Bucharest Academy of Economic Studies
Faculty Of Finance, Insurance, Banking and Stock Exchange
Doctoral School of Finance and Banking

Conditional probability of default methodology

- Msc Student: Tudoran Daniela

Scientific coordinator: PhD Professor Moisă Altăr

Contents

Motivation

- In this paper we propose to ascertain the macroeconomic determinants of the probabilities of loan defaults in Romania's case. In our exercise we use the conditional probability of default methodology (CoPoD), proposed by Miguel A. Segoviano Basurto.
- Moinescu (2012) "A general message that can be drawn from the study of past boom-bust credit cycles, is that these cycles show high resemblances, and therefore lessons can be drawn from an analysis of their commonalities."
- In the recent literature the models that try to predict episodes of financial instability in the financial system use two main approaches:
    - ***early warning systems***
    - ***multivariate probit*** *or* ***logit regression models***
- Logit and probit models use OLS estimation in order to determine credit risk as a function of macroeconomics variables. When the series contains few observation OLS estimators possess large variances and are very sensitive to small changes in the data.
- Segoviano (2006)*: "CoPoD (conditional probability of default methodology) improves the measurement of the impact of macroeconomic developments on loans' credit risk by making a twofold contribution.*
    - *First,* ***econometrically****, the proposed methodology, based on the Jaynes (1957) generalized maximum entropy rule (GME), recovers estimators that in the setting of finite samples are superior to OLS estimators under the Mean Square Error (MSE) criterion.*
    - *Second,* ***economically****, on the basis of a hypothesis that is consistent with economic theory and empirical evidence, a procedure is proposed to select the set of explanatory variables that have a significant effect on loans' credit risk."*

Literature Review

- Wilson (1997) assume that the relation between credit risk and the macroeconomic background is best described by the logistic function.

- Boss (2002) concluded that for Austrian corporate and household sectors the macroeconomic determinants of credit risk are GDP, private consumption, the unemployment rate and industrial production as well as the ratios of equipment investment to GDP and exports to GDP ;

- Segoviano (2006) using the conditional probability of default methodology investigates the probability of default for small business enterprises as function of macroeconomics variables. The dataset used in this study is represented by non-performing loans ratio registered in Mexico and Norway.

- Moinescu (2013) using panel data of nonperforming loans from Central and Eastern European countries have found that : “Real GDP growth and the change in output gap were almost equally important. Moreover, money market interest rate 3M, inflation and exchange rate changes are also statistically relevant. However, fixed effects were not found statistically relevant. This result indicates that credit discipline is homogeneous across CEE region.”

GME–General aspects

- In 1948 Shannon introduce the concept of entropy as a measure of uncertainty in a random variable;
- After a decade E.T. Jaynes lay down the maximum entropy principle according to which if we maximize the Shannon’s entropy function and we take into consideration our current state of knowledge about the possible outcome of a random variable we may select the probability distribution function, which leaves us the largest remaining uncertainty , the maximum entropy, consistent with our constraints, without introducing additional assumptions or biases into our calculations;
- More recently Golan and Judge (1996) introduce the estimators obtained through Generalized Maximum Entropy Rule (GME). According to their studies in this area, applying GME lead us to robust estimators even when we have an ill-conditioned or ill-posed problem;
- Randall and Campbell (2005) in their study about determinants of poverty rate in California used GME and OLS. Based on a Monte Carlo experiment they concluded that GME estimators present the smallest variance.
- Recently Ferreira and Dionisio (2012) used GME, as an alternative to OLS in the estimation of utility function. The study results revealed that GME estimators are more accurate than OLS estimators.

The model (1)

- Starting from Merton model (1957) : borrower's default occurs if the return on a borrower's assets rate of return: s_T, falls below a certain barrier a_t, the default threshold:

$$P(Y_t = 1) = P(s_T < a_t)$$

- As the borrower's assets rate of return asset is standard normally distributed, we may say that the $P(s_T < a_t)$ is also standard normally distributed and we may formulte the probability of default as follows:

$$PoD = \Phi\left(a_t^i\right) \ (1)$$

where $\Phi(.)$ is the standard normal cumulative distribution function (*cdf)*

- Since the observations in the vector of PoD are restricted to be in the interval (0,1) we make the following transformation:

$$a = \Phi^{-1}(PoD)$$

- We want to estimate the default threshold: a as a function of macroeconomic variables: *X,*

$$a = X\beta + e \ (2),$$

- We express each estimator β as discrete variable with 2≤M<∞ possible outcomes:

$$\beta_k = \sum_{m=1}^{M} z_{km} p_{km} \ with \ \sum p_k = 1 \ and \ p_k \in (0,1)$$

The model (2)

- We reformulate equation (2) as follows:

$$a = XZp + Vw \ (3)$$

- Following Judge and Golan (1996) Generalized Maximum Entropy rule (GME): we select p and w in order to maximize the entropy function:

$$\text{E(p,w)} = -\left[\sum_{k=1}^{K}\sum_{m=1}^{M} p_m^k \ln(p)_m^k\right] - \left[\sum_{t=1}^{T}\sum_{j=1}^{J} w_j^t \ln(w)_j^t\right] \ (4)$$

- Taking into consideration the following constraints:

$$a_t = \sum_{k=1}^{K}\sum_{m=1}^{M} x_{tk} z_m^k p_m^k + \sum_{j=1}^{J} v_j^t w_j^t \ (5)$$

$$1 = \sum_{m=1}^{M} p_m^k, \qquad 1 = \sum_{j=1}^{J} w_j^t \ (6)$$

- As we have a maximum problem, we may apply Lagrange function:

$$L = -\left[\sum_{k=1}^{K}\sum_{m=1}^{M} p_m^k \ln(p)_m^k\right] - \left[\sum_{t=1}^{T}\sum_{j=1}^{J} w_j^t \ln(w)_j^t\right] + \sum_{t=1}^{T}\lambda_t\left[a_t - \sum_{k=1}^{K}\sum_{m=1}^{M} x_{tk} z_m^k p_m^k - \sum_{j=1}^{J} v_j^t w_j^t\right] + \sum_{k=1}^{K}\theta_k\left[1 - \sum_{m=1}^{M} p_m^k\right] + \sum_{t=1}^{T}\tau_j\left[1 - \sum_{j=1}^{J} w_j^t\right] \ (7)$$

The model (3)

- In order to recover the probability vector *w* and *p* we maximize the Lagrangian function described in equation (7) and we obtain the following entropy solution:

$$\hat{p}_m^k(\hat{\lambda}) = \frac{\exp\left[-\sum_{t=1}^{T}\hat{\hat{\lambda}}_t x_{tk} z_m^k\right]}{\sum_{m=1}^{M}\left[exp\left[-\sum_{t=1}^{T}\hat{\lambda}_t x_{tk} z_m^k\right]\right]} \quad (8)$$

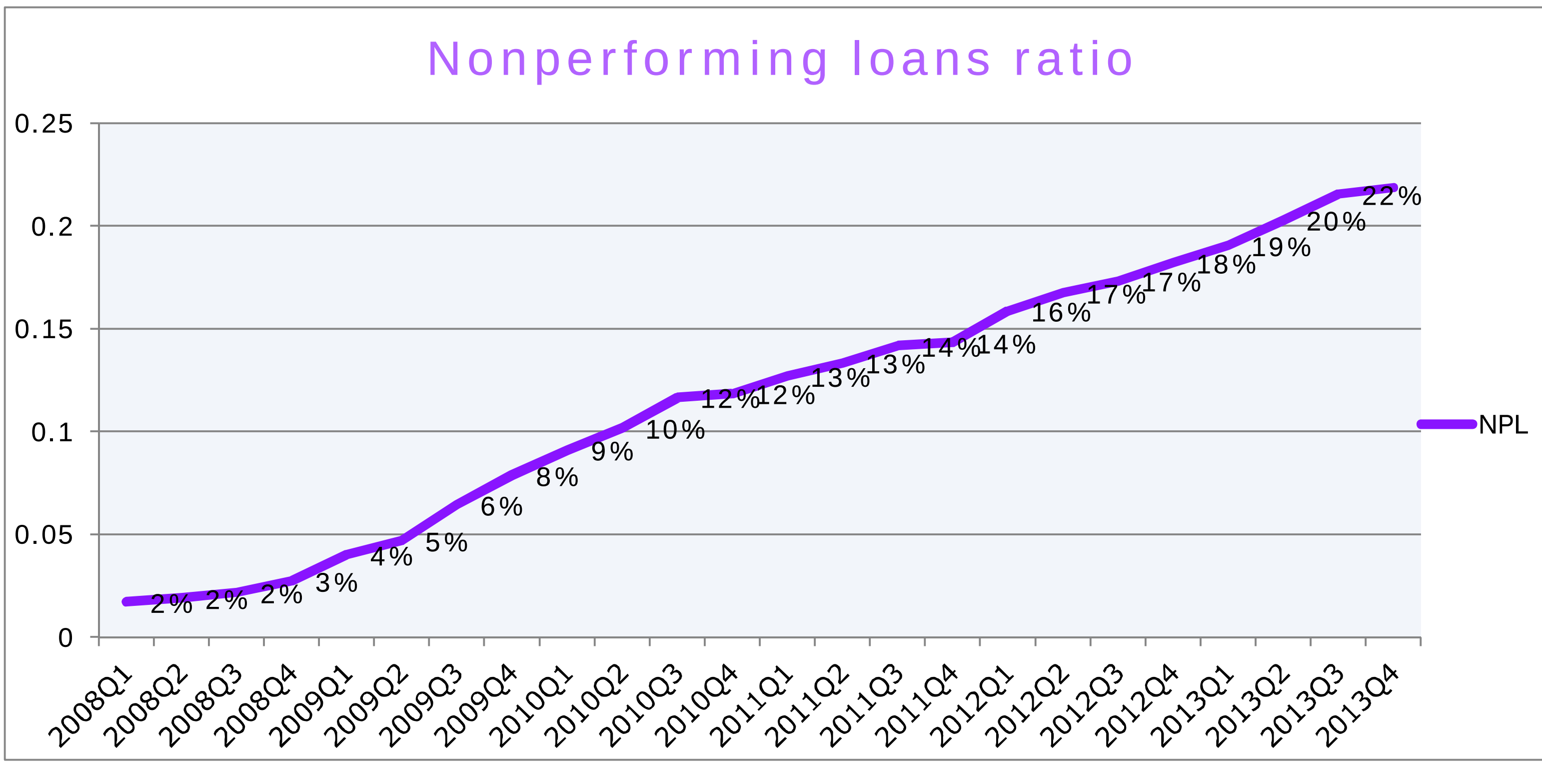
$$\hat{w}_j^t(\hat{\lambda}) = \frac{\exp\left[-\hat{\hat{\lambda}}_t v_j^t\right]}{\sum_{j=1}^{J}\left[exp\left[-\hat{\hat{\lambda}}_t v_j^t\right]\right]} \quad (9)$$

- Now we may state the estimator under GME as:

$$\hat{\beta} = Z\hat{p} \quad (10)$$

Data (1)

- As a proxy for the empirical frequencies of default (PoD) we selected the nonperforming loans ratio, at aggregate level registered by the Romanian credit institutions in the period 2008 Quarter1 -2013 Quarter 4.
- The nonperforming loan ratio is defined as the report between nonperforming loans and the total gross value of loans. A nonperforming loan is a credit with at least 90 days past due or with a debtor declared insolvent.



Data (2)

Period 2005 Quarter 1- 2013 Quarter 4

- Real GDP (GDP)
- Foreign Exchange rate RON/EUR;
- Long Term Interest rate-EMU convergence criterion bond yeld (LTIR)
- Ratio of Government Debt to GDP (GVDOVGDP)
- Ratio of Consumption to GDP (CONOVGDP)
- Share Price Index- Bucharest Stock Exchange Index (BET);
- Real aggregate credit to private sector (CRE);
- Ratio of real aggregate credit to GDP (CREOVGDP);
- Foreign direct investments (FDI);
- Ratio of foreign direct investments to GDP (FDIOVGDP);
- Ratio of M2 monetary aggregate to foreign exchange reserves (M2OVFXR);
- ROBOR 3M interest rate (RBR3);
- Unemployment rate (UNEM);

Implementation (1)

- We have run multiple OLS regression and we explored different combination of variables as number and different combinations for the lag. We selected 3 models based on R-squared criteria, Akaike criteria and models that have economic significance

	Model 1		Model 2		Model 3	
	β OLS	p-value	β OLS	p-value	β OLS	p-value
C	-1.4159	0.0000	-1.0414	0.0000	-1.1871	
GVDovGDP-11					2.1199	0.0024
GVovGDP-10	4.0641	0.0027				
GVDovGDP-8						
GDP-11			-22.5529	0.0000	-11.6000	0.0041
BET-9			-1.3101	0.0009	-1.0952	0.0005
BET-8	-1.1163	0.0000				
FX-7					2.4894	0.0330
FDIovGDP-8			-4.4041	0.0168		
FDI-5					-3.5864	0.0003
M2ovFXR-3			1.7025	0.1360		
R-Squared	0.6150		0.6900		0.8400	
Adjusted R-Squared	0.5780		0.6250		0.7959	
F-statistic	16.7000		10.6000		18.9000	
Prob (F-Statistic)	0.0004		0.0001		0.0000	

Implementation (2)

Based on the previous 3 models we will generate GME estimators as follows:

- For each model we will estimate the vector of coefficients β_{OLS} using equation (2):

$$a = X\beta + e$$

- In order to simulate β_{OLS} parameters distribution we used Bootstrap with 10,000 trials ;
- As we have said earlier in the model specification section we try to express each estimator β as discrete variable with 2≤M<∞ possible outcomes. We choose M=5;
- With the distributions of β_{OLS} coefficients obtained using Bootstrapping, we calculated the standard errors, σ, for each coefficient, and used these standard errors to define the bounds of Z, using the three-sigma rule :

$$z_{k1} = -3\sigma, z_{k5} = 3\sigma, z_{k3} = \mu, z_{k2} = \frac{z_{k1} + z_{k3}}{2} \text{ and } z_{k4} = \frac{z_{k5} + z_{k3}}{2}$$

- We defined the vector *Z* and *W*, we know *X* , the macroeconomics explanatory variables and *a,* vector of observations, transformation of the Proability of default (nonperforming loans ratio)
- We apply the Generalized Maximum Entropy rule in order to recover proabability vectors p and w and then we obtain the GME estimators:

$$\hat{\beta} = Z\hat{p}$$

Implementation (2)

	Model 1		Model 2		Model 3	
	β OLS	β GME	β OLS	β GME	β OLS	β GME
C	-1.4159	-1.3515	-1.0414	-1.0289	-1.1871	-1.1627
GVDovGDP-11					2.1199	1.5872
GVovGDP-10	4.0641	2.9128				
GVDovGDP-8						
GDP-11			-22.5529	-17.9150	-11.6000	-9.6267
BET-9			-1.3101	-0.9608	-1.0952	-0.7908
BET-8	-1.1163	-0.6623				
FX-7					2.4894	1.6825
FDIovGDP-8			-4.4041	-3.3605		
FDI-5					-3.5864	-2.9162
M2ovFXR-3			1.7025	1.1825		

OLS vs GME

In order to quantify the small sample properties of GME estimator we performed the following experiment:

- We used Model 1 specification: the matrix of explanatory variables, *X* and the vector of observations *a* to compute the vector of coefficients βols .
- We used the matrix *X* , the vector of observations *a* and assuming the values of the OLS estimators in order to obtain the vector of residuals:
$$R = a - X\beta_{OLS}$$
- We used Bootstrap with 10,000 random trials. Each trial represent a random value of *X* and *R.*
- With this element we computed simulated values for the vector of observation *a* and we were able to recover the OLS and GME estimators.
- Based on the distribution of OLS and GME estimators obtained we computed Mean Square Error criterion:
$$MSE[\widehat{\beta}] = E[(\widehat{\beta} - \beta)]$$

	β 1	**β** 2	**β** 3
Variance (difference)	10%	37%	36%
MSE (difference)	89%	141%	89%

Conclusions

- This methodology makes a twofold contribution.
  - From the econometric point of view, we managed to produces estimators that, as per small samples properties, are superior to OLS estimators.
  - And from an economic point of view, based on economic arguments and empirical evidence from the speciality literature, selected a set of explanatory variables that have a significant effect on loans' credit risk.
- Our model limitations could be:
  - our model estimates at aggregate level the probability of default and thus we may obtain less accurate estimates than those obtained with industry-specific default rates or with rating portfolio rates.
  - the nonperforming loans ratio may contains some noise.
- Further potential utilisations:
  - Forecast the nonperforming loan ratio
  - Stress testing exercises

References (1)

- Allen, F., Gale, D. (1998), “Bubbles and Crises”, The Economic Journal, 110
- Borio, C., Lowe, P. (2002), “Asset Prices, Financial and Monetary Stability: Exploring the Nexus”, BIS Working Paper 114
- Boss, M. (2002), “A Macroeconomic Credit Risk Model for Stress Testing the Austrian Credit Portfolio”,Financial Stability Report 4, Oesterreichische Nationalbank., p. 64-131.
- Caprio, G., Klingebiel, D., “Bank Insolvency: Cross-Country Experiences”, Washington DC, World Bank.

- Caprio, G., Klingebiel, D., “Bank Insolvency: Cross-Country Experiences”, Washington DC, World Bank.

- Calvo, G., Mendoza, E. (1996), “Petty Crime and Cruel Punishment: Lessons from the Mexican Debacle”, American Economic Review, 86(2), p. 170-175.
- Dooley, M. (1997), “A Model of Crises in Emerging Markets”, NBER, Working Paper 6300.
- Fiori Roberta, Antonella Foglia, and Simonetta Iannotti. 2007. “Estimating Macroeconomic Credit Risk and Sectoral Default Rate Correlations for the Italian Economy”, Paper given at the 2nd Expert Forum on Advanced Techniques on Stress Testing, International Monetary Fund and the Nederlandsche Bank, Amsterdam, The Netherlands, October 23–24, 2007
- Golan, A., Judge, G., Miller, D. (1997), “Maximum Entropy Econometrics”, Series in Financial Economics and Quantitative Analysis, Wiley, p. 85-106.

- Golan, A., Judge, G., Miller, D. (1997), “Maximum Entropy Econometrics”, Series in Financial Economics and Quantitative Analysis, Wiley, p. 85-106.

- Goodhart, C., Hofmann, B., Segoviano, M. (2004), “Bank Regulation and Macroeconomic Fluctuations”, Oxford Review of Economic Policy, Vol. 20, No. 4, p. 591-615.
- Horowitz, J. L. (2001), “The Bootstrap”, Handbook of Econometrics, Vol. 5, J.J. Heckman and E.E. eds., Elsevier Science B.V., Ch. 52, p. 3159-3228.

References (2)

- Jimenez, G., Saurina, J. (2004a), “Credit Cycles, Credit Risk, and Prudential Regulation”, Mimeo, Banco de España.
- Kaminsky, G., Lizondo, S., Reinhart, C. (1998), “Leading Indicators of Currency Crises”, IMF Staff Papers, 5, 1, 1-48.
- Kaminsky, G., Reinhart, C. (1996), “Banking and Balance-of-Payments Crises: Models and Evidence”, Board of Governors of the Federal Reserve, Washington, DC, Working Paper.
- Kaminsky, G., Reinhart, C. (1999), “The Twin Crisis: The Causes of Banking and Balance-of-
- Payments Problems”, American Economic Review, 89, 473-500.
- Kaminsky, G., Schmukler, S. (2003), “Short-Run Pain, Long-Run Gain: The Effects of Financial Liberalization”, NBER Working Paper 9787.
- Kiyotaki, N., Moore, J. (1997), “Credit Cycles”, Journal of Political Economy, 105, 211-248.
- Segoviano M. (2004), “A System for Measuring Credit Risk When Only Limited Data is Available”, in Credit Risk International, London, October 2004.
- Segoviano, Miguel, 2006a, “The Conditional Probability of Default Methodology,” Financial Markets Group, London School of Economics, Discussion Paper 558.
- Virolainen, Kimmo. 2004. “Macro Stress Testing with a Macroeconomic Credit Risk Model for Finland“, Bank of Finland, Discussion Papers, No.18
- Wilson, T. (1997a), “Portfolio Credit Risk I”, in Risk, Vol. 10, No. 9.
- Wilson, T. (1997b), “Portfolio Credit Risk II”, in Risk, Vol. 10, No. 10.

Thank You !