201	0.5198	0.3808	2.3207	17100
5	2013	19.9	-0.2671	5	5.4	0.2882	3.2681	11.3	1.755	1.3798	0.4343	3.5691	17600
5	2014												18500
6	2007	36.5	-0.8162	5	4.8	0.2481	2.9881	10.8	0.8903	0.7711	0.3571	2.0184	15400
6	2008	32	-1.0509	5	5	0.1298	2.9979	11	0.739	0.6624	0.1765	1.5779	15600
6	2009	22.5	-1.651	5	4.5	0.1267	3.0077	11.3	1.523	0.7349	0.2226	2.4806	12900
6	2010	19.1	-2.0813	5	4.3	0.1254	3.0175	11.5	0.7728	1.1339	0.8749	2.7816	13500
6	2011	22.1	-1.8208	5	4.2	0.1881	3.0175	11.5	1.2768	0.8003	0.6428	2.7199	14700
6	2012	25.2	-1.2404	5	4.9	0.1547	3.0175	11.5	1.8296	1.0699	0.5054	3.4049	16000
6	2013	23.3	-1.071	5	5.3	0.2657	3.0175	11.5	1.3094	0.9578	0.3936	2.6608	17000
6	2014												17600

7	2007	28.6	-1.1879	5	4.8	0.035	3.0438	12.1	0.7937	0.6024	0.276	1.6722	15600
7	2008	26	-1.0285	5	4.6	0.0362	3.0608	12.2	0.8294	0.8446	0.3197	1.9937	16200
7	2009	17.9	-1.1103	5	4.9	0.1213	3.0779	12.3	2.7665	1.2877	0.287	4.3413	13800
7	2010	16.9	-2.0969	5	5	0.2707	3.095	12.4	1.3117	1.2616	0.6481	3.2214	15300
7	2011	18.4	-2.2585	5	4.8	0.2399	3.095	12.4	1.2935	1.2142	0.5644	3.0721	17000
7	2012	17.3	-1.3412	5	5.4	0.3119	3.095	12.4	1.6234	1.0914	0.4689	3.1837	18300
7	2013	18.2	-1.012	5	5.7	0.3636	3.095	12.4	1.5313	0.8989	0.5002	2.9305	19400
7	2014												20100
8	2007	22.4	0.3488	5	5.8	1.0619	2.9607	8.9	0.4693	0.2918	0.0919	0.8529	20200
8	2008	19.6	0.6507	5	5.8	1.0442	2.9756	9.1	0.3667	0.271	0.0785	0.7162	20900
8	2009	18.2	0.7539	5	5.2	0.8344	2.9906	9.2	0.316	0.1852	0.0741	0.5753	20500
8	2010	21.4	0.4912	5	5.6	0.855	3.0056	9.4	0.4258	0.4937	0.0797	0.9992	21800
8	2011	17.5	0.4237	5	5.6	0.8344	3.0056	9.6	0.6776	0.3931	0.1568	1.2275	21900
8	2012	18.3	0.7627	5	5.7	0.9643	3.0056	9.9	0.6864	0.3745	0.1239	1.1849	22300
8	2013	18.1	0.93	5	5.6	0.9905	3.0056	9.9	0.6549	0.4469	0.1828	1.2846	22800
8	2014												23300
9	2007	21.9	-0.0543	5	4.2	0.1928	2.883	11.4	0.6787	0.412	0.319	1.4097	13700
9	2008	22.6	0.0136	5	4.6	0.3465	2.8892	11.4	0.6372	0.4596	0.1857	1.2826	14100
9	2009	21.1	0.0678	5	5	0.3701	2.8953	11.6	0.8517	0.7698	0.3214	1.9428	14400
9	2010	19.8	0.0841	5	5.3	0.4132	2.9015	11.7	1.2662	0.6315	0.2687	2.1664	15600
9	2011	20.3	0.9137	5	5.5	0.4851	2.9015	11.8	1.3898	0.7135	0.4404	2.5437	16600
9	2012	19.4	0.0045	5	5.8	0.5855	2.9015	11.8	1.4349	0.8283	0.4471	2.7102	17400
9	2013	18.8	-0.0555	5	6	0.5482	2.9015	11.8	1.3576	0.7605	0.3419	2.4601	17900
9	2014												18600
10	2007	36	-1.4772	5	3.7	-0.1708	2.9673	10.4	0.1432	0.3061	0.0588	0.5081	10700
10	2008	38.4	-1.6664	5	3.8	-0.1572	2.9786	10.5	0.1891	0.4085	0.0776	0.6753	12500
10	2009	26	-0.8331	5	3.8	-0.2667	2.9899	10.5	0.38	0.5002	0.1236	1.0039	11900
10	2010	25.9	-0.594	5	3.7	-0.2154	3.0013	10.6	0.1257	0.3034	0.1351	0.5642	12600
10	2011	27.1	-0.4919	5	3.6	-0.1902	3.0013	10.7	0.3121	0.1705	0.1379	0.6204	13300
10	2012	27.5	-0.4452	5	4.4	-0.2721	3.0013	10.7	0.4481	0.3533	0.0829	0.8843	14000
10	2013	23.8	-0.383	5	4.3	-0.2034	3.0013	10.7	0.8213	0.6246	0.4764	1.9224	14500
10	2014												14600
11	2007	26.9	0.0292	5	4.9	0.3038	3.1616	11.6	0.5778	0.3323	0.217	1.1271	17300
11	2008	25.7	0.0858	5	5	0.3034	3.1632	11.6	0.5685	0.4734	0.1987	1.2406	18500
11	2009	21.8	0.1333	5	4.5	0.2281	3.1649	11.6	0.432	0.322	0.1638	0.9178	17300
11	2010	22.2	0.0932	5	4.3	0.2353	3.1665	11.6	0.8533	0.6881	0.0699	1.6113	18500
11	2011	24.2	0.1289	5	4	0.2434	3.1665	11.6	1.0828	0.197	0.201	1.4808	18900
11	2012	21.3	0.1702	5	4.6	0.0673	3.1665	11.6	1.3054	0.5024	0.3445	2.1523	19600
11	2013	20.4	0.1075	5	4.7	0.058	3.1665	11.6	0.7233	0.2517	0.1464	1.1213	20000
11	2014												20800
12	2007	28.8	0.5592	5	6.6	0.9807	3.2567	11.5	0.1974	0.1797	0.0803	0.4574	22500
12	2008	29.6	0.1581	5	6.7	0.9112	3.2632	11.6	0.1861	0.2785	0.1264	0.591	23100
12	2009	24.3	0.9039	5	6.6	1.0235	3.2697	11.7	0.4313	0.4866	0.1176	1.0355	20700
12	2010	21.2	0.4361	5	6.4	0.8538	3.2762	11.8	0.6977	0.3316	0.29	1.3193	21000
12	2011	20.2	0.2077	5	5.9	0.9006	3.2762	11.9	1.0565	0.1384	0.1994	1.3942	21500
12	2012	19.2	0.21	5	6.1	0.8134	3.2762	11.9	0.9199	0.2826	0.3728	1.5753	21600
12	2013	19.7	0.1357	5	5.7	0.7018	3.2762	11.9	0.7185	0.2251	0.3876	1.3312	21800
12	2014												22600

APPENDIX 3. CROSS-COUNTRY REGRESSION CODE AND OUTPUT

```
. clear all

. cd "C:\Users\User\Dropbox\Thesis\Analysis\Data"
C:\Users\User\Dropbox\Thesis\Analysis\Data

. use dataset_cross.dta

. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
country	12	6.5	3.605551	1	12
gdpg	12	3.338017	2.096174	-.2557	6.7641
y	12	15675	4624.466	9100	24200
inv	12	23.66547	3.262174	19.3571	29.2429
n	12	-.170125	.8033242	-1.4336	1.2124
deltag	12	5	0	5	5
cpi	12	5.152858	.886344	3.8043	6.5357
CC	12	.416325	.4507464	-.2494	1.0884
hc	12	3.099642	.1771446	2.8946	3.4112
educ	12	11.34524	.8447996	9.4286	12.4429
erdf	12	.8380417	.4328125	.1389	1.4499
cf	12	.5617583	.285864	.1106	1.0287
esf	12	.255925	.1238233	.0562	.4533
CP	12	1.655717	.8258019	.3056	2.9164

```
. estpost sum gdpg y inv n deltag cpi CC hc educ erdf cf esf CP
```

	e(count)	e(sum_w)	e(mean)	e(Var)	e(sd)	e(min)	e(max)	e(sum)
gdpg	12	12	3.338017	4.393944	2.096174	-.2557	6.7641	40.0562
y	12	12	15675	2.14e+07	4624.466	9100	24200	188100
inv	12	12	23.66547	10.64178	3.262174	19.3571	29.2429	283.9857
n	12	12	-.170125	.6453298	.8033242	-1.4336	1.2124	-2.0415
deltag	12	12	5	0	0	5	5	60
cpi	12	12	5.152858	.7856058	.886344	3.8043	6.5357	61.8343
CC	12	12	.416325	.2031723	.4507464	-.2494	1.0884	4.9959
hc	12	12	3.099642	.0313802	.1771446	2.8946	3.4112	37.1957
educ	12	12	11.34524	.7136864	.8447996	9.4286	12.4429	136.1429
erdf	12	12	.8380417	.1873266	.4328125	.1389	1.4499	10.0565
cf	12	12	.5617583	.0817182	.285864	.1106	1.0287	6.7411
esf	12	12	.255925	.0153322	.1238233	.0562	.4533	3.0711
CP	12	12	1.655717	.6819488	.8258019	.3056	2.9164	19.8686

```
. esttab using summarystat_cross.rtf, cells("count mean(fmt(2)) sd(fmt(2)) min(fmt(2)) max(fmt(2))") nomtitle nonumber
(output written to summarystat_cross.rtf)

. generate lny=ln(y)

. generate lninv=ln(inv)

. generate deltang=deltag+n

. generate lndeltang=ln(deltang)

. generate lnhc=ln(hc)

. generate lneduc=ln(educ)

. generate lnerdf=ln(erdf)

. generate lncf=ln(cf)

. generate lnesf=ln(esf)

. generate lnCP=ln(CP)

. generate CPI_lnCP=cpi*ln(CP)

. generate CPI_CP=cpi*CP

. generate CC_lnCP=CC*ln(CP)

. generate CC_CP=CC*CP

. generate CPI_lnERDF=cpi*ln(erdf)

. generate CPI_ERDF=cpi*erdf
```

```
. *the following regression is our basic model
. regress gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP

      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  7,     4) =    4.51
       Model |  42.9033373     7  6.12904819           Prob > F      =  0.0819
    Residual |  5.43004233     4  1.35751058           R-squared     =  0.8877
-------------+------------------------------           Adj R-squared =  0.6910
       Total |  48.3333797    11  4.39394361           Root MSE      =  1.1651

------------------------------------------------------------------------------
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.576276   2.685997    -1.33   0.254     -11.0338    3.881246
       lninv |   -9.04315   4.638202    -1.95   0.123    -21.92086    3.834564
      lneduc |   7.896857   6.433319     1.23   0.287      -9.9649    25.75861
   lndeltang |  -6.097307   4.263703    -1.43   0.226    -17.93524     5.74063
         cpi |  -1.158602   .9456389    -1.23   0.288    -3.784116    1.466913
        lnCP |  -16.28306   8.569654    -1.90   0.130    -40.07623    7.510118
    CPI_lnCP |   2.759364   1.427743     1.93   0.125    -1.204686    6.723415
       _cons |    63.5464   21.75735     2.92   0.043     3.138324    123.9545
------------------------------------------------------------------------------

. estimates store m1, title(Model 1)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

       F(  2,     4) =    1.89
            Prob > F =    0.2643

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

       F(  2,     4) =    1.89
            Prob > F =    0.2637


. *the following regression is our basic model with CC instead of CPI
. regress gdpg lny lninv lneduc lndeltang CC lnCP CC_lnCP

      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  7,     4) =    3.81
       Model |  42.0328848     7  6.00469782           Prob > F      =  0.1068
    Residual |  6.30049491     4  1.57512373           R-squared     =  0.8696
-------------+------------------------------           Adj R-squared =  0.6415
       Total |  48.3333797    11  4.39394361           Root MSE      =   1.255

------------------------------------------------------------------------------
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -4.631524   3.076122    -1.51   0.207    -13.17221     3.90916
       lninv |  -8.766083   5.333052    -1.64   0.176    -23.57301    6.040844
      lneduc |   7.782899   7.083579     1.10   0.334    -11.88427    27.45007
   lndeltang |  -6.932377   5.063285    -1.37   0.243    -20.99031    7.125556
          CC |  -.8884661   1.985094    -0.45   0.678     -6.39997    4.623038
        lnCP |  -4.122954   2.844999    -1.45   0.221    -12.02194    3.776028
     CC_lnCP |    4.77501   3.018556     1.58   0.189    -3.605845    13.15586
       _cons |   69.00973   24.61458     2.80   0.049     .6686925    137.3508
------------------------------------------------------------------------------

. estimates store m2, title(Model 2)

. test CC CC_lnCP

 ( 1)  CC = 0
 ( 2)  CC_lnCP = 0

       F(  2,     4) =    1.35
            Prob > F =    0.3559

. test lnCP CC_lnCP

 ( 1)  lnCP = 0
 ( 2)  CC_lnCP = 0

       F(  2,     4) =    1.26
            Prob > F =    0.3758
```

```
. *the following regression is our basic model with hc instead of educ
. regress gdpg lny lninv lnhc lndeltang cpi lnCP CPI_lnCP

      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  7,     4) =    3.82
       Model |  42.0431876     7  6.00616966           Prob > F      =  0.1065
    Residual |  6.29019205     4  1.57254801           R-squared     =  0.8699
-------------+------------------------------           Adj R-squared =  0.6421
       Total |  48.3333797    11  4.39394361           Root MSE      =   1.254

------------------------------------------------------------------------------
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |   .1121674   4.867923     0.02   0.983    -13.40335    13.62769
       lninv |   -3.74367   4.878171    -0.77   0.486    -17.28764    9.800303
        lnhc |  -14.45975    16.6554    -0.87   0.434    -60.70255    31.78304
   lndeltang |  -4.773202   4.530505    -1.05   0.352     -17.3519    7.805496
         cpi |  -1.369314   1.175226    -1.17   0.309    -4.632265    1.893636
        lnCP |  -14.06194   8.722585    -1.61   0.182    -38.27971    10.15584
    CPI_lnCP |   2.638264    1.57199     1.68   0.169    -1.726279    7.002808
       _cons |   45.28758   33.61291     1.35   0.249    -48.03681     138.612
------------------------------------------------------------------------------

. estimates store m3, title(Model 3)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

       F(  2,     4) =    1.42
            Prob > F =    0.3416

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

       F(  2,     4) =    1.45
            Prob > F =    0.3356


. *the following regression is our basic model with ERDF instead of CP
. regress gdpg lny lninv lneduc lndeltang cpi lnerdf CPI_lnERDF

      Source |       SS       df       MS              Number of obs =      12
-------------+------------------------------           F(  7,     4) =    3.68
       Model |  41.8330354     7  5.97614792           Prob > F      =  0.1129
    Residual |  6.50034422     4  1.62508606           R-squared     =  0.8655
-------------+------------------------------           Adj R-squared =  0.6302
       Total |  48.3333797    11  4.39394361           Root MSE      =  1.2748

------------------------------------------------------------------------------
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.304559   3.093187    -1.07   0.346    -11.89262    5.283506
       lninv |  -8.826876   5.396938    -1.64   0.177    -23.81118    6.157426
      lneduc |   7.776273   7.202741     1.08   0.341    -12.22174    27.77429
   lndeltang |  -4.416136   4.136849    -1.07   0.346    -15.90187    7.069599
         cpi |   .5020393   .7597023     0.66   0.545    -1.607232    2.611311
      lnerdf |  -12.62958   8.217932    -1.54   0.199    -35.44622    10.18705
  CPI_lnERDF |   2.198449   1.403499     1.57   0.192    -1.698289    6.095186
       _cons |   48.04328    25.5564     1.88   0.133    -22.91265    118.9992
------------------------------------------------------------------------------

. estimates store m4, title(Model 4)

. test cpi CPI_lnERDF

 ( 1)  cpi = 0
 ( 2)  CPI_lnERDF = 0

       F(  2,     4) =    1.25
            Prob > F =    0.3797

. test lnerdf CPI_lnERDF

 ( 1)  lnerdf = 0
 ( 2)  CPI_lnERDF = 0

       F(  2,     4) =    1.25
            Prob > F =    0.3780

.
. estout m1 m2 m3 m4 using output.rtf, cells(b(star fmt(3)) se(par fmt(2))) starlevels (* 0.1 ** 0.05 *** 0.01) ///
> legend label varlabels(_cons constant)                ///
> stats(r2 r2_a N, fmt(3 3 0) label(R_squared Adj_r2 Number_of_obs))
(output written to output.rtf)
```

```
. *robust SEs
. *the following regression is our basic model
. regress gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP, vce(robust)

Linear regression                                      Number of obs =      12
                                                       F(  7,      4) =   38.29
                                                       Prob > F      =  0.0017
                                                       R-squared     =  0.8877
                                                       Root MSE      =  1.1651

------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.576276    3.10781    -1.15   0.314    -12.20494    5.052387
       lninv |   -9.04315   4.169111    -2.17   0.096    -20.61846    2.532159
      lneduc |   7.896857     5.7015     1.39   0.238    -7.933044    23.72676
   lndeltang |  -6.097307   4.340594    -1.40   0.233    -18.14873    5.954113
         cpi |  -1.158602   .6987288    -1.66   0.173    -3.098584    .7813803
        lnCP |  -16.28306   8.807779    -1.85   0.138    -40.73737    8.171257
    CPI_lnCP |   2.759364   1.448424     1.91   0.129    -1.262105    6.780833
       _cons |    63.5464   19.66369     3.23   0.032     8.951245    118.1416
------------------------------------------------------------------------------

. estimates store m5, title(Model 1)

. *the following regression is our basic model with CC instad of CPI
. regress gdpg lny lninv lneduc lndeltang CC lnCP CC_lnCP, vce(robust)

Linear regression                                      Number of obs =      12
                                                       F(  7,      4) =   39.95
                                                       Prob > F      =  0.0015
                                                       R-squared     =  0.8696
                                                       Root MSE      =   1.255

------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -4.631524   3.059455    -1.51   0.205    -13.12593    3.862884
       lninv |  -8.766083   4.158678    -2.11   0.103    -20.31242    2.780259
      lneduc |   7.782899   6.284675     1.24   0.283    -9.666158    25.23196
   lndeltang |  -6.932377    4.72662    -1.47   0.216    -20.05558    6.190824
          CC |  -.8884661   1.211682    -0.73   0.504    -4.252634    2.475702
        lnCP |  -4.122954   2.889164    -1.43   0.227    -12.14456     3.89865
     CC_lnCP |    4.77501   3.066028     1.56   0.194    -3.737649    13.28767
       _cons |   69.00973    21.1303     3.27   0.031      10.3426    127.6769
------------------------------------------------------------------------------

. estimates store m6, title(Model 2)


. *the following regression is our basic model with hc instad of educ
. regress gdpg lny lninv lnhc lndeltang cpi lnCP CPI_lnCP, vce(robust)

Linear regression                                      Number of obs =      12
                                                       F(  7,      4) =    9.07
                                                       Prob > F      =  0.0250
                                                       R-squared     =  0.8699
                                                       Root MSE      =   1.254

------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |   .1121674   5.406581     0.02   0.984    -14.89891    15.12324
       lninv |   -3.74367   3.964758    -0.94   0.399     -14.7516    7.264263
        lnhc |  -14.45975   16.26872    -0.89   0.424    -59.62898    30.70947
   lndeltang |  -4.773202   5.394762    -0.88   0.426    -19.75146    10.20506
         cpi |  -1.369314   1.209101    -1.13   0.321    -4.726316    1.987688
        lnCP |  -14.06194   10.56053    -1.33   0.254    -43.38266    15.25878
    CPI_lnCP |   2.638264   1.901405     1.39   0.238    -2.640881     7.91741
       _cons |   45.28758   28.40688     1.59   0.186    -33.58256    124.1577
------------------------------------------------------------------------------

. estimates store m7, title(Model 3)
```

```
. *the following regression is our basic model with ERDF instead of CP
. regress gdpg lny lninv lneduc lndeltang cpi lnerdf CPI_lnERDF, vce(robust)

Linear regression                                      Number of obs =      12
                                                       F(  7,     4) =   28.72
                                                       Prob > F      =  0.0029
                                                       R-squared     =  0.8655
                                                       Root MSE      =  1.2748

------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.304559   3.218732    -1.03   0.363    -12.24119    5.632073
       lninv |  -8.826876   4.176092    -2.11   0.102    -20.42157    2.767813
      lneduc |   7.776273   5.733611     1.36   0.247    -8.142784    23.69533
   lndeltang |  -4.416136   3.757247    -1.18   0.305    -14.84793    6.015655
         cpi |   .5020393   .5772859     0.87   0.434    -1.100763    2.104842
      lnerdf |  -12.62958   7.375522    -1.71   0.162    -33.10732    7.848149
  CPI_lnERDF |   2.198449   1.246634     1.76   0.153    -1.262761    5.659659
       _cons |   48.04328   25.28892     1.90   0.130    -22.17002    118.2566
------------------------------------------------------------------------------

. estimates store m8, title(Model 4)
```

APPENDIX 4. PANEL REGRESSION CODE AND OUTPUT

```
. clear all

. cd "C:\Users\User\Dropbox\Thesis\Analysis\Data"
C:\Users\User\Dropbox\Thesis\Analysis\Data

. use dataset_panel_new.dta

. sum if inrange(year, 2007, 2013)

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     country |         84         6.5    3.472786          1         12
        year |         84        2010    2.012012       2007       2013
         inv |         84    23.66548    4.987031       13.4       38.4
           n |         84   -.1701357    .8248811    -2.2585     1.4001
      deltag |         84           5           0          5          5
-------------+---------------------------------------------------------
         cpi |         84    5.155952    .9231104        3.3        6.8
          cc |         84    .4163286      .44382     -.3037     1.2419
          hc |         84    3.099639    .1714929     2.8527     3.4749
        educ |         84    11.34524     .833689        8.9       12.8
        erdf |         84    .8380429    .5692826      .0672     2.7665
-------------+---------------------------------------------------------
          cf |         84    .5617512    .3535842      .0005     1.7363
         esf |         84    .2559238    .1797038      .0303      .8749
          CP |         84    1.655717    .9950895      .1976     4.3413
           y |         84    17807.14    4092.155      10400      27200

. estpost sum y inv n deltag cpi cc hc educ erdf cf esf CP if inrange(year, 2007, 2013)

             |  e(count)   e(sum_w)    e(mean)     e(Var)      e(sd)     e(min)     e(max)     e(sum)
-------------+------------------------------------------------------------------------------------------
           y |        84         84   17807.14   1.67e+07   4092.155      10400      27200    1495800
         inv |        84         84   23.66548   24.87048   4.987031       13.4       38.4     1987.9
           n |        84         84  -.1701357   .6804289   .8248811    -2.2585     1.4001   -14.2914
      deltag |        84         84          5          0          0          5          5        420
         cpi |        84         84   5.155952   .8521328   .9231104        3.3        6.8      433.1
          cc |        84         84   .4163286   .1969762     .44382     -.3037     1.2419    34.9716
          hc |        84         84   3.099639   .0294098   .1714929     2.8527     3.4749   260.3697
        educ |        84         84   11.34524   .6950373    .833689        8.9       12.8        953
        erdf |        84         84   .8380429   .3240827   .5692826      .0672     2.7665    70.3956
          cf |        84         84   .5617512   .1250218   .3535842      .0005     1.7363    47.1871
         esf |        84         84   .2559238   .0322935   .1797038      .0303      .8749    21.4976
          CP |        84         84   1.655717   .9902031   .9950895      .1976     4.3413   139.0802

. esttab using summarystat1.rtf, cells("count mean(fmt(2)) sd(fmt(2)) min(fmt(2)) max(fmt(2))") nomtitle nonumber
(output written to summarystat1.rtf)

. xtset country year
       panel variable:  country (strongly balanced)
        time variable:  year, 2007 to 2014
                delta:  1 unit


. by country: generate lagy = y[_n-1]
(12 missing values generated)

. generate lny=ln(y)

. generate lnlagy=ln(lagy)
(12 missing values generated)

. generate gdpgplus1=(lny-lnlagy)*100
(12 missing values generated)

. generate gdpg=gdpgplus1[_n+1]
(12 missing values generated)

. generate lninv=ln(inv)
(12 missing values generated)

. generate deltang=deltag+n
(12 missing values generated)

. generate lndeltang=ln(deltang)
(12 missing values generated)

. generate lnhc=ln(hc)
(12 missing values generated)

. generate lneduc=ln(educ)
(12 missing values generated)
```

```
. generate lnerdf=ln(erdf)
(12 missing values generated)

. generate lncf=ln(cf)
(12 missing values generated)

. generate lnesf=ln(esf)
(12 missing values generated)

. generate lnCP=ln(CP)
(12 missing values generated)

. generate CPI_lnCP=cpi*ln(CP)
(12 missing values generated)

. generate CPI_CP=cpi*CP
(12 missing values generated)

. generate CC_lnCP=CC*ln(CP)
(12 missing values generated)

. generate CC_CP=CC*CP
(12 missing values generated)

. generate CPI_lnERDF=cpi*ln(erdf)
(12 missing values generated)

. generate CPI_ERDF=cpi*erdf
(12 missing values generated)

.
. *___
. *the following regression is our basic model
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.2631                         Obs per group: min =         7
       between = 0.8820                                        avg =       7.0
       overall = 0.3053                                        max =         7

                                                Wald chi2(7)       =     92.67
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -2.598304    3.28834    -0.79   0.429    -9.043332    3.846724
       lninv |  -12.82564   2.829987    -4.53   0.000    -18.37231   -7.278967
      lneduc |    4.29096   2.949449     1.45   0.146    -1.489855    10.07177
   lndeltang |  -3.068929   1.675014    -1.83   0.067    -6.351896    .2140386
         cpi |  -1.545963   .5905125    -2.62   0.009    -2.703346   -.3885798
        lnCP |  -15.08345    3.75681    -4.01   0.000    -22.44666   -7.720233
    CPI_lnCP |   2.795113   .6412648     4.36   0.000     1.538258    4.051969
       _cons |   70.47681   32.03957     2.20   0.028     7.680399    133.2732
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  4.8272316
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m1, title(Model 1)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.63
         Prob > chi2 =    0.0000

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.71
         Prob > chi2 =    0.0000
```

```
. *the following regression is our basic model with CC instead of CPI
. xtreg gdpg lny lninv lneduc lndeltang CC lnCP CC_lnCP if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.2359                         Obs per group: min =         7
       between = 0.7505                                        avg =       7.0
       overall = 0.2655                                        max =         7

                                                Wald chi2(7)       =    263.74
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -6.121598   3.530158    -1.73   0.083    -13.04058    .7973852
       lninv |  -12.68323   3.991625    -3.18   0.001    -20.50667   -4.859794
      lneduc |   3.387739   4.150986     0.82   0.414    -4.748043    11.52352
   lndeltang |  -3.280847   1.958839    -1.67   0.094    -7.120102     .558408
          CC |  -.4652801   1.334258    -0.35   0.727    -3.080379    2.149818
        lnCP |   -2.52041   1.659419    -1.52   0.129    -5.772811    .7319919
    CC_lnCP |   4.322801   1.623133     2.66   0.008      1.14152    7.504082
       _cons |   99.36574   38.98863     2.55   0.011     22.94943     175.782
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  4.8818746
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m2, title(Model 2)

. test CC CC_lnCP

 ( 1)  CC = 0
 ( 2)  CC_lnCP = 0

           chi2(  2) =    7.22
         Prob > chi2 =    0.0270

. test lnCP CC_lnCP

 ( 1)  lnCP = 0
 ( 2)  CC_lnCP = 0

           chi2(  2) =   19.38
         Prob > chi2 =    0.0001

. *the following regression is our basic model with hc instead of educ
. xtreg gdpg lny lninv lnhc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), vce (cluster country)


Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.2683                         Obs per group: min =         7
       between = 0.8482                                        avg =       7.0
       overall = 0.3047                                        max =         7

                                                Wald chi2(7)       =     74.40
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.026205   4.936719    -0.61   0.540      -12.702    6.649586
       lninv |  -13.04774   3.806685    -3.43   0.001     -20.5087   -5.586772
        lnhc |   6.031086   13.61597     0.44   0.658    -20.65573     32.7179
   lndeltang |  -3.448659   1.982554    -1.74   0.082    -7.334394    .4370767
         cpi |  -1.459716   .6440393    -2.27   0.023     -2.72201   -.1974223
        lnCP |  -14.32177   3.887979    -3.68   0.000    -21.94207   -6.701475
    CPI_lnCP |   2.646841   .7011387     3.78   0.000     1.272635    4.021048
       _cons |   79.08603   42.29679     1.87   0.062    -3.814162    161.9862
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  4.7829185
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. estimates store m3, title(Model 3)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   15.41
         Prob > chi2 =    0.0005

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   14.25
         Prob > chi2 =    0.0008

. *the following regression is our basic model with ERDF instead of CP
. xtreg gdpg lny lninv lneduc lndeltang cpi lnerdf CPI_lnERDF if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.2592                         Obs per group: min =         7
       between = 0.8778                                        avg =       7.0
       overall = 0.3041                                        max =         7

                                                Wald chi2(7)       =     63.82
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                               (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -1.690656   3.288852    -0.51   0.607    -8.136689    4.755376
       lninv |  -11.53178   2.884883    -4.00   0.000    -17.18605   -5.877515
      lneduc |   3.703457   2.835408     1.31   0.192    -1.853841    9.260756
   lndeltang |  -1.807066   1.462207    -1.24   0.217     -4.67294    1.058808
         cpi |   .1150324    .545802     0.21   0.833    -.9547197    1.184785
      lnerdf |    -11.086   2.510042    -4.42   0.000    -16.00559   -6.166408
  CPI_lnERDF |   2.161053   .4933618     4.38   0.000     1.194082    3.128024
       _cons |   48.12642   32.27024     1.49   0.136    -15.12208    111.3749
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |   4.864023
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m4, title(Model 4)

. test cpi CPI_lnERDF

 ( 1)  cpi = 0
 ( 2)  CPI_lnERDF = 0

           chi2(  2) =   19.23
         Prob > chi2 =    0.0001


. test lnerdf CPI_lnERDF

 ( 1)  lnerdf = 0
 ( 2)  CPI_lnERDF = 0

           chi2(  2) =   19.51
         Prob > chi2 =    0.0001

.
. estout m1 m2 m3 m4 using panel.rtf, cells(b(star fmt(3)) se(par fmt(2))) starlevels (* 0.1 ** 0.05 *** 0.01) ///
> legend label varlabels(_cons constant) ///
> stats(r2_o  N, fmt(3 0) label(R_sq_overall Number_of_obs))
(output written to panel.rtf)
```

```
. *---
. *regressions with fixed country and time effects
. *the following regression is our basic model
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.2631                         Obs per group: min =         7
       between = 0.8820                                        avg =       7.0
       overall = 0.3053                                        max =         7

                                                Wald chi2(7)       =     92.67
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -2.598304    3.28834    -0.79   0.429    -9.043332    3.846724
       lninv |  -12.82564   2.829987    -4.53   0.000    -18.37231   -7.278967
      lneduc |    4.29096   2.949449     1.45   0.146    -1.489855    10.07177
   lndeltang |  -3.068929   1.675014    -1.83   0.067    -6.351896    .2140386
         cpi |  -1.545963   .5905125    -2.62   0.009    -2.703346   -.3885798
        lnCP |  -15.08345    3.75681    -4.01   0.000    -22.44666   -7.720233
    CPI_lnCP |   2.795113   .6412648     4.36   0.000     1.538258    4.051969
       _cons |   70.47681   32.03957     2.20   0.028     7.680399    133.2732
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  4.8272316
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m5, title(Model 1)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.63
         Prob > chi2 =    0.0000

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.71
         Prob > chi2 =    0.0000

. *the following regression is our basic model with FE
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), fe vce (cluster country)


Fixed-effects (within) regression               Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.3446                         Obs per group: min =         7
       between = 0.6488                                        avg =       7.0
       overall = 0.1833                                        max =         7

                                                F(7,11)            =     14.70
corr(u_i, Xb)  = -0.9200                        Prob > F           =    0.0001

                                (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |   -18.6919   7.111981    -2.63   0.023    -34.34526   -3.038535
       lninv |  -12.87994   4.319627    -2.98   0.012    -22.38737   -3.372506
      lneduc |  -38.93008   34.60759    -1.12   0.285    -115.1009    37.24071
   lndeltang |  -9.563735   6.226203    -1.54   0.153    -23.26752    4.140046
         cpi |  -.9441008   1.944553    -0.49   0.637    -5.224033    3.335831
        lnCP |  -12.62241   5.990281    -2.11   0.059    -25.80693    .5621073
    CPI_lnCP |   2.910143   1.250102     2.33   0.040     .1586883    5.661598
       _cons |   338.7113   96.41148     3.51   0.005     126.5111    550.9115
-------------+----------------------------------------------------------------
     sigma_u |  6.8476011
     sigma_e |  4.8272316
         rho |  .66802158   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. estimates store m6, title(Model 2)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

       F(  2,    11) =    2.73
            Prob > F =    0.1089

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

       F(  2,    11) =    2.88
            Prob > F =    0.0989

. *the following regression is our basic model with time effects
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP i.year if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.6840                         Obs per group: min =         7
       between = 0.8689                                        avg =       7.0
       overall = 0.6997                                        max =         7

                                                Wald chi2(11)      =         .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =         .

                               (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -.1773299   2.778616    -0.06   0.949    -5.623316    5.268657
       lninv |  -8.381044   2.316373    -3.62   0.000    -12.92105   -3.841037
      lneduc |   1.128557   3.658312     0.31   0.758    -6.041604    8.298717
   lndeltang |  -5.454174   2.515354    -2.17   0.030    -10.38418   -.5241718
         cpi |  -1.287749    .350168    -3.68   0.000    -1.974066   -.6014326
        lnCP |  -12.79248   2.876463    -4.45   0.000    -18.43024   -7.154714
    CPI_lnCP |   2.348934   .4790046     4.90   0.000     1.410102    3.287765
             |
        year |
       2008  |  -11.50773   2.216174    -5.19   0.000    -15.85135   -7.164107
       2009  |  -.4199837   1.025913    -0.41   0.682    -2.430736    1.590768
       2010  |  -1.805047   1.409582    -1.28   0.200    -4.567777    .9576822
       2011  |  -2.707459   1.138913    -2.38   0.017    -4.939687   -.4752311
       2012  |  -3.668192   1.007697    -3.64   0.000    -5.643242   -1.693143
       2013  |  -3.292623   1.172121    -2.81   0.005    -5.589939   -.9953078
             |
       _cons |   46.25471   23.98479     1.93   0.054    -.7546134    93.26404
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  3.3322866
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m7, title(Model 3)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   32.73
         Prob > chi2 =    0.0000

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   27.75
         Prob > chi2 =    0.0000

. *the following regression is our basic model with FE and time effects
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP i.year if inrange(year, 2007, 2013), fe vce (cluster country)
```

```
Fixed-effects (within) regression               Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.7165                         Obs per group: min =         7
       between = 0.5725                                        avg =       7.0
       overall = 0.4475                                        max =         7

                                                F(11,11)           =         .
corr(u_i, Xb)  = -0.7742                        Prob > F           =         .

                              (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -13.50378   19.17744    -0.70   0.496    -55.71305    28.70548
       lninv |  -5.046409   4.296676    -1.17   0.265    -14.50333    4.410512
      lneduc |  -31.17133   28.90707    -1.08   0.304    -94.79536     32.4527
   lndeltang |  -13.36972   5.529128    -2.42   0.034    -25.53925   -1.200193
         cpi |  -.0616941   1.887511    -0.03   0.975    -4.216078     4.09269
        lnCP |  -10.54351   4.013009    -2.63   0.024    -19.37608   -1.710939
     CPI_lnCP |   2.083689   .8642223     2.41   0.035     .1815487     3.98583
             |
        year |
        2008 |  -10.86712   2.361545    -4.60   0.001    -16.06484   -5.669391
        2009 |  -.0482953    1.16392    -0.04   0.968    -2.610066    2.513475
        2010 |  -.6532827   2.879845    -0.23   0.825    -6.991779    5.685214
        2011 |   -.578871   3.759002    -0.15   0.880    -8.852379    7.694637
        2012 |  -1.388848   3.767357    -0.37   0.719    -9.680744    6.903049
        2013 |  -.5153919   4.199904    -0.12   0.905    -9.759318    8.728535
             |
       _cons |   248.6099   181.5549     1.37   0.198    -150.9897    648.2095
-------------+----------------------------------------------------------------
     sigma_u |  4.9136837
     sigma_e |  3.3322866
         rho |    .684975   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m8, title(Model 4)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

       F(  2,    11) =    3.03
            Prob > F =    0.0894

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

       F(  2,    11) =    3.45
            Prob > F =    0.0686

. estout m5 m6 m7 m8 using panel3.rtf, cells(b(star fmt(3)) se(par fmt(2))) starlevels (* 0.1 ** 0.05 *** 0.01) ///
> legend label varlabels(_cons constant) ///
> stats(r2_o  N, fmt(3 0) label(R_sq_overall Number_of_obs))
(output written to panel3.rtf)


. *Hausman test
. quietly xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013)

. estimates store random

. quietly xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), fe

. estimates store fixed

. hausman fixed random, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
         lny |    -18.6919    -2.598304       -16.09359        8.340283
       lninv |   -12.87994    -12.82564       -.0542998        3.929943
      lneduc |   -38.93008      4.29096       -43.22104        39.74926
   lndeltang |   -9.563735    -3.068929       -6.494806        6.472024
         cpi |   -.9441008    -1.545963        .6018622        1.513419
        lnCP |   -12.62241    -15.08345        2.461034         4.19087
    CPI_lnCP |    2.910143     2.795113        .1150298        .8166652
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg
```

```
    Test:  Ho:  difference in coefficients not systematic

                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        9.92
                Prob>chi2 =      0.1933

.
. *___
. *regressions with lags
. by country: gen lnCPlag1=lnCP[_n-1]
(12 missing values generated)

. by country: gen lnCPlag2=lnCP[_n-2]
(24 missing values generated)

. by country: gen lnCPlag3=lnCP[_n-3]
(36 missing values generated)

. generate CPI_lnCPlag1=cpi*lnCPlag1
(24 missing values generated)

. generate CPI_lnCPlag2=cpi*lnCPlag2
(36 missing values generated)

. generate CPI_lnCPlag3=cpi*lnCPlag3
(48 missing values generated)

. *the following regression is our basic model
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        84
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.2631                         Obs per group: min =         7
       between = 0.8820                                        avg =       7.0
       overall = 0.3053                                        max =         7

                                                Wald chi2(7)       =     92.67
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -2.598304    3.28834    -0.79   0.429    -9.043332    3.846724
       lninv |  -12.82564   2.829987    -4.53   0.000    -18.37231   -7.278967
      lneduc |    4.29096   2.949449     1.45   0.146    -1.489855    10.07177
   lndeltang |  -3.068929   1.675014    -1.83   0.067    -6.351896    .2140386
         cpi |  -1.545963   .5905125    -2.62   0.009    -2.703346   -.3885798
        lnCP |  -15.08345    3.75681    -4.01   0.000    -22.44666   -7.720233
    CPI_lnCP |   2.795113   .6412648     4.36   0.000     1.538258    4.051969
       _cons |   70.47681   32.03957     2.20   0.028     7.680399    133.2732
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  4.8272316
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m9, title(Model 1)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.63
         Prob > chi2 =    0.0000

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.71
         Prob > chi2 =    0.0000

. *the following regression is our basic model with 1 lag of CP
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCPlag1 CPI_lnCPlag1 if inrange(year, 2007, 2013), vce (cluster country)
```

```
Random-effects GLS regression                   Number of obs      =        72
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.4205                         Obs per group: min =         6
       between = 0.6896                                        avg =       6.0
       overall = 0.4115                                        max =         6

                                                Wald chi2(7)       =     83.75
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.399272   3.148191    -1.08   0.280    -9.569613    2.771068
       lninv |  -17.76917   6.315569    -2.81   0.005    -30.14746   -5.390879
      lneduc |   11.41407   7.730908     1.48   0.140    -3.738232    26.56637
   lndeltang |   .8723558   2.858944     0.31   0.760    -4.731071    6.475782
         cpi |  -1.691413   .7587187    -2.23   0.026    -3.178474   -.2043517
     lnCPlag1 |  -10.57849   7.081141    -1.49   0.135    -24.45727    3.300295
CPI_lnCPlag1 |   2.102811     1.1648     1.81   0.071    -.1801553    4.385778
       _cons |   69.91331   38.94032     1.80   0.073    -6.408323    146.2349
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  4.5012915
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m10, title(Model 2)

. test cpi CPI_lnCPlag1

 ( 1)  cpi = 0
 ( 2)  CPI_lnCPlag1 = 0

           chi2(  2) =    6.90
         Prob > chi2 =    0.0318

. test lnCPlag1 CPI_lnCPlag1

 ( 1)  lnCPlag1 = 0
 ( 2)  CPI_lnCPlag1 = 0

           chi2(  2) =   10.78
         Prob > chi2 =    0.0046

. *the following regression is our basic model with 2 lags of CP
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCPlag2 CPI_lnCPlag2 if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        60
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.0725                         Obs per group: min =         5
       between = 0.7941                                        avg =       5.0
       overall = 0.4921                                        max =         5

                                                Wald chi2(7)       =    219.06
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -4.042377   3.915621    -1.03   0.302    -11.71685    3.632099
       lninv |  -.9919812   3.136919    -0.32   0.752    -7.140229    5.156266
      lneduc |   3.953599   8.195671     0.48   0.630    -12.10962    20.01682
   lndeltang |   -5.91593   3.029885    -1.95   0.051     -11.8544    .0225352
         cpi |   .1636062   .3044326     0.54   0.591    -.4330708    .7602831
    lnCPlag2 |  -2.800544   3.838869    -0.73   0.466    -10.32459    4.723501
CPI_lnCPlag2 |   .6307718     .62423     1.01   0.312    -.5926965     1.85424
       _cons |   44.85014   34.03847     1.32   0.188    -21.86403    111.5643
-------------+----------------------------------------------------------------
     sigma_u |  .23888538
     sigma_e |   1.806174
         rho |  .01719209   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. estimates store m11, title(Model 3)

. test cpi CPI_lnCPlag2

 ( 1)  cpi = 0
 ( 2)  CPI_lnCPlag2 = 0

           chi2(  2) =    1.74
         Prob > chi2 =    0.4196

. test lnCPlag2 CPI_lnCPlag2

 ( 1)  lnCPlag2 = 0
 ( 2)  CPI_lnCPlag2 = 0

           chi2(  2) =    4.12
         Prob > chi2 =    0.1272

. *the following regression is our basic model with 3 lags of CP
. xtreg gdpg lny lninv lneduc lndeltang cpi lnCPlag3 CPI_lnCPlag3 if inrange(year, 2007, 2013), vce (cluster country)

Random-effects GLS regression                   Number of obs      =        48
Group variable: country                         Number of groups   =        12

R-sq:  within  = 0.1745                         Obs per group: min =         4
       between = 0.8022                                        avg =       4.0
       overall = 0.5681                                        max =         4

                                                Wald chi2(7)       =    175.86
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for 12 clusters in country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lny |  -3.997088    4.98898    -0.80   0.423    -13.77531    5.781133
       lninv |   .0852784   1.141967     0.07   0.940    -2.152935    2.323492
      lneduc |   6.235453   8.116113     0.77   0.442    -9.671837    22.14274
   lndeltang |  -7.339874   3.098766    -2.37   0.018    -13.41334   -1.266405
         cpi |   .1567158   .5540926     0.28   0.777    -.9292857    1.242717
    lnCPlag3 |  -.7840417   4.119302    -0.19   0.849    -8.857725    7.289641
CPI_lnCPlag3 |   .2432464   .7673878     0.32   0.751    -1.260806    1.747299
       _cons |   37.47546   37.60316     1.00   0.319    -36.22538    111.1763
-------------+----------------------------------------------------------------
     sigma_u |  .41106528
     sigma_e |   1.546972
         rho |   .0659517   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store m12, title(Model 4)

. test cpi CPI_lnCPlag3

 ( 1)  cpi = 0
 ( 2)  CPI_lnCPlag3 = 0

           chi2(  2) =    0.65
         Prob > chi2 =    0.7235


 ( 1)  lnCPlag3 = 0
 ( 2)  CPI_lnCPlag3 = 0

           chi2(  2) =    0.67
         Prob > chi2 =    0.7149

. estout m9 m10 m11 m12 using panel_lag.rtf, cells(b(star fmt(3)) se(par fmt(2))) starlevels (* 0.1 ** 0.05 *** 0.01) ///
> legend label varlabels(_cons constant) ///
> stats(r2_o  N, fmt(3 0) label(R_sq_overall Number_of_obs))
(output written to panel_lag.rtf)

.
. *___
. *GMM regressions
. *Arrelano-Bond with restricted instruments
. xtabond gdpg lny lninv lneduc lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), lags(1) maxldep(1) vce(robust) artests(2)
```

```
Arellano-Bond dynamic panel-data estimation  Number of obs         =        60
Group variable: country                      Number of groups      =        12
Time variable: year
                                             Obs per group:    min =         5
                                                               avg =         5
                                                               max =         5

Number of instruments =      13              Wald chi2(8)          =    398.82
                                             Prob > chi2           =    0.0000
One-step results
                                 (Std. Err. adjusted for clustering on country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gdpg |
         L1. |   -.155334   .0864524    -1.80   0.072    -.3247776    .0141096
             |
         lny |  -28.41183   8.622807    -3.29   0.001    -45.31222   -11.51143
       lninv |  -22.08697   5.708402    -3.87   0.000    -33.27523   -10.89871
      lneduc |   40.00978   59.81479     0.67   0.504    -77.22505    157.2446
   lndeltang |  -8.606975    5.57798    -1.54   0.123    -19.53962    2.325666
         cpi |  -2.746985   1.408931    -1.95   0.051    -5.508438    .0144686
        lnCP |   -8.37544   4.188438    -2.00   0.046    -16.58463   -.1662534
     CPI_lnCP |  2.018915   .7335218     2.75   0.006     .5812385    3.456591
       _cons |    278.035   148.4619     1.87   0.061      -12.945     569.015
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/2).gdpg
        Standard: D.lny D.lninv D.lneduc D.lndeltang D.cpi D.lnCP D.CPI_lnCP
Instruments for level equation
        Standard: _cons

. estat abond
artests not computed for one-step system estimator with vce(gmm)

Arellano-Bond test for zero autocorrelation in first-differenced errors
  +-----------------------+
  |Order |  z     Prob > z|
  |------+----------------|
  |   1  |2.0656  0.0389 |
  |   2  |.24612  0.8056 |
  +-----------------------+
   H0: no autocorrelation

. estimates store m13, title(Model 1)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   10.94
         Prob > chi2 =    0.0042

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   11.64
         Prob > chi2 =    0.0030

. *Arrelano-Bond with restricted instruments and endogenous variables
. xtabond gdpg lneduc lndeltang cpi if inrange(year, 2007, 2013), lags(1) maxldep(1) endog(lny lninv lnCP CPI_lnCP, lagstruct(0,1)) vce(robust) artests(2)

Arellano-Bond dynamic panel-data estimation  Number of obs         =        60
Group variable: country                      Number of groups      =        12
Time variable: year
                                             Obs per group:    min =         5
                                                               avg =         5
                                                               max =         5

Number of instruments =      29              Wald chi2(8)          =    762.80
                                             Prob > chi2           =    0.0000
One-step results
                                 (Std. Err. adjusted for clustering on country)
------------------------------------------------------------------------------
             |               Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gdpg |
         L1. |   -.076715   .1273103    -0.60   0.547    -.3262386    .1728085
             |
         lny |  -11.49204   13.18733    -0.87   0.384    -37.33874    14.35465
       lninv |  -25.37244   6.732986    -3.77   0.000    -38.56885   -12.17603
        lnCP |  -.8398101   12.57029    -0.07   0.947    -25.47712     23.7975
    CPI_lnCP |   1.625065   2.168281     0.75   0.454    -2.624688    5.874819
      lneduc |  -99.02284   62.88884    -1.57   0.115    -222.2827    24.23701
   lndeltang |   -13.8076   7.528293    -1.83   0.067    -28.56278    .9475826
         cpi |  -1.878881   1.662713    -1.13   0.258    -5.137737    1.379976
       _cons |   462.1324   134.8491     3.43   0.001     197.8329    726.4318
------------------------------------------------------------------------------
```

```
Instruments for differenced equation
        GMM-type: L(2/2).gdpg L(2/2).lny L(2/2).lninv L(2/2).lnCP L(2/2).CPI_lnCP
        Standard: D.lneduc D.lndeltang D.cpi
Instruments for level equation
        Standard: _cons

. estat abond
artests not computed for one-step system estimator with vce(gmm)

Arellano-Bond test for zero autocorrelation in first-differenced errors

  +-----------------------+
  |Order |  z     Prob > z|
  |------+----------------|
  |   1  |-1.1189  0.2632 |
  |   2  | .06593  0.9474 |
  +-----------------------+
   H0: no autocorrelation

. estimates store m14, title(Model 2)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =    2.18
         Prob > chi2 =    0.3362

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =   22.77
         Prob > chi2 =    0.0000

. *Blundell-Bover with restricted instruments
. xtdpdsys gdpg lny lninv lneduc  lndeltang cpi lnCP CPI_lnCP if inrange(year, 2007, 2013), lags(1) maxldep(1) twostep vce(robust) artests(2)

System dynamic panel-data estimation         Number of obs         =        72
Group variable: country                      Number of groups      =        12
Time variable: year
                                             Obs per group:    min =         6
                                                               avg =         6
                                                               max =         6

Number of instruments =     18               Wald chi2(8)          =    454.93
                                             Prob > chi2           =    0.0000
Two-step results

------------------------------------------------------------------------------
             |              WC-Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gdpg |
         L1. |   .0044208    .279366     0.02   0.987    -.5431264    .5519681
             |
         lny |  -30.88871   22.42862    -1.38   0.168    -74.84799    13.07058
       lninv |  -23.46814   12.02429    -1.95   0.051     -47.0353    .0990325
      lneduc |   12.28863   111.5363     0.11   0.912    -206.3186    230.8958
   lndeltang |   12.45742   28.68991     0.43   0.664    -43.77377    68.68861
         cpi |  -3.686883   3.164361    -1.17   0.244    -9.888916     2.51515
        lnCP |  -7.320309   42.31349    -0.17   0.863    -90.25323    75.61261
    CPI_lnCP |   1.944636   7.605716     0.26   0.798    -12.96229    16.85157
       _cons |   346.3759   320.8222     1.08   0.280    -282.4239    975.1758
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/2).gdpg
        Standard: D.lny D.lninv D.lneduc D.lndeltang D.cpi D.lnCP D.CPI_lnCP
Instruments for level equation
        GMM-type: LD.gdpg
        Standard: _cons

. estat abond

Arellano-Bond test for zero autocorrelation in first-differenced errors

  +-----------------------+
  |Order |  z     Prob > z|
  |------+----------------|
  |   1  |  .3946  0.6931 |
  |   2  | .26829  0.7885 |
  +-----------------------+
   H0: no autocorrelation
```

```
. estimates store m15, title(Model 3)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =    1.42
         Prob > chi2 =    0.4916

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =    1.52
         Prob > chi2 =    0.4671

. *Blundell-Bover with restricted instruments and endogenous variables
. xtdpdsys gdpg lneduc lndeltang cpi if inrange(year, 2007, 2013), lags(1) maxldep(1) twostep endog(lny lninv lnCP CPI_lnCP, lagstruct(0,1)) vce(robust) artests(2)

System dynamic panel-data estimation         Number of obs         =        72
Group variable: country                      Number of groups      =        12
Time variable: year
                                             Obs per group:    min =         6
                                                               avg =         6
                                                               max =         6

Number of instruments =     54               Wald chi2(8)          =     34.31
                                             Prob > chi2           =    0.0000
Two-step results
------------------------------------------------------------------------------
             |              WC-Robust
        gdpg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gdpg |
         L1. |  -.1141655   .2041867    -0.56   0.576     -.514364     .286033
             |
         lny |  -17.85558   25.30018    -0.71   0.480    -67.44303    31.73186
       lninv |   -24.7817   24.25162    -1.02   0.307    -72.31399    22.75059
        lnCP |   .3669201   60.43242     0.01   0.995    -118.0785    118.8123
    CPI_lnCP |   .0990308   9.123206     0.01   0.991    -17.78212    17.98019
      lneduc |   19.48309   117.5436     0.17   0.868    -210.8982    249.8644
   lndeltang |   1.838968   22.69001     0.08   0.935    -42.63264    46.31058
         cpi |  -1.847222   4.217298    -0.44   0.661    -10.11297     6.41853
       _cons |   212.7795    136.189     1.56   0.118    -54.14615    479.7051
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/2).gdpg L(2/2).lny L(2/2).lninv L(2/2).lnCP L(2/2).CPI_lnCP
        Standard: D.lneduc D.lndeltang D.cpi
Instruments for level equation
        GMM-type: LD.gdpg LD.lny LD.lninv LD.lnCP LD.CPI_lnCP
        Standard: _cons

. estat abond


Arellano-Bond test for zero autocorrelation in first-differenced errors
  +-----------------------+
  |Order |  z     Prob > z|
  |------+----------------|
  |   1  |.27878  0.7804  |
  |   2  |.03571  0.9715  |
  +-----------------------+
   H0: no autocorrelation

. estimates store m16, title(Model 4)

. test cpi CPI_lnCP

 ( 1)  cpi = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =    0.34
         Prob > chi2 =    0.8444

. test lnCP CPI_lnCP

 ( 1)  lnCP = 0
 ( 2)  CPI_lnCP = 0

           chi2(  2) =    0.01
         Prob > chi2 =    0.9947

.
. estout m13 m14 m15 m16 using panel_gmm.rtf, cells(b(star fmt(3)) se(par fmt(2))) starlevels (* 0.1 ** 0.05 *** 0.01) ///
> legend label varlabels(_cons constant) ///
> stats(N N_g arm1 arm2 sargan, fmt(0 0) label(Number_of_obs Number_of_groups AR(1) AR(2) Sargan))
(output written to panel_gmm.rtf)
```