Determinants of New Companies' Formation in Romania at Regional Level. A Fixed Effects Model (FEM) Approach

Dalina-Maria Andrei
Institute of Economic Forecasting, Romania
dalinaandrei@yahoo.com

Abstract

This below article will empirically analyse the determinants of new companies' formation basing on a panel data, covering the period 2010-2019, for 42 counties (including Bucharest municipality) of Romania. An OLS regression equation will be employed with new companies' formation as dependent variable, and regional GDP, existing entrepreneurship, density of population, immigrants and unemployment rate as independent ones. We estimate the model with cross sections and/or with time fixed and/or random effects and we will also use an LSDV model using dummy variables in order to observe the similarities. Decisions in using a period fixed effect model FEM will be based on Hausman test and redundancy test under Eviews environment. Most of determinants considered will be found significant in their influences on new companies' formation in Romanian counties.

Key words: new companies formation, entrepreneurship, fixed effect model, dummy variables
J.E.L. classification: C21,C23,M1, M2, M13

1. Introduction

The impact of exogenous on new companies' formation in Romania will be here analysed at the beginning, using a *pooled model*, in which we assume that there is no any individual effect from cross sections (42 Romanian's counties) and/or period involved – i.e. any county is assumed as similar to the others. The next second step will belong to *fixed and random effect models* employed, according to their opportunity or redundancy. The choice of one or another type of model will be made according to results and will be assessed by variables' significance, correlation coefficient R^2 and Durbin Watson statistic. Cross correlation dependence will not be ignored either.

2. Literature review

Since Schumpeter (1950), entrepreneurship is viewed in the literature as a major topic for the theory and practice related to economic growth and development. Its importance explains by that a significant part of new hirings during a period comes to be done by newly entered companies and this, together with "productive innovation" equally brought in (Baumol, 2002). Therefore, it becomes enough important to understand the factors that promote or mitigate entrepreneurial creativity (Acs, 2004). As already mentioned above, our article bases on analysing the influence of factors proper to the creation of new companies: unemployment rate, regional GDP, number of immigrants, existing entrepreneurship and population density. A series of studies on these determinative links have been conducted in Japan, Bulgaria, the Czech Republic, Germany, Poland and the USA, as well as Romania.

2.1 The endogenous will be, of course, the *newly established companies* at *regional level* in Romania. Two methods of measuring this variable are revealed in the literature. The *ecological* one approaches the new companies as a ratio to the whole mass of existing entrepreneurship. The other is the *labour market* method: the total of existing entrepreneurship relates to the number of people employed in the region. A previous study conducted in Bulgaria, on the country’s 28 territorial districts, mentions these two approaches, then preferring the use of the ecological one (Alexandrova, 2015). Another study conducted in the Czech Republic (Hajek and all., 2015), mentions this variable as the number of new companies to 1000 active individuals and this represents a measure for the entrepreneurial climate in the Czech micro-regions. As for here, we preferred the approach through the labour market - the number of new companies, relative to the active population , resulting in 420 observations (42 counties * 10 years*) for each variable, after which the data turn into logarithms.

2.2 The exogenous. *Existent entrepreneurship* actually is the number of existent entrepreneurs and it is taken as favourable for the new entrepreneurs /new entrepreneurship in the literature. It is the appropriate design of a stable business environment in a country. Otsuka (2008) here similarly sees the Japan’s 43 districts through an ‚economic crowding’ that defines a true entrepreneurship social mentality. Here the existing entrepreneurship is seen through the number of establishments related to the one of population in the same region. Basically, the higher the number of companies with their offices, the more the available capital boosting the rest of resources and factors, here including intelligence, talent and opportunities (Ciccone and Hall, 1996).

Then, it is argued in this study for the mass of entrepreneurship with delayed effect on the newly attached business(Alexandrova, 2015). Plus, this effect will limit to past influencing present and does not go to any influence in the future. A presumably positive relation of the future to the existing environment equals the opportunities opened and business encouraged; the negative one equals the same business opportunities rather embarrassed be it in general or in some of details. Hájek and all. (2015) see the high entrepreneurship’s ratio to population as a proxy for the business climate. A quality entrepreneurial climate can positively influence the individual's decision to become an entrepreneur, and other previous studies come to support such an idea (Armington & Acs, 2002; Delfmann, 2014). Similarly, according to Fotopoulos (2014), new business formation would be influenced by entrepreneurial climate that is supposed to have been already settled in the past.

GDP per capita at regional level. Most empirical studies in this field prefer rather the converse relation, i.e. focusing on new business formation effect on regional development. The empirical results of these studies (Fritsch, 2008). How does new business formation affect regional development? Studies show that the effects of new business formation on economic development are not clear enough. Only few of them could provide persuasive evidence of such a positive relationship -- many others fail on this (Fritsch, 2008). On the contrary, the per capita growth as a predictor of new firm formation is found to have a positive effect by Armington & Acs (2002), not too much this way by Lee et all (2004) and even contrary such effect (i.e. of per capita income growth on new firm formation) by Sutaria & Hicks (2004). Back here, in our study the per capita regional GDP is a measure of per capita growth.

Unemployment and unemployment rate. The literature finds unemployment as also influential for the new companies founded or business enlargement. It is here found as a natural labour resource on specific entrepreneurs’ area – i.e. this part of labour is primarily searching for a profit specific to self employment, as primarily compared to unemployment benefits. But in other views the same unemployment rather is negative factor for new companies’ foundation and not only (Delfman, 2014 & Sutaria and Hicks, 2004). Similarly, Fotopoulos (2014) and Bishop (2012) see the unemployment as likely caused by deep structural economic and social causes, the ones equally affecting entrepreneurship. Otsuka (2008) and Hajeck (2015) find the business environment as not compatible with high unemployment.

Storey (1991), Lindh and Ohlsson (1996) note that time-series analyses point to unemployment as positively associated with the creation of new businesses, whereas cross-sectional studies appear to indicate the opposite.

The population density (i.e. inhabitants per square kilometer) adds to determinant factors for new companies born, in the literature’s view. Alexandrova (2015) sees this through ‚savings crowding’. When and where labour and capital do concentrate, on the contrary, specific costs of resources’ and consumers’ distancing lower. Actually, high population in a region means more available labour skilled. Young and educated people around will be also attracted by new business. And so there will be more potential entrepreneurs amongst.

Immigration. Basically, the theory of entrepreneurial choice generally points out positive relationship between the share of immigrants in population and new business formation (Acs, 2004). But just two aspects to be here explained. The one is that according to Romanian Institute of National Statistic the immigrants are assumed as domiciled in Romania, possibly with the change of their former domicile abroad. It is in this last context that most immigrants actually are remarked as former Romanian citizens back to their county and so the other aspect here considered – i.e. more related to our study – would be that such former local emigrants and current immigrants make a good potential source of new entrepreneurs. In a word, when and where we say and note 'immigrants' in our paper below the understanding will limit to former emigrants back home as potential entrepreneurs – i.e. this is for Romania does not belong to usual and traditional immigration countries category.

2. Research methodology

We collected data for variables from ONRC's (National Commerce Romanian Register) data along the 2010-2019 interval and share for those 42 territorial districts. *New companies* as dependent variable is just the number of new companies per year related to the employed population for a better image of the new companies' territorial distribution – namely, this will be new companies to each thousands of employed people in the area (counties). First independent variable in the model, the existing *entrepreneurship* dimension, will equally consist in a number of companies -- i.e. their total number in Romania and by counties each year of our study --, data collected from INSSE (National Institute of Statistics) - i.e. these might be all: legal entities, family business units and/or authorized persons.

As for *by counties GDP, as second dependent variable,* data also come from INSSE in billions of Ron for each of the years (2010-2019) at current prices – i.e. this way a proper GDP volume searched for will need constant/comparable prices and so CPI will be applied on the 2010 basis, then results related to each county's population. Per capita GDP results in each county and year of the interval will be in Ron.

Unemployment, as third dependent variable, will be taken as its rate (unemployment rate) in each of districts by INSSE statistics. *Population density, as fifth dependent variable,* consist in the number of inhabitants per square kilometre and, of course, once more for each of territorial districts (i.e. counties), for which surface stays the same during the whole period analysed. Then, the *immigrants,* the last exogenous here taken, will be seen as their ratio in total population of the same individual county. As such, this ratio is supposed to bring in a plus of distinction-variation of this exogenous throughout the whole country area – i.e. despite the so low weights of immigration in total population all over the country such a differentiation might identify some significance of this variable on our endogenous chosen in the territory. An OLS regression equation will be employed with new companies' formation as dependent variable, and regional GDP, existing entrepreneurship, density of population, immigrants and unemployment rate as independent ones. We will estimate a model with cross sections and/or with time fixed and/or random effects and we will also use an LSDV model using dummy variables in order to find similarities. Decisions in using a period fixed effect model FEM will be based on Hausman test and redundancy test under Eviews environment.

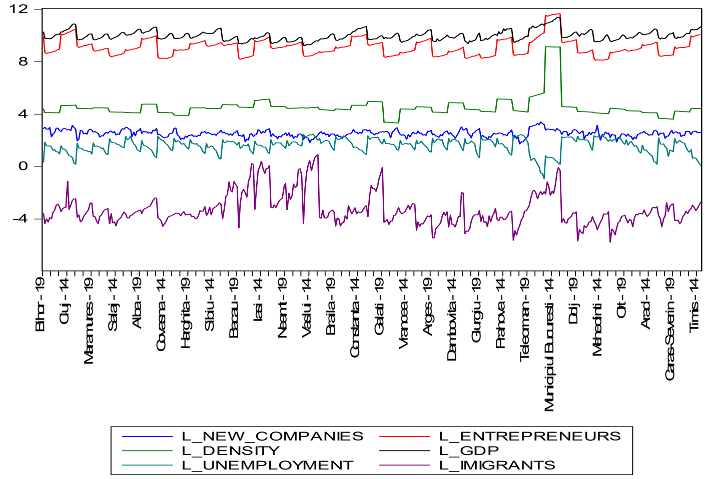
4. Findings: model estimating

Such an analysis to be performed first requires *stationary* and *co-integration* specific tests assigned to data series. The previous of these to be applied assumes the individual series under *Panel unit root* test and *Schwartz Info* criterion for lag length was here apply. This test has null hypothesis and alternative hypothesis:

(a). H0: The panel data of individual variables have unit root (new companies, entrepreneurs, density, immigrants, GDP, unemployment, immigrants);

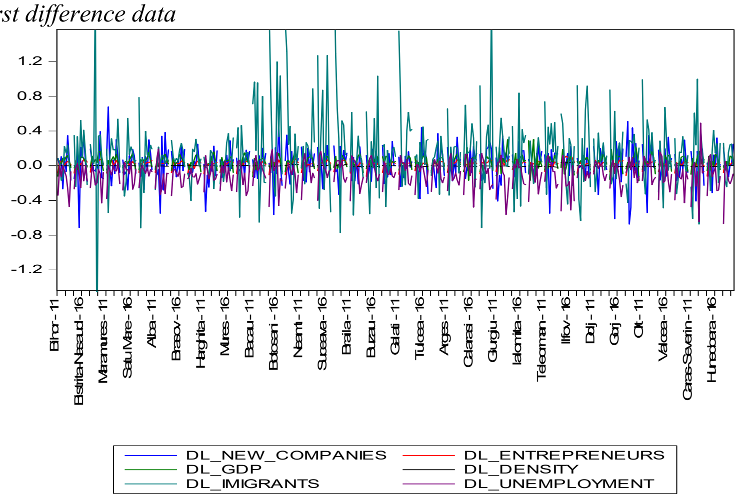
(b). H1: The panel data are stationary for a significance threshold accepted as $\alpha = 5\%$ or 0.05.

Figure no. 1. Level data



Source : Author's own calculations in Eviews

Figure no. 2. First difference data



Source : Author's own calculations in Eviews

We found that our variables have unit root in level data, so they are not stationary at level but they obtain stationarity on first difference. All variables are found to be integrated of order one I(1). The test for co integration was inconclusive. We choose to estimate a model by OLS technique for a static panel testing for fixed or random individual effects.

As already mentioned above in introduction, the estimating method of the regression equation that uses panel data can be done by three approaches: *Common Effect Model* or *Pooled Least Square* (PLS), *Fixed Effect Model* (FEM) and *Random Effect Model* (REM). Let us have them in turn in the following descriptions.

4.1. Pooled Least Square (PLS) model

This type of panel data model assumes homogeneity of all its data in sections – i.e. it does not treat individual sections any differently (Adesete, 2017). All data sections are treated concomitantly, as just one section. No unique characteristic of individuals within the measurement set and no universal effects over time either. *Pooled OLS* makes no difference among the 42 counties in this analyse and neglects both the cross section and the time series nature of given data. See the general form of OLS regression equation for Pooled data (Brooks, 2008):

$y_{it} = \alpha + \beta x_{it} + u$
in which: i = 1, 2, ..., 42 and t = 1, 2, ...,10 are the number of individuals of cross section and the number of time periods and u_{it} is the disturbance term. This is what in our model will be written as:

$DL\_NEW\_COMPANIES_{it}$=C(1)*$DL\_ENTREPRENEURS_{it}$+C(2)*DL_GDP_{it}+C(3)*$DL\_DENSITY_{it}$+ C(4)*$DL\_IMIGRANTS_{it}$ + C(5)*$DL\_UNEMPLOYMENT_{it}$ + C(6)

But ignoring the cross-dimensional and period effects could lead to biased results. We also obtained insignificant values for variables of entrepreneurship, density and immigrants, a low R^2 correlation coefficient (11%) and a Durbin Watson statistic with correlated residuals. Under po*oled OLS* estimation, GDP, unemployment rate and intercept values were found significant(i.e. probability p-value lower than 0,05 significance level). See the table:

Table no. 1 Method: Panel Least Squares, Dependent Variable: DL_NEW_COMPANIES
Periods included: 9, Cross-sections included: 42; Total panel (balanced) observations: 378

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DL_ENTREPRENEURS	-0.270162	0.262731	-1.028283	0.30450
DL_GDP	0.545486	0.134429	4.057802	0.00010
DL_DENSITY	0.276781	1.255295	0.220491	0.82560
DL_IMIGRANTS	0.029441	0.021555	1.365827	0.17280
DL_UNEMPLOYMENT	-0.299947	0.066072	-4.539685	0.00000
C	-0.035754	0.014639	-2.44234	0.01510
R-squared	0.115421	Mean dependent var		0.016415
Adjusted R-squared	0.103532	S.D. dependent var		0.187566
S.E. of regression	0.177591	Akaike info criterion		-0.602921
Sum squared resid	11.73235	Schwarz criterion		-0.540463
Log likelihood	119.9521	Hannan-Quinn criterion.		-0.578133
F-statistic	9.707817	Durbin-Watson stat		2.968418
Prob(F-statistic)	0.00000			

Source: Author's own calculations in Eviews

4.2 Fixed Effect Model (FEM)

This other type of model that could develop here allows for heterogeneity or individuality among different cross-sections – i.e. it allows each cross-section to have its own intercept. So the intercept may be different for the cross sections, but equally time invariant – i.e. the intercept remains the same over time (Adesete, 2017). The error term in a fixed effect model (FEM) is assumed to vary over each entity and each year time. There are unique attributes of individuals which do not vary across time and correlate with independent variables.

See the following specific equation for how FEM works (Brooks, 2008):

$y_{it} = \alpha + \beta^* x_{it} + u_{it}$

in which the disturbance term, u_{it} decomposes into an individual specific effect, μ_i , and the 'remainder disturbance', v_{it} , varies over time and entities (capturing everything that is left unexplained about y_{it}).

$u_{it} = \mu_i + v_{it}$

So we could rewrite equation by substituting in for u_{it} to obtain as follows:

$y_{it} = \alpha + \beta^* x_{it} + \mu_i + v_{it}$

The μ_i term will consist in all of variables that affect Y_{it} cross-sectionally, but do not vary over time. Equation for our panel model with cross section and period fixed effect can be written as follows:

DL_NEW_COMPANIES = C(1)*DL_ENTREPRENEURS + C(2)*DL_GDP + C(3)*DL_DENSITY + C(4)*DL_IMIGRANTS + C(5)*DL_UNEMPLOYMENT + C(6) + [CX=F, PER=F]
in which:
CX=F, fixed cross-section effect
PER=F, period fixed effect

The results in estimating two-way fixed effect model will provide different results. Entrepreneurship environment becomes relevant at county level in influencing the new firms' formation, as expected from the very beginning. The R^2 correlation coefficient rises to 59%. We estimate the fixed effect model with robust standard errors to serial correlation (Arellano,1987and White,1980). We choose the white period as the coefficient covariance method and no degree of freedom for the covariance calculation.

Table no. 2 Model estimation with cross section and period fixed effects

Dependent Variable: DL_NEW_COMPANIES
Method: Panel Least Squares
Date: 03/11/21 Time: 15:08
Sample (adjusted): 2011 2019
Periods included: 9
Cross-sections included: 42
Total panel (balanced) observations: 378
White period standard errors & covariance (no d.f. correction)
WARNING: estimated coefficient covariance matrix is of reduced rank

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DL_ENTREPRENEURS	2.217239	0.597052	3.713647	0.0002
DL_GDP	0.122775	0.120873	1.015736	0.3105
DL_DENSITY	-4.853573	2.553000	-1.901125	0.0582
DL_IMIGRANTS	0.023276	0.019697	1.181710	0.2382
DL_UNEMPLOYMENT	-0.012080	0.043213	-0.279542	0.7800
C	-0.046924	0.013407	-3.500039	0.0005

Effects Specification

Cross-section fixed (dummy variables)
Period fixed (dummy variables)

R-squared	0.590942	Mean dependent var	0.016415
Adjusted R-squared	0.522555	S.D. dependent var	0.187566
S.E. of regression	0.129603	Akaike info criterion	-1.114917
Sum squared resid	5.425416	Schwarz criterion	-0.542380
Log likelihood	265.7193	Hannan-Quinn criter.	-0.887686
F-statistic	8.641098	Durbin-Watson stat	2.586327
Prob(F-statistic)	0.000000		

Source: Author's own calculations in Eviews

And now it is to answer the question whether *fixed effects* here above treated are or not redundant. Or, here the '*Redundant fixed effect likelihood ratio test"* is to be applied. Its H null hypothesis is that fixed effects are redundant and its H alternative hypothesis, on the contrary, is that fixed effects are not redundant. The p value p=0.99 for cross section effects (higher than 0.05 significance level) shows that these cross sections for fixed effect are redundant. Similarly, p = 0.00 for period individual effects, which is lover than 0,05 significance level show that period effects are not redundant.

4.3 Redundant Fixed Effects Tests. Test cross-section and period fixed effects

In its primary steps taken of our study the first difference of data was made in order to deal with no stationary and unit root (see unit root test). Applying the first difference of data, all fixed cross section effects were removed because they do not vary over time (μ_i). The redundancy of cross section fixed effect is tested also with *redundancy fixed effect test.* Redundancy test accounts for separately testing cross-section and period effects and lastly joint significance of all of the effects. This test has the null hypothesis H0: fixed effects are redundant and the alternative hypothesis H1: fixed effect are appropriate. According to *redundancy t*est a *one-way fixed effect model*, with only individual time effects is to be estimated. Under *time-fixed effects model* the average value of "Y" changes over time, but not cross-sectionally. Within time-fixed effects, the intercepts would be allowed to vary over time as such, but assumed to be the same across entities at each given point in time (Brooks, 2008). A *time-fixed effects* model could be written as:

$$y_{it} = \alpha + \beta^* x_{it} + \lambda_t + v_{it}$$

in which λ_t is a *time-varying intercept* that captures all of the variables that affect y_{it} and they vary over time, but stay constant as cross-sectional. In such circumstances, this change of environment will influence *y*, but the same way for all counties. And these last could be assumed, in their turn, to be equally affected by the change.

We check the above individual parameters significance for year influence on dependent variable, through *Wald* tests, where the null hypothesis is: coefficients for years are jointly zero : C(6)…= C(14) =0 We reject the null hypothesis at 1% significance level and confirm that the coefficients are not zero C(6)…= C(14) ≠ 0.

When the *fixed effect model* was estimated through OLS equation an R^2 adjusted correlation coefficient was found as high as 55%. Durbin Watson statistic at the 2.4 level shows a possible absence of autocorrelation at lag 1 of residuals. Estimators of *entrepreneurship* and of *density of population* are significant for 5% confidence level, and the *immigrants*' estimator is significant for 10% of confidence level. The *p-value* probability for intercept is also significant.

4.4 Correlated Random Effects Hausman Test

Robustness of the model was then tested with another version of estimation, as hypothesis of *random time* effects -- i.e. the *Hausman test*. *Random effects* model -- also known as the *variance components* model -- then equally allows for heterogeneity and proves time invariant. Nevertheless, its individual specific effect stays uncorrelated with the independent variables. It can also refer to as a kind of *hierarchical* linear model, assuming data being drawn from several populations made distinct between by a certain hierarchy here referred (Adesete, 2017).

Hausman test keeps the null hypothesis H_0 = *random* effects model is appropriate and the alternative hypothesis H_1 = *fixed* effects model is appropriate. P-value for time random effects is lower than the 0.05 significance level, so the null hypothesis will be rejected and then a confirmation of better approach of a model with an individual time effects, rather than random effects, will correspondingly come up.

Table no. 3 Correlated Random Effects - Hausman Test

Test period random effects			
Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Period random	19.470153	5	0.0016

Source: Author's own calculations in Eviews

4.5 Residual diagnostic - Pesaran CD test

The formal statistical procedures designed to test for *cross-sectional dependence* in low-T, high-N panels, is here the Pesaran (2004) cross-sectional dependence(CD) test. The *CD test* is the *Lagrange multiplier (LM)* test, developed by Breusch & Pagan (1980) often applied when the time-series dimension T of the panel is lower than the cross-sectional dimension N as the case of our data. The null hypothesis Ho: there is no cross sectional dependence and the alternative hypothesis H1: there is cross correlation in the model. *Pesaran CD test* = 0.8066, higher than the 0.05 significance level, so the null hypothesis Ho is accepted: no cross sectional dependence in the model.

Table no. 4 Pesaran CD test

Residual Cross-Section Dependence Test Null hypothesis: No cross-section dependence (correlation) Pool: DATEPANEL Periods included: 9,Cross-sections included: 42,Total panel observations: 378 Note: non-zero cross-section means detected in data Cross-section means were removed during computation of correlations			
Test	**Statistic**	**d.f.**	**Prob.**
Pesaran CD	0.244799		0.8066

Source: Author's own calculations in Eviews

5. Conclusions

Basically, in the panel type model cases estimators are likely substantial – e.g. see both the above given and actually all random effects models. Besides, the same estimators are efficient for fixed effect models. It is in such a context that this article argues in favour of the model with *fixed time effects* for the case of the '*new companies' formation*' facing its proper influence factors – i.e. as here identified: *entrepreneurial climate, regional GDP, immigration* and *unemployment* – since its efficient *estimations* and high *determination coefficient*. These above variables were converted into their natural logarithms to minimize variability and were *first differenced* to get rid of the unit root problem and so to obtain *stationarity*. Then three types of models were developed – (i) *Pool OLS model*, (ii)*Fixed effect model* and (iii) *Random effects model* – and statistical tests employed indicate that the fixed time effect is here appropriate. Actually, there was a double purpose to talk about, from the very beginning of this approach: (a) of course, a model to be shaped as able to find and explain relationship between variables and (b) as such, searching for possible individual effects able to identify differences among Romanian counties -- i.e. to make them really different behaviours from one-another. Time effects could capture the impact of crisis periods, as well as, on the contrary, the new firms' formation phenomenon as specific for the economic recovery. *Redundancy test* was employed to see whether the fixed effects are redundant against *Pooled OLS* model and *Hausman test* helped to choose between the fixed effect and the random effect models.

No specific individuality among the 42 Romanian counties was found – i.e. multiple similarities of these areas prove here implied, e.g. so different time effects similarly affecting all counties at the same during each year. It was this way that the *fixed effects* model with only *time effects* proven the most appropriate in this case.

The cross-sectional dependence *Pesaran CD* test considered appropriate for N>T (number of cross section> years) was employed and the results revealed that the panel variables had not exhibited cross-sectional dependence.

Results of the final model show that existing *entrepreneurship's and density of population coefficients are* significant at the 0.05 level, and *immigration* at the 0.10 level. As expected, the entrepreneurship environment keeps a positive impact on new companies' formation – i.e. a 1% increase of entrepreneurship leads to a 1.75 % increase in new companies' formation. Namely, new companies are likely to prefer stable business environments with business relationships truly shaped and strong.

Immigration, in its turn, has a positive impact on new companies' formation, as expected – but this under the punctual observation that since 2014 the National Institute of Statistic (INSSE) sees as immigrants just the Romanian people back home from abroad and here remaking their main residence. Shortly, those Romanian i*mmigrants* reach a not too high, but positive influence on the new business creation: a 1% increase of immigrants' number will lead to a 0.026% increase in the number of Romanian new business.

As for *density of population, once more,* it keeps a significant, but negative coefficient (i.e. unexpected by the literature): a 1% increase in density of population will lead to a 1.95% decrease a new companies' formation.

Both the *unemployment's* and *regional GDP's* coefficients appear not highly significant – i.e. p-value>0.05 significance levels. Unemployed people, when see themselves constrained to choose between their low but certain benefit and the entrepreneurship, actually get able to notice that the last is rather risky. Regional GDP proves not too significant either for future companies. Besides, each year considered in our model proves its individual impact and so it has been added in equation. Fixed effects were separately highlighted – e.g. in the FEM approach. The R squared determination coefficient is 0.57(adjusted R-squared = 0,55) and expresses that 57% of new companies' formation could be explained by given exogenous: existing entrepreneurs, regional GDP, unemployment, population density and immigration. The rest of 43% identifies the percentage of total variation of endogenous that explains by factors other than those here above considered. The intercept value of (-0.3) represents the intersection between the OY axis and the regression line or the average value of Y endogenous (new companies), the other factors being zero.

6. References

- Adesete, A., 2017. Panel Data Regression Models In Eviews: Pooled OLS, Fixed or Random effect model ?, [online] Available at: https://researchsolutionblog.wordpress.com/2017/12/06/panel-data-regression-models-in-eviews-pooled-ols-fixed-or-random-effect-model/ [Accessed March 2021].
- Alexandrova M., 2015. Effects of Regional Determinants on New Businesses Formation: The Case of Bulgaria. *International Journal of Economic Theory and Application*, 2(4), pp.33-39
- Anderson, T. W., C. Hsiao, 1982. Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics* 18, pp.67–82.
- Arellano, M.,1987. Computing Robust Standard Errors for Within-Groups Estimators*, Oxford Bulletin of Economics and Statistics* 49, pp.431–434.
- Arellano, M., Bond S.R. ,1991. Some Specification Tests for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58, pp.277–298.
- Arellano, M., Bover O., 1995. Another Look at the Instrumental Variables Estimation of Error Component Models. *Journal of Econometrics* 68, pp. 29–51.
- Arellano, M., Honore B.E., 1995. *Panel Data: Some Recent Developments Handbook of Econometrics*, Volume 5, Amsterdam, North Holland.
- Armington C., Acs, Z., 2002.The Determinants of Regional Variation in New Firm Formation, *Regional Studies*, Vol.36.1, pp.33-45.
- Arrelano M., 2003. *Panel Data Econometrics*. Oxford University Press.
- Baltagi, B. H. ,1995. *Econometric Analysis of Panel Data*. New York: Wiley.
- Baltagi B.H., 2008. *Econometrics*. Springer
- Baumol, W J.,2002. Entrepreneurship, Innovation and Growth: The David-Goliath Symbiosis. *Journal of Entrepreneurial Finance and Business Ventures,* 7(2), pp. 1-10.
- Bishop P.,2012. Knowledge, diversity and entrepreneurship: a spatial analysis of new firm formation, *Entrepreneurship & Regional Development,* 24:7-8, pp.641-660.
- Bourlakis, Michael et all, 2014. Firm size and sustainable performance in food supply chains: Insights from Greek SMEs. *International Journal of Production Economics*, Elsevier, vol. 152, pp.112-130.
- Breusch, T. S., G. E. Mizon, P. Schmidt ,1989. Efficient Estimation Using Panel Data*, Econometrica* 57, pp. 695–700.
- Brooks C.,2008. *Introductory Econometrics for Finance*,2nd Edition. New York: Cambridge University Press,
- Chamberlain, G.,1984. Panel Data, *Handbook of Econometrics*, volume 2, pp.1247–1318.
- Ciccone A.,Hall R.E.,1996. Productivity and the Density of Economic Activity. *American Economic Review*, 86, no. 1, pp. 54-70.
- Delfmann H at all, 2014. Population Change and New Firm Formation in Urban and Rural Regions. *Regional Studies, Taylor & Francis Journals,* vol. 48(6), pp.1034-1050.
- Fotopoulos G.,2014. On the spatial stickiness of UK new firm formation rates, *Journal of Economic Geography*, 14(3), pp. 651-679.
- Frees E., 2004, *Longitudinal and Panel Data*. New York: Cambridge University Press.
- Fritsch M., Mueller, P.,2004. Effects of New Business Formation on Regional Development over Time *Regional Studies*, 38(8), pp. 961–975.
- Giannetti M., Simonov A, 2004. On the determinants of entrepreneurial activity: Social norms, economic environment and individual characteristics*, Swedish Economic Policy Review* 11, pp.269-313
- Hájek, O., Nekolová, J., Novosák, J. (2015), Determinants of new business formation – some lessons from the Czech Republic, Economics and Sociology, Vol. 8, No 1, pp. 147-156
- Hausman, J.A., Taylor W.E.,1981. Panel Data and Unobservable Individual Effects*, Econometrica* 49, pp.1377–1398.

- Hsiao, C. ,1986. *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- Hsiao C., 1990, *Analysis of Panel Data*. New York: Cambridge University Press
- Hurlin C., 2002. *L'Econométrie des Données de Panel*. Paris: Université Dauphine,
- Kao, C.,1999. Spurious Regression and Residual-Based Tests for Cointegration in Panel Data*. Journal of Econometrics* 90, pp.1–44.
- Lee, S.Y., Florida, R., Acs, Z.J., 2004. Creativity and entrepreneurship: a regional analysis of new firm formation, *Regional Studies*, 38(8), pp. 879-891.
- Lindh, T. and Ohlsson, D.N.,1996. Self-employment and Windfall Gains: Evidence from the Swedish Lottery, *Economic Journal*, 106, pp.1515-1526
- Storey, D.J.,1991. The birth of new firms. Does unemployment matter? A review of the evidence. *Small Business Economics*, 3, pp.167–178.
- Otsuka A., 2008.Determinants of new firm formation in Japan: A comparison of the manufacturing and service sectors, *Economics Bulletin*, 18(5), pp. 1-7.
- Sutaria, V., Hicks D., 2004. New firm formation: Dynamics and determinants. *Ann Reg Sci* 38, pp.241–262.
- Wooldridge J.M., 2010, *Econometric Analysis of Cross-Section and Panel Data.* 2nd Edition. New York, MIT Press
- Wooldridge J.M., 2008, *Introductory Econometrics: A Modern Approach.* 5th Edition. South-Western, Cengage Learning