

DE GRUYTER
OPEN

The role of entrepreneurship in explaining the real Gross Domestic Product per capita: Regression model selection

Anita Harmina

*VERN University of Applied Sciences, Zagreb, Croatia
anita.harmina@vern.hr*

Abstract

In this research, the impact of total early-stage entrepreneurial activity and competitiveness of the economy on the real gross domestic product (GDP) per capita is analyzed in a cross-section of world economies using the methods of correlation and multiple regression analysis. In the attempt to select between the linear and the double-logarithmic model, the regression diagnostics and quality of the relationship between the dependent and the independent variables were analyzed. The functional form of the model was tested by the MacKinnon, White and Davidson test. Model selection methods regarding the comparison of coefficients of determination and the Akaike information criterion were used. The results of the analysis show that independent variables have a statistically significant impact on the real GDP per capita, and that the real GDP per capita is elastic to the changes of competitiveness but inelastic to the changes of total early-stage entrepreneurial activity.

Keywords: Akaike information criterion, correlation, double-logarithmic regression model, entrepreneurship, multiple linear regression model.

JEL classification: C31, L26, M13, O11.

DOI: 10.1515/crebss-2016-0007

Received: April 20, 2016

Accepted: August 25, 2016

Introduction

Among the long list of variables that have been considered as the determinants of economic growth (Sala-i-Martin, 1997), in the last two decades the focus has increasingly turned to entrepreneurship (Holcombe, 1997, Wennekers, Thurik, 1999, Carree, Thurik, 2003, Audretsch, Keilbach, 2004, van Stel, Carree, Thurik, 2005, Wennekers, van Stel, Thurik, 2005, Wong, Ho, Autio, 2005, Acs, 2006, Audretsch, 2007, Stem, van Stel, 2009, Valliere, Peterson, 2009, Martin, Picazo, Navarro, 2010). There is ever growing scientific evidence that entrepreneurial activities affect employment, productivity and economic development (Wennekers, Thurik, 1999, Carree, Thurik, 2003). Therefore, the institutional environment that helps with the development and realization of entrepreneurship is crucial for the economic growth (Holcombe, 1997).

Ever since entrepreneurship has been conceptualized by the Global Entrepreneurship Monitor (GEM) project in the 1997, empirical data on entrepreneurship has enabled the exploration of the relationship between

entrepreneurial activities and the level of economic development (Carree, Thurik, 2003, Acs, 2006, Audretsch, 2007). The results of the research concerning the relationship between GEM's entrepreneurship data and economic development show that this link depends upon the motives that drive individuals into entrepreneurship (necessity and opportunity motivated entrepreneurship), and that the level of economic development in addition to the motives for entrepreneurship determines the resulting effects on economic growth (van Stel, Carree, Thurik, 2005, Wennekers et al., 2005). Also, the research showed that only some types of entrepreneurial activities like nascent, start-up or entrepreneurial activity of employees have a positive relationship with economic growth, but that this relationship again is very sensitive and varies according to the development phases (van Stel et al., 2005, Wennekers et al., 2010). Entrepreneurship does not contribute to economic growth in less developed countries, in contrast to developed countries where high-growth entrepreneurship seems to have a strong effect on economic growth rates (Stam, van Stel, 2009, Wong, Ho, Autio, 2005, Valliere, Peterson, 2009). However, high-growth entrepreneurship is very rare so it is important to explain individual and contextual level of appearance of this type of entrepreneurship (Stam et al., 2012).

Economic development can be explained by a long list of other important factors. Some of those are education, macroeconomic environment, investments, quality of institutions, technology and others. However, those factors are not independent in their nature, so some of them can simultaneously be significant (Schwab, Sala-i-Martin, 2011). Ever since 2005, the analysis of competitiveness relies on the World Economic Forum's (WEF's) Global Competitiveness Index (GCI). This index is "a comprehensive tool that measures the microeconomic and macroeconomic foundations of national competitiveness" (Schwab, Sala-i-Martin, 2011, p. 4). Since the small size of the sample does not allow the inclusion of many independent variables in this research, the GCI is used in the models as a variable that combines those other factors important for explaining economic development.

Hence, based on the GEM's indicator of Total early-stage Entrepreneurial Activity (TEA) and WEF's data on GCI, the aim of this paper is to analyze whether this indicator of entrepreneurial activity could be considered a statistically significant factor of economic development (measured as the level of real GDP per capita), and to investigate and test the shape of their relationship. More formally, the analysis in this research starts from the following hypothesis: Total early-stage entrepreneurial activity and competitiveness have a statistically significant and double-logarithmic relationship with the real GDP per capita. The methodology underlying this research is based on the methods of multiple regression analysis with the emphasis on the analysis of linear versus double logarithmic regression model.

After brief introduction, in the second chapter of this paper the relevant literature is elaborated. In the third chapter the data and methods are described. The regression and model selection results along with the discussions are presented in the fourth chapter. The fifth chapter ends the paper with final conclusions.

Literature review

The GEM project was originally started with the objective of exploring the complex mechanism that relates entrepreneurship and economic development (Carree, Thurik, 2003, Acs, 2006, Audretsch, 2007), where entrepreneurship is defined as "any attempt at new business or new venture creation, such as self-employment, a new business organization, or the expansion of an existing business, by an individual, a team of individuals, or an established business" (Bosma et al., 2012, p. 20). Hence,

the GEM scientists defined the conceptual framework that determines the key elements of the link between entrepreneurship and economic development and the channels through which these elements interact. That model is based upon the concept that according to the phase of development of a country (factor-driven phase, efficiency-driven phase or innovation-driven phase (Porter, Sachs, McArthur, 2002)), the impact of entrepreneurship on economic development differs (Wennekers et al., 2005). The research also showed that this impact varies due to the motives for entering entrepreneurship: opportunity or necessity motivated entrepreneurship (Acs, Arenius, Minniti, 2005, Wennekers et al., 2005). In countries with higher necessity motivated entrepreneurship rates, the GDP per capita is lower, while the opposite conclusion stands for opportunity-motivated entrepreneurship (Kelley, Bosma, Amoros, 2011).

Wennekers et al. (2005) analyzed the GEM 2002 data for 36 countries that participated in the survey that year and found that nascent entrepreneurship has a U-shaped relationship (second-degree polynomial) with economic development (which they measured by per capita index or by an innovative capacity index). Opportunity and necessity entrepreneurship explain this U-shaped relationship (Wennekers et al., 2005). Wong, Ho and Autio (2005) also analyzed the 2002 GEM data set for 37 countries based on the adjusted Cobb-Douglas production function. Their results show that only high-potential entrepreneurial activity influences economic growth rates.

Acs, Arenius and Minniti (2005) found a U-shaped link between TEA rate and per capita GDP (expressed in US\$) for 2004 GEM data collected from 34 economies. This U-shaped relationship shows that the TEA rates are highest in factor-driven economies with lowest GDP per capita and that they decline as GDP per capita rises and economies enter the efficiency-driven phase of development. After reaching certain GDP per capita level, TEA rates start to rise as the per capita GDP further increases (in innovation-driven economies). Using GEM 's Total early-stage Entrepreneurial Activity rates for the year 2007, Wennekers et al. (2010) found a U-shaped relationship between TEA rate and per capita income to be statistically significant.

However, the results of the GEM research based on data for 54 countries in 2011 do not support the U-shaped relationship of TEA and GDP per capita (expressed in purchasing power parity (PPP) \$). The positive correlation for countries in innovation-driven phase of development (the right hand tail of the U-shape curve) is not present in the 2011 data; instead the TEA rate is now declining as the GDP per capita level raises (Bosma, Wennekers, Amoros, 2012). Harmina, Dumičić and Cingula (2014) analyzed the GEM data for 2010 for 54 countries in the sample and found a double-logarithmic relationship of TEA rate with the real GDP per capita (expressed in 2005 international PPP\$) for 2011. For this log-log model they reported the $R^2=0.521$.

Recent research showed that TEA rates are highest for countries in the factor-driven phase of development, and that they decrease as the per capita GDP raises (Amoros, Bosma, 2014, Singer, Amoros, Moska, 2015). Based on the 2014 GEM data Singer, Amoros and Moska (2015) report a linear-log model with $R^2=0.3742$ for the relationship between TEA rate and GDP per capita in PPP\$.

The results of empirical research indicate that the relationship between entrepreneurship and economic development moves from positive to negative and varies in shape throughout the years, but that all these results very much depend upon the types of entrepreneurial activities, the motivations for these activities and the level of economic development.

Research methodology

Data

For the purpose of empirical research, the data on three variables has been collected: the real GDP per capita, the total early-stage entrepreneurial activity and the global competitiveness index (Table 1). With respect to the year of observation for the real GDP per capita as the dependent variable, the values used for independent variables (total early-stage entrepreneurial activity and global competitiveness index) are one year time lagged. The author thus simply assumes a short-run impact of changes in entrepreneurship and competitiveness values on the real GDP per capita.

Table 1 The list of variables

Variable name	Variable description	Variable units	Data source	Sample size (n)
Y_{GDPpc}	real GDP per capita for 2011	international 2005 PPP\$	The World Development Indicators (World Bank, 2013)	173
X_{TEA}	total early-stage entrepreneurial activity for 2010	% of 18-64 population	Key Indicators (Global Entrepreneurship Research Association, 2016)	59
X_{GCI}	global competitiveness index for 2010	1-to-7 (best) scale	The Global Competitiveness Report 2010/2011 (Schwab, Sala-i-Martin, 2011)	139

Source: author's work

The main independent variable used for analysis is the total early-stage entrepreneurial activity. This variable measures the country's entrepreneurial activity through entrepreneurial activity of nascent entrepreneurs and new business owner-managers. Nascent entrepreneurs are individuals engaged in the process of starting a new business that paid salaries or wages for up to three months. Owners-managers of new business are those individuals whose entrepreneurial activity is measured by the payment of wages or salaries for the period from 3-42 months. The total early-stage entrepreneurial activity (the TEA rate) is then the ratio of the total number of those two types of entrepreneurs to the size of the sample of the adult population aged 18-64 years (Global Entrepreneurship Research Association, 2016).

Additional independent variable, the global competitiveness index (GCI) represents the weighted average of a number of different factors, all of which are important for competitiveness from the different perspective. Those factors are organized in twelve pillars of competitiveness (Schwab, Sala-i-Martin, 2011): institutions, infrastructure, macroeconomic environment, health and primary education, higher education and training, goods market efficiency, labor market efficiency, financial market development, technological readiness, market size, business sophistication and innovation. Even though all these pillars of competitiveness are in a certain amount important for all economies, they influence different economies in different ways due to the stages of development of these countries. Therefore, for the calculation of the GCI the phases of development are taken into consideration: higher the importance of pillars for a particular phase, higher the weights (Schwab, Sala-i-Martin, 2011).

Methods

The influence of chosen independent variables on the real GDP per capita in 2011 is analyzed by the means of correlation analysis and multiple regression analysis. The correlation analysis is based on the Pearson correlation coefficient (Wooldridge,

2013). The regression parameters were estimated by ordinary least squares (OLS) method for the linear multiple regression model given by:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{ij}, \quad i = 1, 2, \dots, n, \quad (1)$$

and, also for the double logarithmic (double-ln, or log-log) regression model (Asteriou, 2006) given by:

$$y = \beta_0 \cdot \prod_{j=1}^k x_j^{\beta_j} \cdot e^{\varepsilon} \Rightarrow \ln y = \ln \beta_0 + \sum_{j=1}^k \beta_j \ln x_j + \varepsilon. \quad (2)$$

In search for the model that gives a better fit to the data, hypothesis testing and methods of model selection analysis were used. For both linear and double-logarithmic model the regression diagnostics (Asteriou, 2006) was performed. Multicollinearity of the independent variables was tested with the average variance inflation factor (VIF) (Wooldridge, 2013). Serial correlation of residuals was tested with the Breusch-Godfrey serial correlation test (Asteriou, 2006). Further, the White test was used to test for (pure) heteroscedasticity of residuals and to test for model specification error (Gujarati, 2004). If cross-product terms are excluded from the White test procedure, then it is a test of pure heteroscedasticity, and if cross-product terms are included, then it is a test of both heteroscedasticity and specification bias (Harris, 1995). The normality of residuals was tested by the Jarque-Bera test (Maddala, Lahiri, 2009). Finally, based on the overall F-test, the statistical significance of each regression model was tested.

Further, the functional form of the regression was tested by the MacKinnon, White, and Davidson test (MWD test) (Gujarati, 2004, Verbeek, 2012). In this test, both the linear and double logarithmic models are first estimated by OLS to obtain the estimates \hat{y}_i and $\widehat{\ln y}_i$. Then the linear model is tested against its double logarithmic alternative by assuming the null hypothesis $H_0: \delta_{LIN} = 0$ (i.e. the model is linear) for the test regression

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \delta_{LIN} Z_1 + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (3)$$

where $Z_1 = (\ln \hat{y}_i - \widehat{\ln y}_i)$. One should reject H_0 if the coefficient δ_{LIN} is statistically significant by the usual t test. Alternatively, for the double logarithmic model the test assumes the null hypothesis $H_0: \delta_{LOG} = 0$ (i.e. the model is double logarithmic) for the test regression

$$\ln y_i = \ln \beta_0 + \sum_{j=1}^k \beta_j \ln x_{ij} + \delta_{LOG} Z_2 + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (4)$$

where $Z_2 = (\hat{y}_i - \exp(\widehat{\ln y}_i))$.

Similarly, one should reject H_0 if the coefficient δ_{LOG} is statistically significant by the usual t test. The possible problem with using MWD test is that in a given situation either hypothesis cannot be rejected.

The problem of choosing between two alternative specifications of the model was also analyzed regarding the model selection criteria. When choosing among candidate models (for comparing purposes and/or for forecasting purposes), it is important to distinguish between in-sample forecasting performance of a model (fit of the data in a given sample) and out of-sample forecasting performance of a model (forecasts of future values based on the fitted model). Usually, the R2 and adjusted R2, Akaike information criterion (AIC), Schwarz Information criterion (SIC), Mallow's Cp criterion, and forecast χ^2 (chi-square) are used as model selection criteria and are all based on minimizing the residual sum of squares (Gujarati, 2004).

Model selection techniques that are used for out-of-sample model estimation usually refer to cross-validation methods (Maddala, Lahiri 2009). In this research the model selection analysis was conducted based on the comparison of coefficients of determination and Akaike info criterion.

The comparison of the coefficients of determination for linear and double logarithmic model is based on the following procedure (Gujarati, 2004), described in two steps. In the first step the value $\exp(\widehat{\ln y_i})$ from double-ln model for each observation is obtained. Then the R^2 between these values and actual y_i according to the equation

$$R^2 = \frac{(\sum y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i^2)} \quad (5)$$

is computed. This R^2 value can now be compared to the R^2 value of the linear model. Alternatively, in the second step, assuming all y values are positive, $\ln y_i$ is computed and the values for $\ln \hat{y}_i$ from the linear model are obtained, and finally the R^2 between these values in the manner indicated by the equation 5 is computed. This R^2 value can now be compared to the R^2 value of the double-ln model.

The Akaike information criterion allows comparison and ranking of competing models and estimation of the best approximate of the "true" process underlying the data. Based on information theory, the Akaike information criterion (AIC) (Akaike, 1973) measures the relative distance of models to the "truth" model for a given set of data. AIC is calculated as (Burnham, Anderson, 2002)

$$AIC = -2 \log L + 2k, \quad (6)$$

where k is the number of estimated parameters, and L is the maximum value of the likelihood function for the model. AIC represents the compromise between bias and variance (goodness of fit of the model and the complexity of the model). In the special case of ordinary least squares (OLS) estimation, AIC can be expressed as a simple function of the residual sum of squares. If all the models in the set assume normally distributed errors with a constant variance, then AIC can be computed as (Gujarati, 2004)

$$AIC = e^{2k/n} \frac{\sum \hat{u}_i^2}{n} = e^{2k/n} \frac{RSS}{n}, \quad (7)$$

where k is the number of estimated parameters (including the intercept and variance) and n is the number of observations. The selection of the "best" model is made based on the lowest value of AIC. AIC is helpful for in-sample and out-of-sample forecasting performance of a regression model. For small sample sizes (n/k is roughly less than 40) a modified version of AIC (AIC_c) is recommended (Hurvich, Tsai, 1989). The AIC_c is given by

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}. \quad (8)$$

In the next chapter the results of the empirical analysis are presented.

Results and discussion

Correlation analysis

The first aim here is to verify the relationship between real GDP per capita in 2011 (in PPP\$) and the chosen explanatory variables. The second aim is to examine the shape of the relationship between the dependent and the independent variables. The third aim is to investigate the possible relationship between total early-stage entrepreneurial activity and global competitiveness index.

All three variables were analyzed in pairs with taking the maximum cross-section of all available observations for the pair under consideration into account. The scatterplot at Figure 1 presents the results of the correlation analysis conducted by

Harmina, Dumičić and Cingula (2014), who analyzed the relationship between real GDP per capita in 2011 (int. 2005 PPP\$) and total early-stage entrepreneurial activity in 2010 for the sample of 54 countries that were found in cross-section of available data on these two variables. They found a strong negative correlation between real GDP per capita and TEA rate and observed the higher coefficient of determination ($R^2 = 0.521$) with the power function used as the functional form of the simple regression model than with the simple linear regression model ($R^2 = 0.357$) (Harmina, Dumičić, Cingula, 2014).

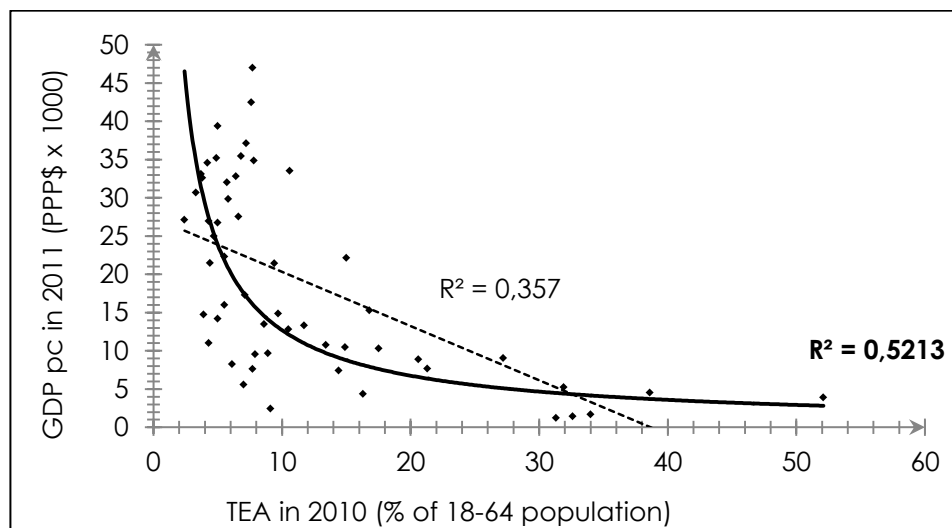


Figure 1 Scatterplot for real GDP per capita (int. 2005 PPP\$) and total early-stage entrepreneurial activity

Source: Harmina, Dumičić, Cingula, 2014.

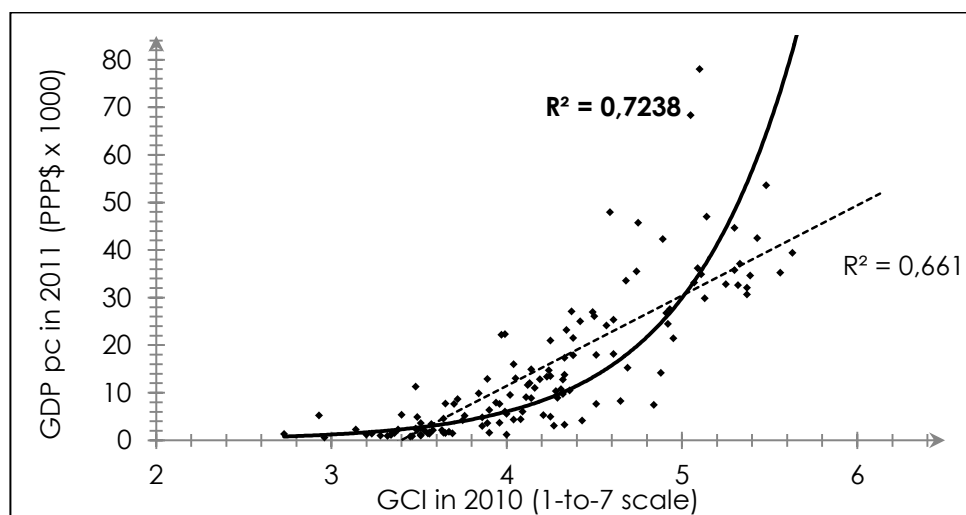


Figure 2 Scatterplot for real GDP per capita (int. 2005 PPP\$) and global competitiveness index

Source: author's work

The scatterplot in Figure 2 shows that the correlation between GCI and real GDP per capita (the cross-section of data on these two variables resulted in sample size $n=128$) could also be double logarithmic (the power function) ($R^2 = 0.724$) rather than linear.

The Pearson correlation coefficients for the natural log values of all pair of variables has also been calculated, taking into account the different sample sizes for different pairs. The results are summarized in Table 2.

Table 2 Correlation coefficients for log of real GDP per capita in 2011 (int. 2005 PPP\$) and logs of values of TEA and GCI

	logGDPpc11	logTEA10	logGCI10
logGDPpc11	1 n=173		
logTEA10	-0.72 n=54	1 n=59	
logGCI10	0.86 n=128	-0.63 n=57	1 n=139

Source: author's work

The linear correlation coefficient between the natural log of real GDP per capita in 2011 and the natural log of the TEA rate in 2010 (-0.72) shows that lower values of real GDP per capita (int. 2005 PPP\$) correspond to the economies with a higher TEA rate. The negative sign of the correlation coefficient for TEA is in line with previous findings and is explained by the different effects that opportunity and necessity-motivated entrepreneurship have on economic development (Acs, Arenius, Minniti, 2005).

Based on their log values, the total early-stage entrepreneurial activity and the global competitiveness index are in a negative correlation (-0.63). This result could mean a problem with multicollinearity of independent variables. Though, Baltagi (2008, p. 76) writes that "Maddala (2001) argues that high intercorrelation among the explanatory variables are neither necessary nor sufficient to cause the multicollinearity problem", and that "in practice, multicollinearity is sensitive to the addition or deletion of observations".

Still, because of this moderately high correlation coefficient, the correlation analysis was conducted for the sample of following 36 countries that were found in cross-section of all available data (all three variables) and that form the total sample size used for the regression analysis: Australia, Belgium, Bosnia & Herzegovina, Brazil, Chile, China, Colombia, Croatia, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Japan, Korea (South), Latvia, Malaysia, Mexico, Netherlands, Pakistan, Peru, Portugal, Romania, Russia, Slovenia, South Africa, Spain, Sweden, Switzerland, Trinidad & Tobago, Turkey, United Kingdom, United States of America and Uruguay. The results are shown in Figures 3 and 4 and in Tables 3 and 4.

Based on the coefficients of determination for linear and double-ln model, from scatterplot in Figure 3 it is obvious that the performance of linear and double logarithmic functional form of the relationship between real GDP per capita and entrepreneurship is almost equally good, only a little in the favor of the double-ln form. Though, this result for the sample of 36 countries compared to the diagram in Figure 1 is much weaker.

Scatterplot in Figure 4 shows that the linear function would be a better choice for describing the relationship between the real GDP per capita and global competitiveness index than the power function. This result is opposite to the result suggested by scatterplot in Figure 2. The coefficients of determination shown in Figure 4 are also much lower than those presented in Figure 2.

It seems that both linear and double logarithmic model could be potentially a good start for modeling the influence of total early-stage entrepreneurial activity

and global competitiveness index on the real GDP per capita for the sample of these 36 countries, but that double logarithmic functional form of the model would be a better choice if a larger sample would be available.

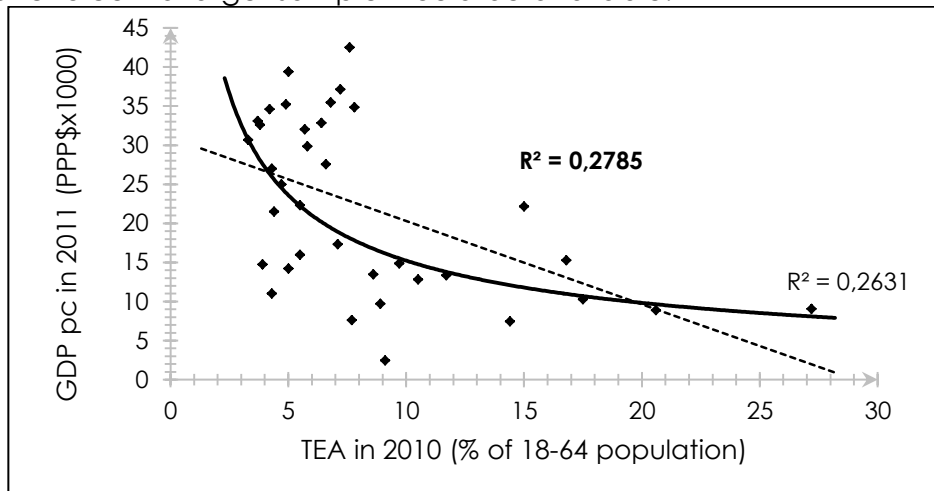


Figure 3 Scatterplot for the sample of 36 countries for real GDP per capita in 2011 (int. 2005 PPP\$) and TEA rate

Source: author's work

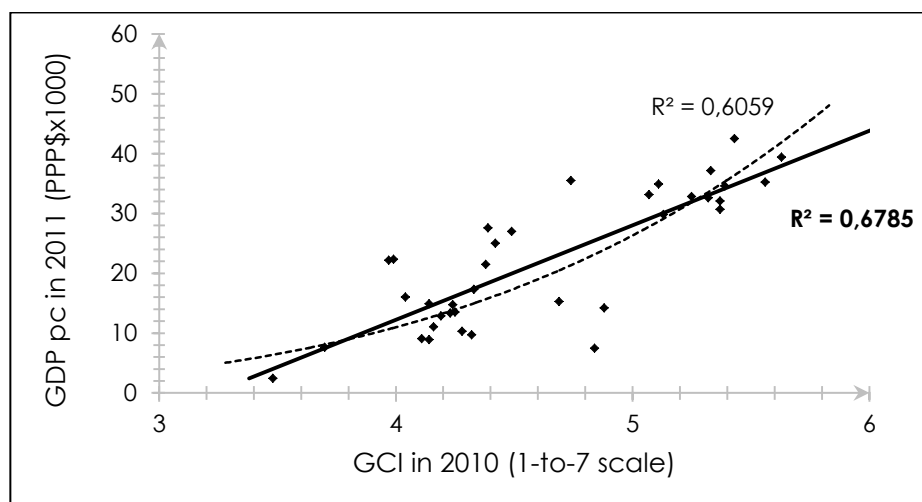


Figure 4 Scatterplot for the sample of 36 countries for real GDP per capita in 2011 (int. 2005 PPP\$) and GCI

Source: author's work

By calculating the Pearson's correlation coefficients for both competing model forms, the potential multicollinearity issue observed in larger sample is also addressed within the final smaller sample (n= 36). The linear correlation coefficients for original values are presented in Table 3 and for their log transformed values in Table 4.

Table 3 Correlation matrix for original values of all variables (n=36)

	GDPpc11	TEA10	GCI10
GDPpc11	1		
TEA10	-0.51	1	
GCI10	0.82	-0.37	1

Source: author's work

Table 4 Correlation matrix for logs of values of all variables (n=36)

	logGDPpc11	logTEA10	logGCI10
logGDPpc11	1		
logTEA10	-0.53	1	
logGCI10	0.78	-0.41	1

Source: author's work

From the correlation coefficient -0.37 (Table 3) between TEA rate and the GCI index in original values and from the coefficient -0.41 (Table 4) between the log of TEA and log of GCI, it can be seen that within this sample it is unlikely to expect serious problems with multicollinearity of independent variables.

Regression analysis results

The multiple regression analysis as well as the regression model evaluation and regression diagnostics (Asteriou, 2006) was conducted for linear and double-logarithmic model for the sample of 36 countries. The results are shown in Table 5.

Table 5 Results of linear and double-logarithmic regression for real GDP per capita in 2011 (int. 2005 PPP\$) and diagnostics tests

Model		Dependent variable	
		Y _{GDPpc11}	ln Y _{GDPpc11}
		linear	double-ln
Independent variables (SE)	C	-38831.38* (9483.84)	5.2540* (0.9949)
	X _{TEA10}	-497.23** (203.40)	-
	ln X _{TEA10}	-	-0.3007** (0.1337)
	X _{GCI10}	14073.45* (1874.91)	-
	ln X _{GCI10}	-	3.4005* (0.5615)
	R ²	0.728	0.658
	\bar{R}^2	0.711	0.638
	V _y	27.13	3.88
	F (p-value)	44.12 (0.0000)	31.78 (0.0000)
	n	36	36
	k	2	2
VIF average		1.161	1.202
Breusch-Godfrey test (lags=2) (p-value)		0.5317 (0.7666)	1.1854 (0.5528)
White test - included cross-terms (p-value)		5.0699 (0.4074)	9.5843 (0.0879)
White test - excluded cross-terms (p-value)		0.0673 (0.9669)	4.1265 (0.1270)
Jarque-Bera test (p-value)		0.6832 (0.7106)	5.1136 (0.0776)
Akaike info criterion		20.31021	0.989442

Source: author's work. Note: Significant at 1%(*), 5%(**), 10%(***).

In both regression models, all regression coefficients are statistically significant at 5% significance level. The results of the diagnostics tests show that both models satisfy the regression assumptions (Asteriou, 2006).

Multicollinearity of the independent variables does not exist since the average VIF is smaller than 5 in linear as well as in double-ln model. The Breusch-Godfrey serial correlation test was performed and the results show that the null-hypothesis that the serial correlation of residuals up to order 2 does not exist might not be rejected at the

5% significance level (p-values for test statistic in both models are higher than 0.05). The White test was used to test for (pure) heteroscedasticity of residuals and to test for model specification error. The White test results with cross-terms excluded show that at the 5% significance level the null-hypothesis might not be rejected, so there is no pure heteroscedasticity in either model. Also, with cross-terms included in the test procedure, at 5% significance level no heteroscedasticity and specification bias was found in linear, as well as in double-In model. Based on the Jarque-Bera test it was determined that at the 5% significance level the null-hypothesis that the residuals are normally distributed might not be rejected. Finally, based on the overall F-test, both regression models are statistically significant. Hence, both models have good diagnostics so neither model can be eliminated from further analysis in this research.

In the next paragraph the functional form of the regression model was tested in order to investigate which model fits the data better.

For the purpose of choosing between a linear regression model and a double logarithmic regression model the MWD test was used.

Table 6 MWD test for linear and double-In model

Independent variables (SE; p-value)	Dependent variable	
	$Y_{GDPpc11}$	$\ln Y_{GDPpc11}$
	linear model	double-In model
C	-36478.83* (10488.38; 0.0015)	4.8885* (1.0289; 0.0000)
X_{TEA10}	-537.44** (218.07; 0.0193)	-
$\ln X_{TEA10}$	-	-0.3178** (0.1333; 0.0232)
X_{GCI10}	13685.97* (2020.56; 0.0000)	-
$\ln X_{GCI10}$	-	3.6444* (0.5900; 0.0000)
Z1	-4417.02 (7994.16; 0.5844)	-
Z2	-	-0.00003559 (0.00002847; 0.2203)
R^2	0.730	0.674
\bar{R}^2	0.705	0.644
F (p-value)	28.90 (0.0000)	22.07 (0.0000)
n	36	36
k	3	3

Source: author's work. Note: Significant at 1%(*), 5%(**), 10%(***).

The results of the MWD test (see Table 6) show that the null hypothesis that the model is linear cannot be rejected (the coefficient of the variable Z1 is not statistically significant since the p-value of the estimated t equals 0.5844). On the other hand, the hypothesis that the model is double-In also cannot be rejected (p-value for parameter next to variable Z2 equals 0.2203). Based on the MWD test the decision which functional form of the model is better cannot be made.

In the next subchapter the choice between two alternative specifications of the model was conducted based on the model selection criteria.

Regression model selection

In this research, the appropriate regression model selection was conducted based on the comparison of coefficients of determination and Akaike info criterion.

For comparison of two models based on the coefficient of determination (R^2) the assumption that the sample size n and the dependent variable are the same must be fulfilled, while the independent variables can take any form (Gujarati, 2004). Thus for linear and double logarithmic models the computed R^2 terms cannot be compared directly (i.e. conclusion that the $R^2_{LIN} = 0.728 > R^2_{LOG} = 0.658$ cannot be made). The results of coefficients of determination comparison procedures are summarized in the Table 7.

Table 7 Comparing R^2 and V_y between linear and double-ln model

	Dependent variable		
	$Y_{GDPPc11}$	$\ln Y_{GDPPc11}$	
R^2 based on $Y_{GDPPc11}$ and $\hat{Y}_{GDPPc11}$	0.728 (linear)	0.658 (double-ln)	R^2 based on $\ln Y_{GDPPc11}$ and $\widehat{\ln Y}_{GDPPc11}$
Procedure 1 R^2 based on $Y_{GDPPc11}$ and $\exp(\widehat{\ln Y}_{GDPPc11})$	0.698 (double-ln)	0.667 (linear)	Procedure 2 R^2 based on $\ln Y_{GDPPc11}$ and $\ln(\hat{Y}_{GDPPc11})$

Source: author's work

Computation of R^2 for double-ln model by following the procedure 1 resulted in $R^2 = 0.698$ that is now comparable to the original $R^2 = 0.728$ of the linear model. Since the R^2 value for the linear model is a little bit higher than that derived for the double-ln model, it seems that linear model gives a better fit. Alternatively, by following procedure 2 the $R^2_{LIN} = 0.667$ was obtained, which is higher than original $R^2_{LOG} = 0.658$. Hence, using either procedure, seems that the linear model gives a slightly better fit. However, since the coefficient of determination measures in-sample goodness of fit, there is no guarantee that (based on the relatively high value of coefficient of determination) the "better" model will forecast well out-of-sample observations.

Models can be compared using the AIC (AIC_c) criterion only when they have been estimated for exactly the same data set (Burnham, Anderson, 2002). Therefore, in our research the AIC for linear (20.31021) and AIC for double-ln model (0.989442) cannot be directly compared.

In the spirit of a Cobb-Dougllass production function and economic development theory, the usage of double logarithmic transformations of variables when modelling productivity levels of economies is supported by economic theory (Samuelson, Nordhaus, 2007). The correlation analysis results performed in this research showed that the relationship between real GDP per capita and independent variables in a bigger sample could indeed be double logarithmic. Further, from the statistical point of view the analysis results that are in favour of the linear model are only slightly better than those in the favour of the double-ln model. Taking all this into account, the final choice for modelling the relationship between entrepreneurship, competitiveness and the real GDP per capita is the double logarithmic model given in Table 5.

For the double-ln model, the regression coefficient $\hat{\beta}_1 = -0.3007$ implies that a 1% increase in the TEA rate (while keeping the GCI value constant) is followed by the average decrease in the real 2011 per capita GDP regression value for 0.3007%. Hence, the real per capita GDP is inelastic to the changes of the TEA rate (keeping GCI constant).

The regression coefficient $\hat{\beta}_1 = 3.4005$ shows that if the global competitiveness index value would increase for 1% (without changing the TEA rate) the real GDP per capita would on average react with an increase of 3.4005%. This result implies that the real GDP per capita is elastic to the changes of the GCI.

The adjusted coefficient of determination in double-ln model shows that 63.8% of the variation in logs of real GDP per capita is explained by the variations of logs of TEA and GCI.

Conclusions

In this research modeling the relationship between total early-stage entrepreneurial activity, global competitiveness index and real GDP per capita with the methods of multiple regression analysis was supported by the correlation analysis. The results of the correlation analysis show that both linear and double logarithmic model form are possible in the sample used for regression, but that double logarithmic model would be a better choice when bigger sample is analyzed. The regression analysis resulted in two statistically valid models: the linear and double-ln model. It has been found that both variables, TEA and GCI, are statistically significant at 5% significance level in each model, and that both models satisfy regression assumptions. Further the functional form of the model was tested in order to determine which of the two models gives a better fit to the data. For that purpose the MWD test was used. The results of the MWD test showed that the hypothesis that the model is linear cannot be rejected. Also, the test showed that the hypothesis that the model is double logarithmic cannot be rejected as well. Therefore, the analysis was continued with model selection methods. First, the procedure that compares the coefficients of determination of linear and double-ln model in case when the dependent variable in those two models comes in different forms was conducted. It has been found that the R^2 value for the linear model is higher than the R^2 value of double-ln model. This led to the conclusion that the linear model rather than the double logarithmic model explains the variations in real GDP per capita better, but only slightly better. Second, the usage of Akaike information criterion did not lead to the choice of a "better" model since its values are not comparable across different functional forms of the model. Finally, taking all these statistical results alongside economic theory and previous research on the topic into account, the choice has been made in favor of the double logarithmic regression model. From the estimated double logarithmic model, it was found that TEA has a negative, and GCI has a positive impact on real GDP per capita. This means that if a country's competitiveness remains unchanged and total early-stage entrepreneurial activity rises, on average it causes a decrease in GDP per capita level. On the other hand, if the total early-stage entrepreneurial activity rate in an economy stays the same, but its competitiveness rises, it will (on average) lead to a rise in GDP per capita level. Also, the real GDP per capita is found to be elastic to the changes of global competitiveness index, but inelastic to the changes of total early-stage entrepreneurial activity.

Since the main limitation of this research concerns the limited number of observations and time frame which is actually a point in time, a panel data analysis over a longer time period could be a good start for the future research. Also, in this research the detection and treatment of outliers, leverage points and influence points has been ignored, but should be addressed in the future research. An important issue that is also to be resolved in the future analysis is the predictive quality of a model, so model selection analysis from the perspective of cross-validation techniques should be performed.

References

1. Acs, Z. J. (2006). How Is Entrepreneurship Good for Economic Growth? *Innovations: Technology, Governance, Globalization*, Vol. 1, No. 1, pp. 97-107.
2. Acs, Z. J., Arenius, P., Hay, M., Minniti, M. (2005). *Global Entrepreneurship Monitor 2004 Executive Report*. Babson College, Babson Park, MA, USA; London Business School, London, UK.
3. Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, and F. Csaki, (eds.) *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest. pp. 267-281.
4. Amoros, J. E., Bosma, N. (2014). *Global Entrepreneurship Monitor 2013 Global Report*. Universidad del Desarrollo: Santiago, Chile; Babson College: Babson Park, MA, USA; Universiti Tun Abdul Razak, Kuala Lumpur, Malaysia; London Business School: London, UK.
5. Asteriou, D. (2006). *Applied Econometrics: A Modern Approach using EViews and Microfit*. Palgrave Macmillan, Hampshire.
6. Audretsch, D. B. (2007). Entrepreneurship capital and Economic Growth. *Oxford Review of Economic Policy*, Vol. 23, No. 1, pp. 63-78.
7. Audretsch, D. B., Keilbach, M. (2004). Entrepreneurship capital and economic performance. *Regional Studies*, Vol. 38, No. 8, pp. 949-959.
8. Baltagi, B. H. (2008). *Econometrics*, 4th Ed. Springer-Verlag, Berlin Heidelberg.
9. Bosma, N., Coduras, A., Litovsky, Y., Seaman, J. (2012). *GEM Manual. A report on the design, data and quality control of the Global Entrepreneurship Monitor*. Babson Park, MA, US: Babson College; Santiago, Chile: Universidad del Desarrollo; Kuala Lumpur, Malaysia: Universiti Tun Abdul Razak.
10. Bosma, N., Wennekers, S., Amoros, J. E. (2012). *Global Entrepreneurship Monitor 2011 Extended Report: Entrepreneurs and Entrepreneurial Employees Across the Globe*. Universidad del Desarrollo : Santiago, Chile; Babson College: Babson Park, MA, USA; Universiti Tun Abdul Razak, Kuala Lumpur, Malaysia; London Business School: London, UK.
11. Burnham, K. P., Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.), Springer-Verlag.
12. Carree, M. A., Thurik, A. R. (2003). The impact of entrepreneurship on economic growth. In *Handbook of entrepreneurship research*, Acs, Z. J., Audretsch, D. B. (Eds.), Springer, pp. 437-471.
13. Global Entrepreneurship Research Association (2016). Key Indicators. Available at <http://www.gemconsortium.org/key-indicators> [07 March 2016].
14. Gujarati, D. N. (2004). *Basic Econometrics*. McGraw-Hill, Boston.
15. Harmina, A., Dumičić, K., Cingula, M. (2014). The relationship between types of entrepreneurial activities and GDP per capita. In Bendekovic, M., Klacmer Calopa, M. and Filipovic, D. (Eds.) "*Economic and Social Development*", *Book of Proceedings of the 6th International Scientific Conference on Economic and Social Development and 3rd Eastern European ESD Conference: Business Continuity, Vienna, 25-25 April 2014*. Varazdin, Croatia: Varazdin Development and Entrepreneurship Agency. ISBN: 978-953-6125-10-4. pp. 507-516.
16. Harris, R. (1995). *Using Cointegration Analysis in Econometrics Modelling*. Prentice Hall/Harvester Wheatsheaf, Upper Saddle River.
17. Holcombe, R. G. (1997). Entrepreneurship and Economic Growth. *Quarterly Journal of Austrian Economics*, Vol. 1, No. 2, pp. 45-62.
18. Hurvich, C.M., Tsai, C-L. (1989). Regression and time series model selection in small samples. *Biometrika*, Vol. 76, pp. 297-307.
19. Kelley, D., Bosma, N., Amoros, J. E. (2011). *Global Entrepreneurship Monitor 2010 Global Report*. London Business School, London, Universidad del Desarrollo, Santiago, Babson College, Babson Park.
20. Maddala, G. S., Lahiri, K. (2009). *Introduction to Econometrics*. John Wiley & Sons, New York.
21. Martin, M. A. G., Picazo, M. T. M., Navarro, J. L. A. (2010). Entrepreneurship, income distribution and economic growth. *International Entrepreneurship and Management Journal*, Vol. 6, No. 2, pp. 131-141.

22. Porter, M. E., Sachs, J. J., McArthur, J. (2002) Executive Summary: Competitiveness and Stages of Economic Development. In Porter, M., J. Sachs, P.K. Cornelius, J.W. McArthur, and K. Schwab (Eds), *The Global Competitiveness Report 2001-2002* (pp 16-25). New York: Oxford University Press.
23. Sala-i-Martin, X. X. (1997). I Just Ran Two Million Regressions. *American Economic Review*, Vol. 87, No. 2, pp. 178-183.
24. Samuelson, P. A., Nordhaus, W. D. (2007). *Ekonomija*. Mate, Zagreb.
25. Schwab, K., Sala-i-Martin, X. X. (2011). *The Global Competitiveness Report 2011-2012*. Available at http://www3.weforum.org/docs/WEF_GCR_Report_2011-12.pdf [07 March 2016].
26. Singer, S., Amoros, J. E., Moska, D. (2015). *Global Entrepreneurship Monitor 2014 Global Report*. Universidad del Desarrollo : Santiago, Chile; Babson College: Babson Park, MA, USA; Universiti Tun Abdul Razak, Kuala Lumpur, Malaysia; Tecnológico de Monterrey, Monterrey, Mexico; London Business School: London, UK.
27. Stam, E., Bosma, N., van Witteloostuijn, A., de Long, J., Bogaert, S., Edwards, N., Jaspers, F. (2012). Ambitious Entrepreneurship. A review of the academic literature and new directions for public policy. Available at <http://udoc.eu/reader/604762> [07 March 2016].
28. Stam, E., van Stel, A. (2009). *Types of Entrepreneurship and Economic Growth*. Available at <https://www.wider.unu.edu/sites/default/files/RP2009-47.pdf> [07 March 2016].
29. Valliere D., Peterson R. (2009). Entrepreneurship and economic growth: Evidence from emerging and developed countries. *Entrepreneurship and Regional Development*, Vol. 21, No. 5-6, pp. 459-480.
30. van Stel, A., Carree, M., Thurik, R. (2005). The Effect of Entrepreneurial Activity on National Economic Growth. *Small Business Economics*, Vol. 24, No. 3, pp. 311-321.
31. Verbeek, M. (2012). *A Guide to Modern Econometrics*. John Wiley & Sons, New York.
32. Wennekers, S., Thurik, R. (1999). Linking entrepreneurship and economic growth. *Small business economics*, Vol. 13, No.1, pp. 27-56.
33. Wennekers, S., van Stel, A., Carree, M., Thurik, A. R. (2010). The relationship between entrepreneurship and economic development: Is it U-shaped? *Foundations and Trends in Entrepreneurship*, Vol. 6, No. 3, pp. 167-237.
34. Wennekers, S., van Stel, A., Thurik, R., Reynolds, P. (2005). Nascent entrepreneurship and the level of economic development. *Small Business Economics*, Vol. 24. No. 3, pp. 239-309.
35. Wong, P. K., Ho, Y. P., Autio, E. (2005). Entrepreneurship, Innovation and Economic Growth: Evidence from GEM Data. *Small Business Economics*, Vol. 24, No. 3, pp. 335-350.
36. Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western College Publishing, Mason.
37. World Bank (2013). *GDP per capita, PPP (constant 2005 international \$)*. Available at <http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD> [05 June 2013].

About the author

Anita Harmina is a University Specialist for Statistical methods for economic analyses and forecasting. She finished her university specialist postgraduate study programme at the Faculty of Economics and Business, University of Zagreb, in 2013. Formerly, Anita became the Engineer of Mathematics after finishing the graduate university programme in Financial and Business Mathematics at PMF - Department of Mathematics at University of Zagreb in 2007. Anita works as a higher education lecturer at VERN' University of Applied Sciences in Zagreb since 2007, where she teaches the subjects Business mathematics and Business statistics to undergraduate students. Anita Harmina can be contacted through the LinkedIn or at anita.harmina@vern.hr.