9. FORECASTS OF ROMANIAN INDUSTRY EMPLOYMENT USING SIMULATIONS AND PANEL DATA MODELS

Mădălina Ecaterina ANDREICA*
Marin ANDREICA**

Abstract

In this paper, we have built several panel data models at industry level for the period 2000-2011, by considering the impact of various macroeconomic variables, such as labour cost and average gross earnings, upon employment in Romania. A forecasting scenario was then built to forecast the employment in Romania. The results of the econometric analysis were consistent with the empirical evidence, while the analysis offered relevant inside information about the performances of the economic activities of the Romanian industry.

Keywords: simulation, panel data estimation, employment, industry

JEL Classification: C15, C23

I. Introduction

In this paper, we focused on quantifying the impact of labour cost, net investments and average gross earnings upon employment in the Romanian industry. In order to do that, several econometric models were estimated and a forecasting scenario was then built to predict employment in Romania.

The econometric models estimated in this paper were based on panel data series for the period 2000-2011 for 32 economic activities of the Romanian industry, which correspond to the Romanian Nomenclature of Economic Activities (NACE 2), for which data were available.

First, a *hierarchical cluster analysis* was applied, in order to identify the main distinctive groups of economic activity based on macroeconomic performance. Based on the main four industry clusters obtained, the econometric study continued with

* The Bucharest University of Economic Studies, National Scientific Research Institute for Labor and Social Protection, E-mail: madalina.andreica@gmail.com.
** The Bucharest University of Economic Studies, E-mail: marinandreica@gmail.com.

several distinctive panel data models assigned to each of the resulted industry clusters, which implied better prediction performance of employment in the Romanian industry.

The econometric results were consistent with the empirical evidence, which indicated a negative impact, expected in the econometric analysis based on the classic theory, concerning the relationship between the labour costs and wages upon employment (see also Andreica *et al*., 2010; Antonie *et al*., 2010; Seyfried, 2005; Padalino and Vivarelli, 1997; Schaafsma and Walsh, 1983).

Several other studies were made to predict macroeconomic variables (see Albulescu, 2010; Dobrescu, 2010; Nicolae *et al*., 2010; Andreica M. *et al*., 2006; Andreica C. *et al*., 2007). However, the novelty of this paper consists in the way the forecast scenario was structured by considering four main industry clusters, based on the performance of the Romanian economic activities.

The paper is organized as follows: Section 2 describes the data and methodology, while the econometric results are presented in Section 3. In Section 4, the forecasting scenario was designed to predict the employment in Romania, while the last section concludes.

II. Data and Methodology

The panel data analysis was based on annual data for the period 2000- 2011 to predict the Romanian employment in industry. The following variables were used in this study: the number of employed population (*employ*), the average gross earnings (*wage*), the labour force cost (*labor_cost*) and net investments (*Inv*).

The nominal average gross earnings, the net investments and the labour force cost were first deflated using the Consumer Price Index and then the natural logarithm was applied to all the variables in order to ensure higher similarities between the measurement units and more comparative values.

The main data sources were the Romanian National Institute of Statistics and the Ministry of Labour, Family and Social Protection.

The econometric analysis implied several panel data estimations, where the individual effects can be either assumed to be correlated with the explanatory variables as in the case of fixed effects model (FEM) or to be incorporated into the error term as in the case of random effects model (REM) and assumed uncorrelated with the explanatory variables (Baum, 2001).

In our estimation, we first checked whether a panel data model is more appropriate than a simple pooled OLS model, and the results indicated a need for panel data model estimation. We then assumed the presence of fixed effects between the Romanian industrial activities and tested the assumption that a FEM is more appropriate than a REM using the Hausman test. For the FEM, the most used estimator is the “within estimator”.

The panel data model was then estimated assuming that the errors are independent and identically distributed (Cameron and Trivedi, 2009). We then tested for the absence of both heteroskedasticity and serial autocorrelation of the residuals. When heteroskedasticity is present the standard errors of the estimates will be biased and one needs to compute robust standard errors. Another problem is the serial correlation of the idiosyncratic error term, but Wooldridge proposed a simple test to check the autocorrelation of the residuals. In order to overcome these problems, one should estimate the regression model using robust standard errors (Drukker, 2003, Baum, 2001, Green, 2000), which is implemented in the STATA software.

III. Econometric Results

The analysis was conducted over a period of 12 years (2000-2011), using macroeconomic data for 32 economic activities of the Romanian industry. Starting with a basic data description, a cluster analysis was first made to identify the main distinctive groups of economic activity based on macroeconomic performance. This was required in order to see whether there are significant differences between the 32 cross-sections of the panel data structure.

The *hierarchical cluster analysis* is actually a method of unsupervised learning that allows assigning a set of observations into subsets (named clusters) so that the observations in the same cluster are similar.

The cluster technique was built with the Between Groups Linkage cluster method, whereas the intervals were calculated using the squared Euclidean distance. Based on the dendrogram, we notice that the 32 economic activities coded from 1 to 32 can be easily assigned to four main clusters, as shown in Figure 1.

The allocation of the 32 economic activities among the four clusters can be summarized as follows:

- Cluster 1 includes the following economic activities codes: 4, 13, 28, 5, 17, 15, 31, 14, 3 and 8.
- Cluster 2 includes the following economic activities codes: 20, 23, 18, 30, 19, 16, 26, 9, 22, 29, 1, 7 and 2.
- Cluster 3 includes the following economic activities codes: 12, 32, 24, 27, 11 and 21
- Cluster 4 includes the following economic activities codes: 6, 10 and 25.

Since several dissimilarities were noticed between the industrial activities of every two distinct clusters, we decided to structure the econometric analysis into four sections, by considering and building four panel data models corresponding to each of the four resulted industry clusters.

The following general employment equation was considered:

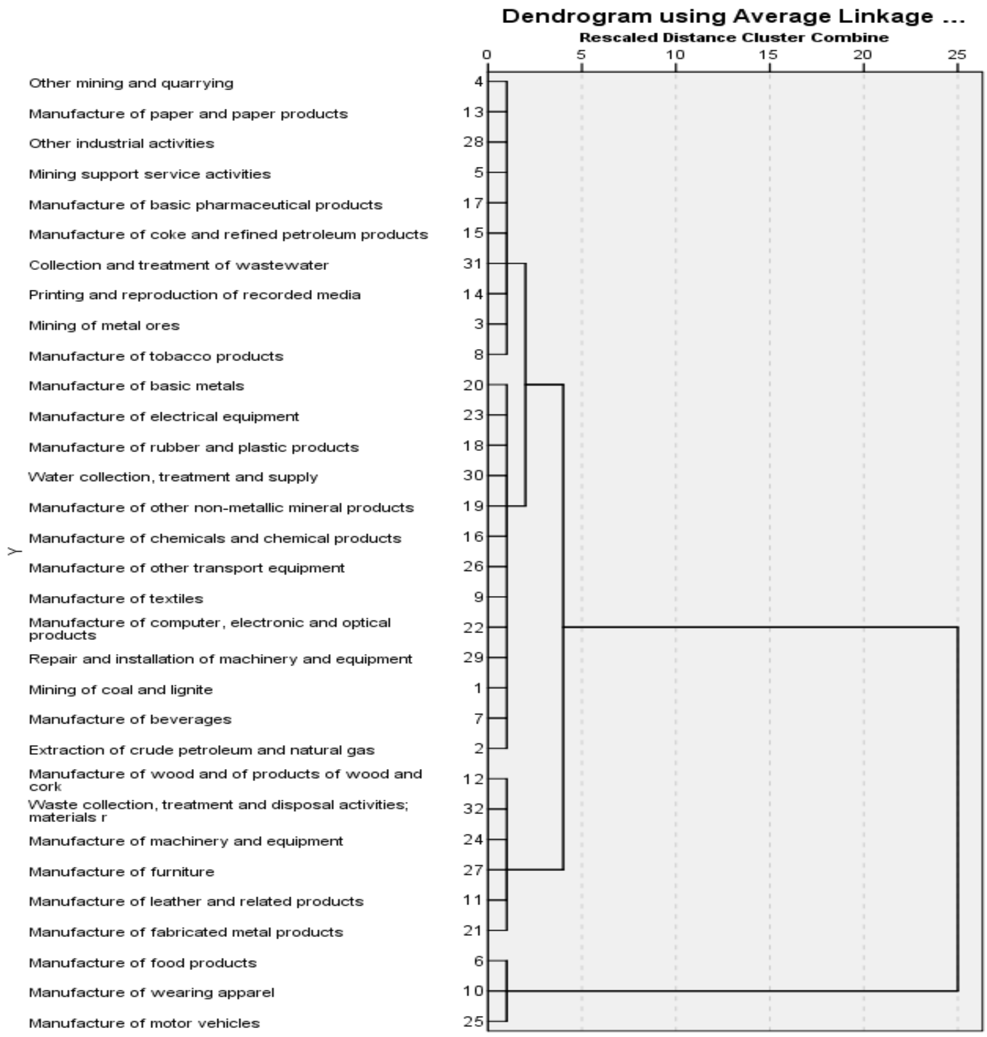
$$employ_{it} = a + b_1 * lwage_{it} + b_2 * linv_{it} + b_3 * llabor\_cost_{it} \quad 1)$$

The econometric analyses of all four industry clusters were built on a similar structure.

We first checked whether a panel data model is more appropriate than a simple pooled OLS model, and the results suggested that panel data models were more suitable (see Table 1). When running the Hausman test in order to decide whether a RE model is more appropriate than a FE model, the probability was less than 5% in all cases. Concluding that we are dealing with fixed effects, we estimated the four models at industry-cluster level using the within estimator.

Figure 1

Economic Activities Dendrogram



Secondly, when applying for each of the four clusters both the modified Wald test for groupwise heteroskedasticity in the FEM and the Wooldridge serial correlation test implemented in the STATA software, it resulted that the errors were both auto correlated and heteroskedastic. That is why, in order to ensure the validity of the statistical results, we had to estimate robust fixed-effects regressions with the Driscoll and Kraay standard errors (see Tables 2-5).

Table 1

Preliminary Test Results

	1st Cluster Equation	2nd Cluster Equation	3rd Cluster Equation	4th Cluster Equation
F test for Pooled OLS versus panel data model	F(9,109)=27.9 Prob.=0.00	F(12,142)=32 Prob.=0.00	F(5,64)=16.1 Prob.=0.00	F(2,32)=56 Prob.=0.00
Hausman test Probability	0.017	0.00	0.026	0.00
Wald modified test	Chi2(10)=18837 Prob.=0.00	Chi2(13)=67.5 Prob.=0.00	Chi2(6)=57.2 Prob.=0.00	Chi2(3)=212.4 Prob.=0.00
Wooldridge serial correlation test	F(1,9)= 97.7 Prob.=0.00	F(1,12)= 709 Prob.=0.00	F(1,5)= 741 Prob.=0.00	F(1,2)= 26.7 Prob.=0.036
Conclusions	Robust fixed-effects estimation required			

Source: Authors' calculations.

The employment equation for the first cluster of the Romanian industry activities is:

$$Lemploy_{it} = 11.19432 - 0.2994^{*}\ lwage_{it} \quad (2)$$

Table 2

Robust Fixed-Effects Regression Model for Cluster 1

Regression with Driscoll- Kraay standard errors				Obs.	120
Method: Fixed-effects regression				Groups	10
Dependent variable:	lemploy			F(1,9)=	32.17
Explanatory variables	**Coef.**	**Driscol/Kraay Std. Err.**	**t**	**[95% Conf. Interval]**	
lwage	-0.2994*	0.0528	-5.67	-0.4188	-0.1799
_cons	11.194*	0.3699	30.27	10.358	12.031
within R-squared	0.304				
where * stands for 1% significance.					

Source: Authors' calculations using STATA software.

The only statistically significant explanatory variable when estimating the employment equation for the first panel data model turned out to be the average gross earnings.

As expected, the average gross earnings influence employment in a negative way, with a coefficient that indicates a decrease by about 0.3% in the number of employed population in case the average gross earnings increase by one percent.

We can notice that our results are consistent with the neoclassic assumption of a negative correlation between employment rates and wage levels.

When estimating the robust panel data model *for the second industry cluster*, the employment equation is:

$$Lemploy_{it} = 11.994 - 0.2047^{*}\ llabor\_cost_{it} \quad (3)$$

In this case, the impact of the average gross earnings was statistically insignificant and replaced by the impact of the labour force cost. The negative coefficient of the labour cost indicates that an increase by 1% in labour cost implies a reduction by almost 0.205% in the employed population.

Table 3

Robust Fixed-Effects Regression Model for Cluster 2

Regression with Driscoll- Kraay standard errors				Obs.	156
Method: Fixed-effects regression				Groups	13
Dependent variable:	lemploy			F(1,9)=	28.48
Explanatory variables	**Coef.**	**Driscol/Kraay Std. Err.**	**t**	**[95% Conf. Interval]**	
llabor_cost	-0.2047*	0.0384	-5.34	-0.288	-0.121
_cons	11.994*	0.269	44.58	11.408	12.58
within R-squared	0.43				
where * stands for 1% significance.					

Source: Authors' calculations using STATA software.

The employment equation for the third cluster of the Romanian industry activities is:

$$Lemploy_{it} = 16.085- 5*llabor\_cost_{it} +4.4845* lwage_{it} \quad (4)$$

Although we still notice the absence of the stimulating effect of net investments upon employment, in this case, both the average gross earnings and the labour cost turned out to be statistically significant. The strongest relation is that between the labour cost and the employed population, while an increase in the level of average gross earnings propels the employed population growth by almost 4.5%, keeping all other variables constant. This time, the sign of the wages is opposite to what we would have expected. The explanation might be based on an offer effect supported by an increase in the labour demand during the expansion period, meaning that an increase in wage might have supported an increase in employment by attracting new echelons of labour force.

Table 4

Robust Fixed-Effects Regression Model for Cluster 3

Regression with Driscoll- Kraay standard errors				Obs.	72
Method: Fixed-effects regression				Groups	6
Dependent variable:	lemploy			F(1,9)=	4.7
Explanatory variables	**Coef.**	**Driscol/Kraay Std. Err.**	**t**	**[95% Conf. Interval]**	
lwage	4.485**	1.644	2.73	0.258	8.711
llabor_cost	-5.001**	1.811	-2.76	-9.656	-0.346
_cons	16.085*	1.655	9.72	11.831	20.339
within R-squared	0.36				
where *, ** stand for 1% , respectively 5% significance.					

Source: Authors' calculations using STATA software.

When estimating the robust panel data model *for the last industry cluster*, the employment equation is:

$$Lemploy_{it} = 12.45 - 0.08^{*} \, lwage_{it} \quad (5)$$

Table 5

Robust Fixed-Effects Regression Model for Cluster 4

Regression with Driscoll- Kraay standard errors				Obs.	36
Method: Fixed-effects regression				Groups	3
Dependent variable:	lemploy			F(1,9)=	7.73
Explanatory variables	**Coef.**	**Driscol/Kraay Std. Err.**	**t**	**[95% Conf. Interval]**	
lwage	-0.08***	0.029	-2.78	-0.204	0.044
_cons	12.448*	0.172	72.18	11.71	13.19
within R-squared	0.092				
where *, *** stand for 1%, respectively 10% significance.					

Source: Authors' calculations using STATA software.

The only statistically significant explanatory variable for the employment equation of the fourth panel data model turned out to be the average gross earnings with a very slight negative impact upon employment.

IV. The Forecasting Scenario

Based on the econometric relations highlighted by the estimated panel data models presented in the previous section, the study continued with a forecasting scenario of the evolution of the explanatory variables in order to predict the total number of employed population on the 2012-2013 horizon in Romania.

For that, we used the panel data models previously estimated for each of the four main Romanian industry clusters and had to forecast the evolution of the two explicative variables of the model, meaning the average gross earnings and the labor cost for the same period.

Since the uniform distribution is one of the simplest and most commonly used distributions in macroeconomic predictions, working very well in most cases, we assumed that the random variation of the annual percentage growth of the two explanatory variables will be best descried by a uniform distribution between the following intervals, as presented in Table 6.

The inferior and the superior limits of each interval of the uniform distributions were established after considering the main features of the four clusters, the annual percentage variations of the two explanatory variables during the 2000-2011 period and our own expectations concerning future industrial activities fluctuations.

Since the interval limits are assumed to be the annual percentage growths of the macroeconomic indicators as compared to the previous year, when predicting for the year 2013 the variations were determined on the basis of the average values that were previously estimated for the year 2012.

For each cluster and explanatory variable, an average annual percentage growth was separately calculated by equalizing it to the theoretical mean of the uniform distributions described in Table 6.

Table 6

Statistical Hypothesis for the Forecasting Scenario

	WAGE (%)	LABOR_COST (%)
Cluster 1	[0%;10%]	
Cluster 2		[5%;15%]
Cluster 3	[3%; 9%]	[3%; 9%]
Cluster 4	[5%;10%]	

Source: Authors' calculations.

After forecasting the average gross earnings and the labor cost for the year 2012 for each cluster individually, we entered the data into the panel data models and forecasted the employment population for the year 2012. The same steps were repeated in order to extend the prediction for the year 2013 and the predicted average levels are presented in Table 7.

Table 7

Predictions for the Period 2012-2013

	EMPLOYMENT (persons)		WAGE (RON)		LABOR COST (RON)	
	2012	2013	2012	2013	2012	2013
Cluster 1	6914	6822	2798	2925	3600	3777
Cluster 2	23209	22778	2413	2656	3513	3855
Cluster 3	54593	52417	1558	1654	2019	2148
Cluster 4	141383	140579	1646	1768	2184	2347

Source: Authors' calculations.

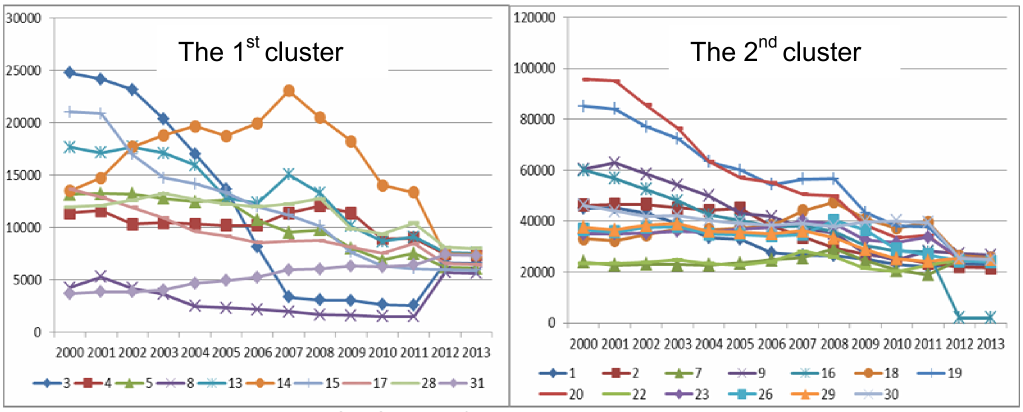
The results suggest that, based on this scenario, the employed population in Romania is more likely to face only slight reductions during the forecasting horizon, in comparison to the previous years (see Figures 2 and 3).

Moreover, the descending trend seems to be more noticeable for the economic activities assigned to the first and second industry clusters as compared to the other two industry clusters.

The only economic activities that tend to keep a positive trend in employment, however, during the forecasting horizon are the following: *Collection and treatment of waste water* assigned to the first industrial cluster and *Manufacture of motor vehicles* assigned to the fourth cluster.

Figure 2

The Forecast of the 1st and 2nd Cluster Employment Evolution



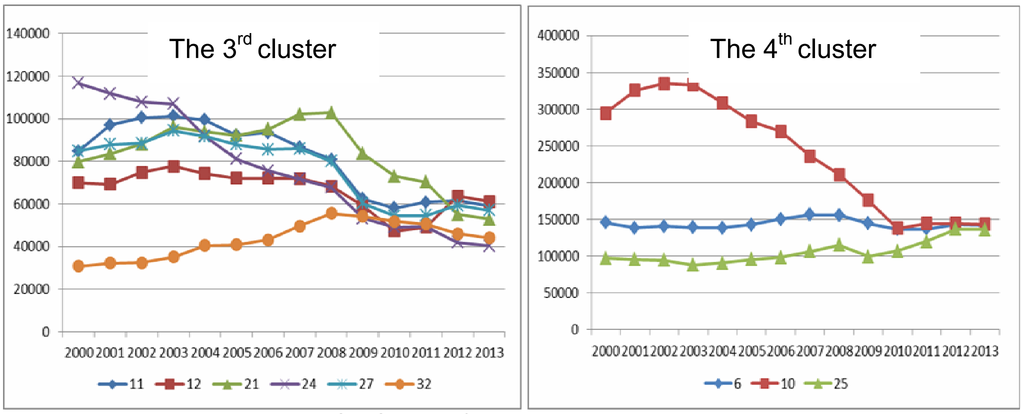
Note: The numbers are the NACE Codes of the sectors included in a cluster.

The industrial activities that had a constant evolution during the sample period were the following: *Manufacture of food products* assigned to the fourth cluster, *Manufacture of beverages* and *Manufacture of computer, electronic and optical products*, both belonging to the second industrial cluster.

The final results of the forecasting scenario indicated that at industry level the employment is expected to decrease by 5.5% in 2012 as compared to 2011 and by only 2% in 2013 as compared to the previous year.

Figure 3

The Forecast of the 3rd and 4th Cluster Employment Evolution



Note: The numbers are the NACE Codes of the sectors included in a cluster.

V. Conclusions

In this paper, we investigated the impact of various macroeconomic variables, such as labour cost, average gross earnings and net investments upon employment in Romania. For that, several econometric models were estimated, based on panel data series for the period 2000-2011 for 32 main activities of the Romanian industry, for which data were available.

First, a cluster analysis was applied in order to classify the economic activities into several different groups. Based on the main four resulted industry clusters, the econometric study continued with four more distinctive panel data models, which allowed for a more complex analysis of the performance of the Romanian industry sectors and provided additional information regarding the Romanian labour market fluctuations.

When estimating the panel data models for the four industry clusters identified with the hierarchical cluster analysis, several results were obtained. Although net investments turned out to be statistically insignificant in all four cases, one can notice that our results were consistent with the neoclassic assumption of a negative correlation between employment rates and wage levels and labour costs.

More precisely, for the first and the fourth cluster, the only significant explanatory variable when estimating the employment equation turned out to be the average gross earnings, with a negative impact upon employment. However, in the case of the second cluster the employment variation is explained only by the labour force cost evolution, while the employment equation of the third cluster includes both the impact of the wages and labour force cost. This time, the sign of the wages is opposite to what we would have expected. The explanation might be based on an offer effect supported by an increase in the labour demand during the expansion period, meaning that an increase in wages might have supported an increase in employment by attracting new echelons of labour force.

Based on the estimated panel data models, a forecasting scenario was then built in order to predict the total number of employed population on the 2012-2013 horizon in the Romanian industry.

The novelty of this paper in the international literature consists, in fact, in the proposed forecasting procedure that was based on several panel data models and hierarchical cluster analysis. Its main advantage is the fact that it has a large potential of applicability, since it can easily be similarly applied to other macroeconomic models as well.

Acknowledgment

This paper was co-financed from the European Social Fund, through the Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields -EXCELIS", coordinator The Bucharest University of Economic Studies.

References

Albulescu, C.T., 2010. Forecasting the Romanian financial system stability using a stochastic simulation model", *Romanian Journal of Economic Forecasting*, 1(1), pp. 81-98.

Andreica, C. *et al*., 2007. Conclusions on Using the Statistical Methods in Forecasting the Structure Evolution of an Economic Indicator System. *Economic Computation and Economic Cybernetics Studies and Research*, 41(1-2), pp. 147-157

Andreica, M., 2006. A Model to Forecast the Evolution of the Structure of a System of Economic Indicators. *Romanian Journal of Economic Forecasting*, 3(1), pp. 65-73.

Andreica, M.E. *et al*. 2010. A Panel Data Analysis of the Romanian Labour Market. *Proceedings of the 5th WSEAS International Conference EMT '10*, Timişoara, pp. 406-411.

Andreica, M.E. *et al*., 2010. Models of the Minimum Wage Impact upon Employment, Wages and Prices: The Romanian Case. *Proceedings of the 11th WSEAS International Conference MCBE '10*, Iasi, pp.104-109.

Antonie, M.D., Cristescu, A., Cataniciu, N., 2010. A Panel Data Analysis of the Connection between Employee Remuneration, Productivity and Minimum Wage in Romania. *Proceedings of the 11th WSEAS Int. Conf. MCBE*, pp.134-139.

Baltagi, B.H., 2008. *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd.

Baum, C.F., 2001. Residual diagnostics for cross-section time series regression models. *The Stata Journal*, 1(1), pp.101–104.

Cameron, A.C. Trivedi, P.K., 2009. *Microeconometrics Using Stata*. Stata Press Publishing.

Dobrescu, E., 2010. Macromodel simulations for the Romanian economy. *Romanian Journal of Economic Forecasting*, 2010(2), pp. 7-28.

Drukker, D.M., 2003. Testing for serial correlation in linear panel-data models. *The Stata Journal*, 3(2), pp.168–177.

Hoechle, D., 2007. Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence. *The Stata Journal*, 7(3), pp. 281-312.

Jula, D. Jula, N.M., 2013. Economic Growth and Structural Changes in Regional Employment. *Romanian Journal of Economic Forecasting*, 16(2), pp. 52-69.

Nicolae, D. *et al*., 2010. The progress offered by the subtle sets theory (SST) in the investigation of the socio-economic processes. *Proceedings of the 9th WSEAS Int. Conf. AIKED '10*, University of Cambridge, UK, pp. 251-256.

Seyfried, W., 2005. Examining the Relationship between Employment and Economic Growth in the Ten Largest States. *Southwestern Economic Review*, 35, pp. 13-24.

Schaafsma, J. Walsh, W.D., 1983. Employment and Labour Supply Effects of the Minimum Wage. *The Canadian Journal of Economics*, 16(1), pp. 86-97.

Wooldridge, J.M., 2003. *Introductory econometrics – A modern approach*. Second Edition, South-Western Publishing Co, pp. 501-528.