Professor Miruna MAZURENCU-MARINESCU, PhD
The Bucharest Academy of Economic Studies
E-mail: miruna@ase.ro
Professor Peter NIJKAMP, PhD
VU University Amsterdam
E-mail: p.nijkamp@vu.nl
Lecturer Daniel Traian PELE, PhD
The Bucharest Academy of Economic Studies
E-mail: danpele@ase.ro

AN ECONOMETRIC ANALYSIS OF ICT – ORIENTED COMPANIES' BANKRUPTCY. A CASE STUDY ON ROMANIA

Abstract. *Bankruptcy and insolvency risk analysis with all their respective methodologies and techniques are well covered topics in the literature. Investigating companies operating in very dynamic or specific industries from an insolvency risk point of view using traditional tool packs may become rather cumbersome for various reasons. The purpose of this paper is to suggest a much simplified and relatively straightforward method using SAS System software, as well as discussing some potential pitfalls.*
Keywords*: unbalanced panel data, insolvency, IT-oriented businesses, bankruptcy.*

JEL classification: L62, L86, M15

1. Introduction

Over the past decade, the ICT sector in Romania has witnessed an unprecedented increase, not only in terms of business performance, but also in controversies surrounding its viability and sustainability after Romania's EU accession in 2007. There have been various studies which did look deeper into the Romanian ICT sector, ranging from how to appraise an IT based business to bankruptcy risk for this type of companies.

Unfortunately, there were several main limitations to this research related to the scarcity of reliable data on various accounting indicators, both cross-sectional and time series, not to mention companies reporting "inadvertencies" (such as conflicting definitions of the declared main type of activity). Some of these difficulties will be addressed in the present research, which offers a new model for assessing the performance (i. e., bankruptcy) of ICT-based firms in Romania. We will first start with a concise review of the literature.

2. Literature review

The ICT-sector has employed a great deal of scientific interest. To the best of our knowledge, there is virtually no study on the survival of ICT based businesses operating in Romania.

The present study is an attempt to design an econometric model for assessing the bankruptcy risk of this type of businesses, while taking into account a multitude of constraints both data and methodology related.

The main international literature stream is dedicated to the analysis of duration of a life expectancy of firms in the ICT sector, the main focus is on firms that are publicly owned.

Private firms do not release their financial information, and therefore cannot be handled by a more complex research design like Bayesian dynamic models with time-varying coefficients, which, in turn, surpass the explanatory power of the Cox proportional hazards model by examining how the impact of the explanatory variables changes over time, based on firm age and calendar time. This calls for an alternative approach.

The prediction of the bankruptcy of an e-business can be approached differently, as follows:

- from an accounting point of view, by using financial rates and cash-flow components of firms;
- from an economic point of view, by using duration models to examine the impact of various industry and company related factors on firm survival;
- from an information system point of view, by applying survival analysis and Bayesian statistics to the quantitative analysis of duration.

In the early stages of an e-business, its survival is closely related to initial set-up conditions, while when it grows its survival is increasingly associated with their financial capital and size.

In exploring the business models of this type of companies, it comes out that Internet firms which sell digital products and services are better positioned and equipped for long-term survival than those which sell physical goods through e-reformation (e.g. e-tailers, e-booksellers etc.)

Financial and accounting research has used quite a number of financial ratios to predict bankruptcy, but, among all used indicators, debt ratio (i.e. the ratio of firm's total liabilities to its assets) has consistently been found to offer a significant predictor of firm survival among mature firms.

Economists have proposed the product life cycle (PLC) theory to explain the growth patterns of many industries concluding that as the rates of entry and exit reach equilibrium, the number of firms in the industry stabilizes.

They have also tested the impact of various industry and company related factors like the rate of new firm entry and the technological regime prevailing in the industry. Company related factors include firm size, financial capital and founding time. In addition, a higher initial return on investment and a higher rate of ownership by its insiders can increase the likelihood of survival.

Some studies have also shown that the use of skilled labour is a good predictor of firm survival.

We will now proceed by describing the data base and next the model for a concrete applied research in this field.

3. Data base

The data on Romanian ICT-firms was organized in a panel framework where each firm represented a cross-sectional element, while its data over time comprised the time series counterpart. A balanced panel is hard to acquire in company data. When considering the evolution of a company over the years, some may have been there for all the studied period, while others might have appeared and disappeared without any traceable pattern. There may be sometimes an opportunity to group specific firms together, but, never is not a guarantee for a balanced time series.

As suggested in the literature, a Bayesian approach using dynamic models with time varying coefficients might have been appropriate to model ICT based businesses survival as opposed to their bankruptcy, but, this would have required prior estimation involving OLS with corrections for autocorrelation and heteroskedasticity given a balanced panel, which generally is impossible to obtain.

Therefore, the data set was obtained from the National Institute of Statistics database through interrogation of all companies which have declared information technology as main business activity. The time horizon used in the statistical research is 2002 – 2008. The initial set had more than 68,000 records. After examining the data, its array has been rearranged into a panel framework.

The database, provided by the Romanian National Institute of Statistics, contains 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 (profit/Loss), and Net Working Capital.

4. Methodology

Panel or longitudinal data are cross-sectional and time-series. There are multiple entities, each of which has repeated measurements at different time periods. Panel data may have group effects, time effects, or both, which are analysed by fixed effect and random effect models.

A panel data set contains *n* entities or subjects (e.g. firms), each of which includes T observations measured at time period 1 through *t* (year 2002 to year 2008). Thus, ideally, the total number of observations is *nT*. Panel data should be measured at regular time intervals, otherwise, they should be analysed with caution. Panel data can be balanced or unbalanced. The most fortunate situation for a researcher is to deal with a balanced panel, in which all entities have measurements in all time periods. In a contingency table of cross-sectional and time-series variables, each cell should have only one frequency. When each entity in a data set has different numbers of observations due to missing values, the panel data are not balanced, thus the total number of observations is no longer *nT*.

Unbalanced panel data entail some considerable computational and estimation issues. Panel data models examine fixed and/or random effects of entity (individual or subject) or time. The main difference between fixed and random effect models lies in the role of dummy variables. If dummies are considered as a part of the intercept, this is the case of a fixed effect model. In a random effect model, the dummies act as an error term. A fixed group effect model examines group differences in intercepts, assuming the same slopes and constant variance across entities or subjects. Fixed effect models use least squares dummy variables (LSDV) and within-effect estimation methods. A random effect model, by contrast, estimates variance components for groups (or times) and errors, assuming the same intercept and slopes. The difference among groups (or time periods) lies in their variance of the error term, not in their intercepts. A random effect model is estimated by generalized least squares (GLS) when the Ω matrix, a variance structure among groups, is known. The feasible generalized least squares (FGLS) method is used to estimate the variance structure when Ω is not known.

Fixed effects are tested by the incremental F test, while random effects are examined by the Breusch – Pagan Lagrange Multiplier (LM) test. The Hausman specification test helps in deciding what type of model to use by testing whether the null hypothesis, that is, if the individual effects are uncorrelated with the other regressors in the model is not rejected, so that then a random effect model is better than its fixed counterpart.

If one cross-sectional or time-series variable is considered (e.g. a firm), this is called a one-way fixed or random effect model. Two-way effect models have two sets of dummy variables for group and/or time variables (e.g. firm and year).

SAS provides procedures and commands that estimate panel data models in a convenient way. SAS has the TSCSREG and PANEL procedures to estimate one-way and two-way fixed/random effect models. These procedures estimate the within-effect model for a fixed effect model and by default employ the Fuller-Battese method to estimate variance components for group, time and error for a random effect model. PROC TSCSREG and PROC PANEL also support other estimation methods such as Parks autoregressive model and Da Silva moving average method, although they cannot be used with an unbalanced panel. PROC TSCSREG can handle balanced data only, whereas PROC PANEL is able to deal with balanced and unbalanced panel data. PROC PANEL requires each firm to have more than one observation. The present research uses SAS 9.13, in which PROC PANEL is a mere experimental procedure, and therefore it had to be downloaded and installed from http://www.sas.com/apps/demosdownloads/setupintro.jsp.

The Hausman test results indicated as more appropriate a two-way random-effects model for the particular case of the present research.

The specification for the two-way random-effects model is:

$$u_{it} = v_i + e_t + \varepsilon_{it}$$

As in the one-way random-effects model, the PANEL procedure provides four options for variance component estimators. Unlike the one-way random-effects model, unbalanced panels present some special concerns.

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables arranged by time and by cross section within each time period. Let M_t be the number of cross sections observed in time t and $\sum_t M_t = M$. Let $\mathbf{D}_t$ be the $M_t \times N$ matrix obtained from the $N \times N$ identity matrix from which rows that correspond to cross sections not observed at time t have been omitted. Consider Z=(Z_1, Z_2)
where
$\mathbf{Z}_1 = (\mathbf{D}_1', \mathbf{D}_2', \ldots..\mathbf{D}_T')'$ and $\mathbf{Z}_2 = \text{diag}(\mathbf{D}_1\mathbf{j}_N, \mathbf{D}_2\mathbf{j}_N, \ldots\ldots \mathbf{D}_T\mathbf{j}_N)$.

The matrix $\mathbf{Z}$ gives the dummy variable structure for the two-way model.

For notational ease, let:

$$\Delta_N = \mathbf{Z}_1'\mathbf{Z}_1, \Delta_T = \mathbf{Z}_2'\mathbf{Z}_2, \mathbf{A} = \mathbf{Z}_2'\mathbf{Z}_1$$
$$\bar{\mathbf{Z}} = \mathbf{Z}_2 - \mathbf{Z}_1\Delta_N^{-1}\mathbf{A}'$$
$$\bar{\Delta}_1 = \mathbf{I}_M - \mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'$$
$$\bar{\Delta}_2 = \mathbf{I}_M - \mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'$$
$$\mathbf{Q} = \Delta_T - \mathbf{A}\Delta_N^{-1}\mathbf{A}'$$
$$\mathbf{P} = (\mathbf{I}_M - \mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1') - \bar{\mathbf{Z}}\mathbf{Q}^{-1}\bar{\mathbf{Z}}'$$

The Wansbeek and Kapteyn (1989) method for estimating variance components can be obtained by setting VCOMP = WK. The following methodology, outlined in Wansbeek and Kapteyn is used to handle both balanced and unbalanced data. The Wansbeek and Kapteyn method is the default for a RANTWO model with an unbalanced panel. If RANTWO is requested without specifying the VCOMP= option, PROC PANEL proceeds under the Wansbeek and Kapteyn method if the panel is unbalanced.

The estimator of the error variance is

$$\hat{\sigma}_\epsilon^2 = \tilde{\mathbf{u}}'\mathbf{P}\tilde{\mathbf{u}}/(M - T - N + 1 - (K-1))$$

where the vector $\tilde{\mathbf{u}}$ is given by
$\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{j}_M\mathbf{j}'_M/M)(\mathbf{y}_* - \mathbf{X}_{*s}(\mathbf{X}'_{*s}\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}{}'\mathbf{P}\mathbf{y}_*)$ if there is an intercept and by $\tilde{\mathbf{u}} = (\mathbf{y}_* - \mathbf{X}_{*s}(\mathbf{X}'_{*s}\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}'_{*s}\mathbf{P}\mathbf{y}_*$ if there is not.

The estimation of the variance components is performed by using a quadratic unbiased estimation (QUE) method that involves computing on quadratic forms of the residuals $\tilde{\mathbf{u}}$, equating their expected values to the realized quadratic forms, and solving for the variance components.

Let

$$q_N = \tilde{\mathbf{u}}'\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'\tilde{\mathbf{u}}$$
$$q_T = \tilde{\mathbf{u}}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\tilde{\mathbf{u}}$$

The expected values are:

$$E(q_N) = (T + k_N - (1 + k_0))\sigma^2 + (T - \frac{\lambda_1}{M})\sigma_v^2 + (M - \frac{\lambda_2}{M})\sigma_e^2$$

$$E(q_T) = (N + k_T - (1 + k_0))\sigma^2 + (M - \frac{\lambda_1}{M})\sigma_v^2 + (N - \frac{\lambda_2}{M})\sigma_e^2$$

where:

$$k_0 = \mathbf{j}_M'\mathbf{X}_{*s}(\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{j}_M/M$$

$$k_N = \mathrm{tr}((\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'\mathbf{X}_{*s})$$

$$k_T = \mathrm{tr}((\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\mathbf{X}_{*s})$$

$$\lambda_1 = \mathbf{j}_M'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{j}_M$$

$$\lambda_2 = \mathbf{j}_M'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{j}_M$$

The quadratic unbiased estimators for σ_v^2 and σ_e^2 are obtained by equating the expected values to the quadratic forms and solving for the two unknowns.

When the NOINT option is specified, the variance component equations change slightly. In particular, the following is true (Wansbeek and Kapteyn, 1989):

$$E(q_N) = (T + k_N)\sigma^2 + T\sigma_v^2 + M\sigma_e^2$$

$$E(q_T) = (N + k_T)\sigma^2 + M\sigma_v^2 + N\sigma_e^2$$

The input data set used by the PROC PANEL must be sorted by cross-section and then by time within each cross-section. It also verifies that the time series ID values are the same for all cross-sections.

PROC PANEL was used in the following SAS code for the dataset *panel*:

```
proc sort DATA=panel;
by FICTIV YEAR;
run;
proc PANEL DATA=panel;
id FICTIV YEAR;
model net_result_euro=CA_Euro TA_Euro TCP-Euro D1_Euro nrsal/
rantwo vcomp=wk;
run;
```

In the above code we have used the following notations

- *CA_Euro* stands for Net Turnover;
- *TCP_Euro* stands for total owners' equity;
- *D1_Euro* stands for long term debt;
- *nrsal* stands for number of employees;
- *FICTIV* stands for company identifier;
- *YEAR* stands for year of record.

All the above financial indicators were expressed in Euro to ensure comparability among time.

5. Main findings

Our statistical-econometric analysis shows that the evolution of companies did not vary dramatically from 2002 to 2007. Every year the records show more or less the same number of companies, but not the same ones. As the initial assumption indicated and was confirmed, not many companies that there were present in 2002 were still striving in 2008. As it can be gleaned from the graph the dynamics of the apparition and disappearance of companies is quite pronounced.

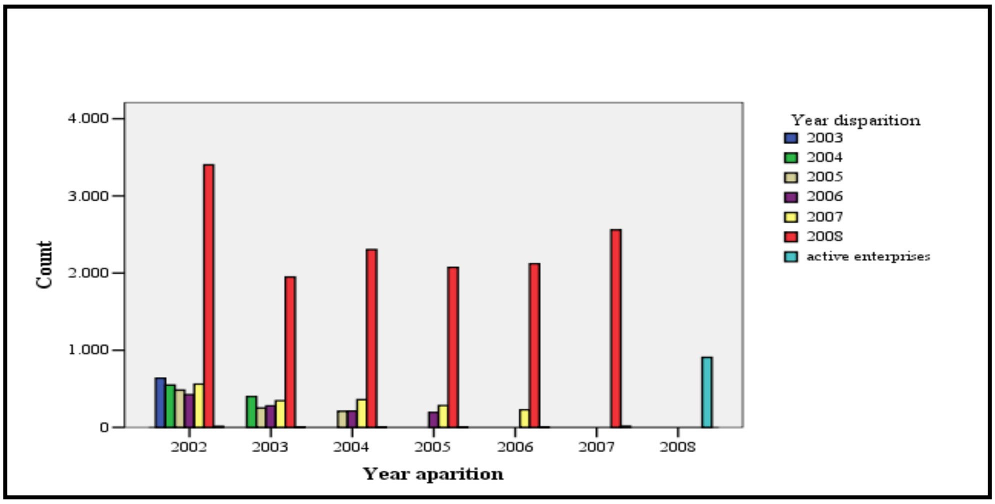

Figure 1. Companies dynamics by year of apparition and disappearance

There was a dramatic decrease in the observed number of companies in 2008 explained by the change in regulation of SMEs that lead to bankruptcy of a considerable number of companies. The companies that have been set up in 2008 were ruled out from the database, as we could not have had the model running without at least two years of observation for each company.

The main results of the model are shown in the following PROC PANEL output:

Table 1: Output of PANEL PROCEDURE

Variable	Estimate	Standard Error	t Value	Pr > |t|
Intercept	5808.32	872.5	6.66	<.0001
CA_Euro	0.1	0	132.23	<.0001
TA_Euro	-0.15	0	-113.94	<.0001

TCP_Euro	0.2	0	107.26	<.0001
D1_Euro	0.14	0.01	26.43	<.0001
nrsal	-276.02	16.69	-16.54	<.0001

Fit Statistics			
SSE	60349.29	**DFE**	61882.00
MSE	0.98	**Root MSE**	0.99
R-Square	0.32		

Hausman Test for Random Effects		
DF	**m Value**	**Pr > m**
5.00	678.05	<.0001

There is a direct effect of sales and owner's equity on the net result that confirms for the Romanian ICT based businesses the main findings in the literature. Paradoxically the long term debt increases the profitability of the companies, but as suggested in the literature this type of companies use long term debt to increase their market share and sustain the customer base. As the ICT based companies are human capital intensive, it is not surprising that the total assets do not directly affect the dependent variable. The only bizarre fact is the indirect relationship between the number of employees and the company's net result. Usually the number of employees is a good proxy for the size of the company and from previous research resulted that the bigger the company is therefore stable, the bankruptcy risk is lowered. But, given the fact that the majority of the most gifted employees are not recorded as such, but paid as "authorized natural person" to take advantage of some tax benefits, the obtained result does not contradict the main findings from the literature review.

6. Conclusions and limitations

Although some previous identified limitations have been alleviated, namely the current panel design of data, the use of the PANEL procedure that accommodates an unbalanced panel, there are still some modelling limitations to be further researched. A more elaborate model should be built to account for the newly created variable 'years of existence" that will allow the identification of certain regressors that differentiate between the companies which have survived for at least 7 years and the ones who have survived merely two.

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] **Banerjee S., Kauffman J. R., Wang, B. (2006),** ***Modeling Internet firm Survival Using Bayesian Dynamic Models with Time-varying Coefficients***; *Electronic Commerce Research and Applications 6, 332-342;*
[2] **Greene , C., Ah-Sen J. (1996),** ***Unbalanced Panel Data Estimation Using SAS/IML Software***, *Statistics, Data Analysis and Modelling;*
[3] ***Hempell, T., van Leewen, G., van der Wiel, H., (2004), ICT, Innovation and Business performance in Services: Evidence for Germany and the Netherlands,*** *ZEW Discussion paper no. 04-06;*
[4]**Kaniovski S., Peneder, M. (2008),** ***Determinants of Firm Survival; A Duration Analysis Using the Generalized Gamma Distribution*** ;*Empirica* 35:41-58*;*
[5]**Kauffman J. R., Wang B. (2002),** ***Duration in the Digital Economy;*** *Proceedings of the 36th Hawaii International Conference on System Sciences*;
[6]**Kennedy, P. (2009),** ***A Guide to Econometrics***; *6E, Blackwell Publishing;*
[7]**Park, H.M. (2009),** ***Linear Regression Models for Panel Data Using SAS, Stata, LIMDEP, and SPSS***; *Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University;*
[8]**Wansbeek, T. and A. Kapteyn (1989),** ***Estimation of the Error Components Model with Incomplete Panels***. *Journal of Econometrics 41, 341-361;*
[9]http://support.sas.com/documentation/cdl/en/etsug/60372/HTML/default/panel_toc.htm.