Some Conceptual Aspects of the Multilevel Modeling for the Study of Social-Economic Phenomena

***Miruna MAZURENCU MARINESCU*[1]**

Abstract: *Multilevel modeling has been mainly used in education research, health studies, psychology and sociology, its application to the study of social-economic phenomena and especially to business studies is rather scarce. Hence, its application to business studies might produce interesting new insights on business performance, especially due to the micro-macro interactions, for all the stakeholders. Therefore, the research is focused on business, namely to the analysis of the Romanian ICT sector. The current paper aims at illustrating a new assessment tool for business analysis, namely the assessment of company performance through a new approach called multilevel modeling. Given the nested structured of the database comprising the Romanian companies operating in ICT industry, a multilevel random coefficient model is suggested and how it can be fitted in R software. The conclusions of the research can be used by various stakeholders, including policy makers.*

Keywords**:** *multilevel modeling; socio-economic phenomena; performance; business analysis*

JEL**:** *CO1; C52; C87.*

Introduction

The current paper introduces a more advanced methodological approach for assessing corporate performance, which, if implemented can benefit various stakeholders from entrepreneurs to bankers and policy makers. The need for such an approach emanates from both academe and practitioners to identify a superior methodological approach in general and very industry specific in particular.

The proposed methodology is a multi-level modeling approach that to the best of our knowledge has not been used in business yet and even more to an industry, with a very rapid pace of change like ICT. The multilevel approach is a fruitful methodological framework in which to formulate the micro-macro relationships existing between entities and their context. The subsector in which a particular ICT based business operates can be taken as proxy for context.

The underlying idea is that interactions between entities (companies) and their context (the subsector they operate within) influence individual behaviour and, in turn, shape up subgroup characteristics and properties.

[1] Professor Ph.D., MBA, Fulbright scholar, Bucharest University of Economic Studies, Economic Cybernetics, Statistics and Informatics Faculty, Bucharest, Romania, e-mail: miruna@ase.ro

As ICT industry has been the driving force of the Romanian economy lately, it is important to understand how sustainable its companies are, now, once the EU accession requirements for e-governance are no longer fuelling the same the business opportunities and compounding further, the entire economy has suffered from the current financial crisis.

Given the current background, and the fact that ICT industry is highly heterogeneous, the following questions arise:

1. How much do Romanian ICT based companies vary in their performance (defined as profit/net result/rate of return);
2. Given the fact that ICT industry is human capital intensive, does a larger number of employees predict a better performance?
3. How and to what extent turnover/equity/assets/long term debt/net working capital explain the variation between companies within the same subsector and between subsectors with the same average rate of return?
4. How the relationship between performance and turnover, equity, assets, long term debt, net working capital changes over time across companies and subsectors?
5. Is the connection between company size (defined by the number of employees) and performance similar across companies? Or does the relationship show substantial variation? Same can be applied for all the other predictors.
6. How do subsectors compare in terms of performance and the strength of association between size and performance after we control for the average size?

The purpose of this study is how to address these substantive research questions with multilevel modeling and further more, to briefly illustrate the proposed models in R software. Multilevel analysis or modeling is a term used to describe a set of analyses also referred to as multilevel random coefficient models or mixed-effects models (Bryk & Raudenbush, 1992; Kreft & De Leeuw, 1998; Snijders & Bosker, 1999).

Usually, the definition of multilevel modeling reflects a wide range of interrelated multilevel topics (see also Klein & Kozlowski, 2000), like within-group agreement and reliability, contextual OLS models, covariance theorem decomposition, Random Coefficient Modeling and Random Group Resampling. For the purpose of this current paper, we will restrict to the application of multilevel random coefficients models by using a data set gathered from the Romanian National Institute of Statistics.

1. The data

The data set comprises approximately 68000 validated records, from 2002 to 2010. The resulting collection of companies included cross-sectional records for the following variables: Company fictional ID, Activity domain (subsector), Net

Turnover, Total Owner's Equity, Long term debt, Number of employees, Net result and Net working capital (figure 1). The evolution of the ICT companies' performance along the years can be depicted in the following graph:

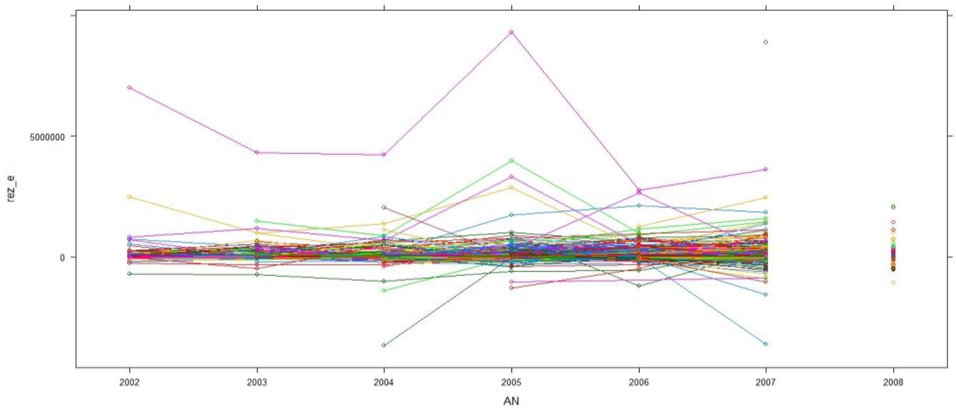

Figure 1. Net result of the Romanian ICT companies between 2002 to 2008

As it can be gleaned from the above graph, one can notice the variability of the companies' performance, not only in time, but also from one to the other. If the analysis is deepened even further, the variability is even more pronounced if we cluster the companies according to their activity code. (Figure 2)

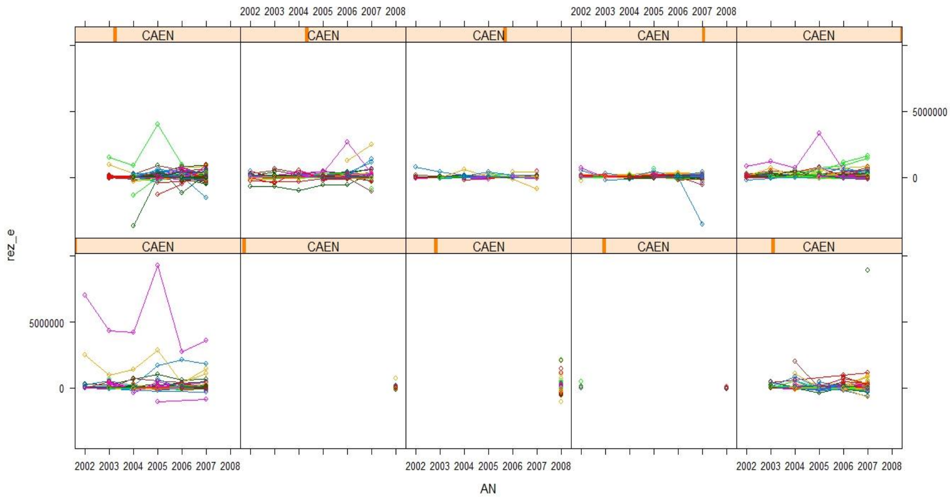

Figure 2. The evolution of the net result of the Romanian ICT companies between 2002 to 2008, according to activity code

All these database manipulations indicated as the most appropriate tool for assessing the business performance the multi-level methodology.

2. The methodology and the model

The key to applying multilevel random coefficient models (MRC) is understanding how group membership can lead to additional sources of variance in one's model. The first variance term that distinguishes a MRC model from a regression model is a term that allows groups to differ in their mean values (intercepts) on the dependent variable. When this variance term, τ_{00}, is non-zero, it suggests that groups differ on the dependent variable. When groups differ by more than chance levels one can potentially model why some groups have high average dependent variable values while other groups have low average dependent variable values. One predicts group-mean differences with group-level variables. These are variables that differ across groups, but do not differ within-groups. Group-level variables are often called "level-2" variables.

The second variance term that distinguishes a MRC model from a typical regression model is the term that allows slopes between independent and dependent variables to differ across groups (τ_{11}). Simple regression models generally assume that the relationship between the independent and dependent variable is constant across groups. In distinction, MRC models allow the slope to vary from one group to another. If slopes randomly vary, one can attempt to explain this slope variation as a function of group differences – again, one uses level-2 variables to explain why the slopes within some groups are stronger than the slopes within other groups.

A third variance term is common to both MRC and regression models. This variance term, σ^2, reflects the degree to which an individual score differs from its predicted value within a specific group. One can think of σ^2 as an estimate of within-group variance. One uses individual-level or level-1 variables to predict within-group variance, σ^2. Level-1 variables differ among members of the same group.

Typically, in a complete MRC analysis, one wants to know (1) what level-1 factors are related to the within-group variance σ^2?; (2) what group-level factors are related to the between group variation in intercepts τ_{00}?; and (3) what group-level factors are related to within-group slope differences, τ_{11}? More complex models, as 3 level ones can also be designed and fitted, depending on how many layers of data are nested within each other.

As we are not working with a SAMPLE of subsectors (Activity Codes) we cannot fit a three level model, with measures nested within companies, and companies nested within subsectors. As we have every subsector in our data set, we would treat subsector as a Level 2 covariate (a feature of the company), and include fixed effects of the subsectors in the models.

The outcome variable $Y_{ij}$is the NET RESULT. There are six potential level-1 predictors: turnover (T/O), total assets (TA), net working capital (NWC), total owner's equity (TE), long term debt (LTD) and the number of employees (E) of individual companies. At level 2, there are two potential (subsector level) predictors: SSECTOR and MEANROR, which is the average rate of return for the subsector.

In this section, the theoretical models are presented, followed by their testing using R software.

Model A.

Level-1: NET RESULT_$e_{ij}=\pi_{0i}+\pi_{1i}$*$TIME_{ij}+\varepsilon_{ij}$, $\varepsilon_{ij}\sim N(0,\sigma_\varepsilon^2)$

Level-2: $\begin{cases}\pi_{0i}=\delta_{00}+\zeta_{0i}\\ \pi_{1i}=\delta_{10}+\zeta_{1i}\end{cases}$

$$\begin{matrix}\delta_{00}\\ \delta_{10}\end{matrix}\sim N\begin{pmatrix}0 & \sigma_0^2\ \sigma_{01}\\ 0 & \sigma_{10}\ \sigma_1^2\end{pmatrix}$$

By substituting level 2 variables into the level 1 model, one obtains the following form with one equation:

Model A composed: NET RESULT_$e_{ij=}$ $\delta_{00}+\delta_{10}$* $TIME_{ij}$+[$\zeta_{0i}+\zeta_{1i}$* $TIME_{ij+}$ ε_{ij}]

Model B.

Level-1: NET RESULT_$e_{ij}=\pi_{0i}+\pi_{1i}$*$TIME_{ij}+\pi_{2i}$*$E_{ij}+\varepsilon_{ij}$, $\varepsilon_{ij}\sim N(0,\sigma_\varepsilon^2)$

Level-2: $\begin{cases}\pi_{0i}=\delta_{00}+\zeta_{0i}\\ \pi_{1i}=\delta_{10}+\zeta_{1i}\\ \pi_{2i}=\delta_{20}\end{cases}$

whereas E stands for the number of employees.

By substituting level 2 variables into the level 1 model, one obtains the following form with one equation:

Model B composed: NET RESULT_$e_{ij=}$ $\delta_{00}+\delta_{10}$* $TIME_{ij}+\delta_{20}$ *E_{ij}+[$\zeta_{0i}+\zeta_{1i}$* $TIME_{ij+}$ ε_{ij}]

Model C.

Level-1: rez_$e_{ij}=\pi_{0i}+\pi_{1i}$*$TIME_{ij}+\pi_{2i}$*$E_{ij}+\varepsilon_{ij}$, $\varepsilon_{ij}\sim N(0,\sigma_\varepsilon^2)$

Level-2: $\begin{cases}\pi_{0i}=\delta_{00}+\zeta_{0i}\\ \pi_{1i}=\delta_{10}+\zeta_{1i}\\ \pi_{2i}=\delta_{20}+\zeta_{2i}\end{cases}$

By substituting level 2 variables into the level 1 model, one obtains the following form with one equation:

Model C composed: NET RESULT_$e_{ij=}$ δ_{00}+ δ_{10}* $TIME_{ij}$+ δ_{20} *E_{ij}+[ζ_{0i}+ ζ_{1i}* $TIME_{ij}$+ ζ_{2i}* E_{ij}+ ε_{ij}]

Model D.

Level-1: NET RESULT_e_{ij}=π_{0i}+ π_{1i}*$TIME_{ij}$+ π_{2i}*E_{ij}+π_{3i}*T/O_e_{ij}+ π_{4i}*TE_e_{ij}+ π_{5i}*DTL_e_{ij}+ π_{6i}*ACDCN_e_{ij}+ε_{ij}, ε_{ij}~N(0,σ_ε^2)

Level-2:
$$\begin{cases} \pi_{0i}=\delta_{00}+\zeta_{0i} \\ \pi_{1i}=\delta_{10}+\zeta_{1i} \\ \pi_{2i}= \delta_{20}+ \zeta_{2i} \\ \pi_{3i}= \delta_{30}+ \zeta_{3i} \\ \pi_{4i}= \delta_{40}+ \zeta_{4i} \\ \pi_{5i}= \delta_{50}+ \zeta_{5i} \\ \pi_{6i}= \delta_{60}+ \zeta_{6i} \end{cases}$$

whereasT/O_e stands for turnover in euro;
TE_e stands for total equity in euro;
LTD_e stands for long term debt in euro;
NWC_e stands for net working capital in euro;

By substituting level 2 variables into the level 1 model, one obtains the following form with one equation:

Model D composed: NET RESULT_$e_{ij=}$ δ_{00}+ δ_{10}*$TIME_{ij}$+ δ_{20}*E_{ij}+ δ_{30}*T/O_e_{ij}+ δ_{40}*TE_e_{ij} + δ_{50}*LTD_e_{ij} + δ_{60}*NWC_e_{ij}+ [ζ_{0i}+ ζ_{1i}* $TIME_{ij}$+ ζ_{2i}* E_{ij}+ ζ_{3i} *T/O_e_{ij}+ ζ_{4i} *TE_e_{ij} + ζ_{5i} *LTD_e_{ij} + ζ_{6i} *NWC_e_{ij+} ε_{ij}]

In the following section we shall illustrate how can the models be fitted by using R software.

3. The R software

Due to the wide variety of topics covered in the definition of multilevel modeling, it is necessary to use several “packages” written for R. The first of these packages is the “base” package that comes with R. This package is automatically loaded and provides the basic structure of R along with routines to estimate ANOVA and regression models important in contextual OLS models. (Bliese, P. 2009)

The analysis begins by attaching the multilevel package, which also loads nlme package and making the dataset in the multilevel package available for the analysis.

```
➢ library(multilevel)
➢ library(nlme)
➢ companyinfo<-read.csv("d:\\infostatnew.xlsx",
  "header=T")
➢ companyinfo
```

In `lme` we specify the cross-level interaction by adding an interaction term. Specifically, the model is:

```
➢ Model.3<-
  lme(NETRESULT~TO+TA+NWC+TE+LTD+E+SSECTOR+MEANROR
  , random=~SSECTOR|MEANROR, DATA=companyinfo,
control=list(opt="optim"))
➢ Summary(Model.3)
```

As the interpretation of the results and the discussion of other technical pitfalls fall beyond the scope of the present paper, we shall limit to the illustration of the technique.

4. Conclusions

This proposed business tool can not only serve as a danger sign prediction tool, but also as an intelligent and complex method for monitoring the total health of a company, its business, or a division thereof. In the day to day running of a business, managers and senior executives are normally tasked with monitoring business activity to ascertain adherence to strategic objective set, presented in various measurable performance criteria that are only relevant to a particular function or a sub-division of business line.

Data is normally collected and aggregated in a bottom-up approach then condensed into a set of "management accounts" to be presented to board members, senior managers, in conformity with, and normally "aligned" to accounting policies and standard adopted by the company.

Most accounting policies and standards internationally practiced and approved are mostly concerned with and almost too narrowly focused on: cash-flows. Whilst cash-flows is very critical to survival of any business, other various business ratios derived or related to cash-flow in most businesses monitoring and evaluation tools, procedures and accounting practice should be used.

Therefore, the proposed new approach still relies on business ratios but they are used to gauge the subtle differences that differentiate companies between themselves, in time and across various subsectors.

5. Limitations of the current research

This paper was meant to merely lay the grounds of a much more complex study on applying the multilevel modeling into business research, with a particular focus on subsectors variation of the Romanian ICT industry. Besides the already introduced models, there are much more complex ones that still need validation through using empirical data.

6. Further research

The econometric analysis can be taken to a further refinement, that anticipates that the individual net result is partly dependent on measurement occasions (time points) and the subsectors to which they belong. This nested (hierarchical) structure in individual net result can be modelled by separating the time, company and subsector sources of variation. Antweiler (2001) compares the nested maximum likelihood (ML) estimator results with conventional non-nested ML estimator and corresponding Monte Carlo simulation and reports a downward bias in the standard errors of the regresses when the conventional (non-nested) random effect estimator is applied to a nested/hierarchical panel.
One might consider the following multilevel/nested unbalanced panel data linear regression model:

$$y_{ijt} = X'_{ijt}\,\beta + SC'_i\,\delta + t'\theta + u_{ijt}$$

Where i=1, …., M, j=1…., N_i and t=1,………., T_{ij}. The dependent variable y_{ijt} denotes the net result of the company *j* operating in the subsector *i* in time period t.

In this specification, the overall error term u_{ijt} is decomposed into $\mu_i + \nu_j + e_{ijt}$, where $\mu_i$is the error random term for the *i*th subsector, ν_j denotes the nested effect of the *j*th company in the *i*th subsector, and e_{ijt} is the remaining disturbance term (error term for the *t*th observation time of of the *j*th company in the *i*th subsector). Each error term should be assumed as independently and identically distributed (IID) with mean zero and their respective variances. These disturbance terms assumed to be independent of each other.

This model allows for an unequal number of companies in each subsector as well as different numbers of observed time periods across companies (Baltagi et al., 2001; Snijders and Bosker, 1999). In the above mentioned equation, y_{ijt} represents the net result (the dependent variable) which is related to a vector of individual-level explanatory variables X and the subsector-level variable SC, which varies between subsectors, but is fixed over time for each subsector and fixed among companies operating in the same subsectors. Y indicated year dummies.

Since the dependent variable is continuous, a multilevel linear model can be used. As the ML estimator provides a convenient estimation method (Antweiler, 2001), the model can be fitted using the Restrictive Iterative Generalized Least Squares (RIGLS) ML estimator and implemented within R software.

Among other advantages, multilevel regression analysis allows to identify and quantify the extent to which differences in the net results are attributable to time and the subsector in which the company has operated. This technique enables to determine if subsector level variable affects individual performance (net result) over and above individual characteristics. Multilevel regression analysis also allows us to investigate how much of the subsector differences in performance are explained by differences in the individual composition of subsectors in a sequence of multilevel models.

The ability to partition variance at different levels (e.g. subsectors, companies and measurements) is a unique feature of multilevel regression analysis, and its consideration is relevant for better estimation (unbiased) and quantification of the relative importance of individual compositional and contextual effects for understanding individual performance variations.

To illustrate the relevance of the subsector differences (variance) for understanding individual performance differences, there are two measures: (1) the intra-class correlation, and (2) the proportional change in variance or total variance explained. The intra-class (cluster) correlation (ICC) or the variance partition coefficient (VPC) expresses the proportion of the individual differences in the performance (i.e. individual variance) that is at the subsector level (contextual). The closer the ICC/VPC is to 0%, the smaller the proportion of the total variance at the subsector level and the lower the relevance of the areas for understanding individual disparities in performance. The ICC is calculated by the general formula:

$$ICC = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\vartheta^2}$$

The ICC expresses the proportion of the individual-level variance $(\sigma_\mu^2 + \sigma_\vartheta^2)$ that is at the subsector level σ_μ^2.

Several models can be fitted. Using a Student's t-statistic (Wald test), one can test statistical significance for each regression coefficient and for each variance between subsectors, between companies and between time.

Acknowledgements

This work was supported by the project "Post-Doctoral Studies in Economics: training program for elite researchers - SPODE" co-funded from the European Social Fund through the Development of Human Resources Operational Programme 2007-2013, contract no. POSDRU/89/1.5/S/61755.).

References

1. **ANTEWEILER, W.** (2001), *Nested random effects estimation in unbalanced panel data*, Journal of Econometrics, 101:295-313;
2. **BALTAGI, B.H., SONG, S.H. & JUNG, B.C.** (2001), *The unbalanced nested error component regression model*, Journal of Econometrics, 101: 357-381;
3. **BALTAGI, B.H.** (2001), *Econometric Analysis of Panel Data*, John Wiley and Sons;
4. **BLIESE, P.** (2009), *Multilevel Modeling in R(2.3)*
5. **BRYK, A. S., & RAUDENBUSH, S. W.** (1992).*Hierarchical linear models*. Newbury Park, CA: Sage
6. **ISLAM, K.M.,** and al. (2006), *Does it really matter where you live? A panel data multilevel analysis of Swedish municipality-level social capital on individual health*-related quality of life, health economics, policy and law, 1:209-235
7. **KLEIN, K. J. & KOZLOWSKI, S. W. J.** (2000).*Multilevel theory, research, and methods in organizations*. San Francisco, CA: Jossey-Bass, Inc
8. **KREFT, I. & DE LEEUW, J.** (1998). *Introducing multilevel modeling*. London: Sage Publications
9. **MAZURENCU MARINESCU, M., PELE D.T.** (2012). Applying multilevel modeling for assessing business performance in R software: case study--the companies operating in Romania's ICT sector, *Journal of International Business and Economics*, ISSN: 1544-8037, Volume 12, Number 3, pp. 99-102
10. **SNIJDERS, T. A. B. & BOSKER, R. J.** (1999). *Multilevel analysis: An introduction to basic andadvanced multilevel modeling*. London: Sage Publications