5th International Conference
of the Financial Engineering and Banking Society (FEBS)

PD – LGD Correlation Study

Evidence from the Russian Corporate Bond Market

Ermolova Maria,
Bachelor, Higher School of Economics

Henry Penikas, PhD.
Project Manager, Alfa-Bank
Associate Professor, Higher School of Economics

June, 13, 2015,
Nantes, France

The authors would like to thank to Vadim Arzamasov for his valuable comments and helpful discussion and to Ekaterina Antonova for the data provided from the research about parameter LGD for Russian corporate borrowers.

Presentation Outlay

1) Problem Setting

2) Review of the Literature

3) Correlation (PD, LGD) in Regulatory Requirements

4) Research Objectives

5) Methodology

6) Estimation of PD-LGD correlation:

   - Data
   - LGD estimates
   - PD model
   - Sample to estimate corr (PD, LGD)
   - Results

7) Conclusion

© Penikas, Ermolova 2015

Problem Setting

- $Capital\ Adequacy\ Ratio\ (CAR) = \frac{K}{RWA}$, where
- $RWA = 12.5 * EAD * \left\{LGD * N\left(\sqrt{\frac{1}{1-R}} * G(PD) + \sqrt{\frac{R}{1-R}} * G(0.999)\right) - PD * LGD\right\} * \frac{1+(M-2.5)*b}{1-1.5*b}$
  - Source: http://www.bis.org/publ/bcbs128.pdf

Key assumption:

- Correlation (PD, LGD) = 0, or
- Correlation (PD, 1 - RR) = 0 ↔ Correlation (PD, RR) = 0

Note:
RWA = risk weighted assets; **K** = capital;
PD = probability of default; **LGD** = loss given default; **EAD** = exposure at default**;**
M = effective maturity; **b** = maturity adjustment; **R =** assets' correlation;
RR = recovery rate;
PLC = correlation between PD and LGD;
N(x) denotes the cumulative distribution function of a standard normal random variable (i.e. the probability that a normal random variable with mean zero and variance equal to 1, less than or equal to x). **G(z)** denotes the inverse cumulative distribution function of a standard normal random variable (i.e. the value x such that N(x) = z).

© Penikas, Ermolova 2015

Review of Literature

Research	Data	Result
Vasicek, 1987, 1991, 2002	Actual bank portfolio, 479 loans: maturities 6 months - 6 years; default probabilities from 0.0002 to 0.064; LGD averaged 0.54. The asset returns were generated with 14 common factors.	Model does not account for systematic correlation between PD & LGD
Frye, 2000	Moody's Default Risk Service database: data on defaults and recoveries from U.S. corporate bond data for the 1982-1997 period	• Suggested quantifying estimates of collateral damage on the example of 2 loans • Showed that bond recoveries might decline 20–25 percentage points from their normal-year average. Loan recoveries may decline by a similar amount but from a higher level. • The result is inverse relationship between realized default rates and recovery rates
Hu, Perraudin, 2002	Moody's historical bond market data	Corr (quartely RR, default rates) = -0.22 for 1983-2000 Corr (quartely RR, default rates) = -0.19 for 1971-2000
Carey, Gordy, 2003	• Moody's Default Risk Service database of bond • defaults and LGDs • Society of Actuaries database of private placement defaults and • LGDs • Standard & Poor's database of bond defaults and LGDs • Portfolio • Management Data's database of LGDs • *Time period: 1970 - 1999*	Their preliminary results: Estimates of simple default rate-LGD correlation are close to zero. Limiting the sample period to 1988-1998, estimated correlations are more in line with main results about negative correlation between PD and RR (0.45 for senior debt and 0.8 for subordinated debt). Concluded that a weak or asymmetric relationship suggests that default rates and LGDs may be influenced by different components of the economic cycle.

© Penikas, Ermolova 2015

Review of Literature (continued)

Research	Data	Result
Dullmann, Trapp, 2004	Time series of default rate and recovery rates from Standard and Poor's Credit Pro database: bond and loan information from 1982 to 1999	Found that estimates of recovery rates based on market prices at default are sensitive to changes of the systematic risk factor than are recovery rates determined at restructuring from default. Showed that choice between these 2 methods has a stronger impact on economic capital
Altman et al., 2004	Defaulted bonds' data for the period 1982-2000, which includes the relatively high-default years of 1999 and 2000,	A negative correlation between default rates and RRs Showed that both the expected loss and the unexpected loss are vastly understated if one assumes that PDs and RRs are uncorrelated
Giese, 2005	American bonds over the period from 1982 to 2001	Positive relationship between annual default frequency and average LGD
Miu, Ozdemir, 2006	Portfolio consisting of 2000 loans made 2000 different borrowers is used. The size of each loan is the same and equal to 1$.	Correlation between systematic PD and LGD was estimated on data of large corporations, therefore it was overestimated (53%). If the average market portfolios are examined where loans are secured by collateral, and that means it less connected with systemic risk, the correlation will be much weaker.
Meng, et al., 2010	1,000 long-term loans. It is assumed that the maturity of each loan is 3 years, PD=1%, the correlation between the value of assets$R_A^2 = 0{,}25$, the correlation between the value of assets and the level of the recovery is $\rho_{A,RR} = 0{,}4$; $R_{RR}^2 = 0{,}34$, average LGD=0,58.	It has been concluded the following. MTM will be lower with non-zero correlation between the parameters PD and LGD. Order price of asset's purchase decreases to reduce systemic risk return. As a consequence, spread of a credit instrument increases.
Witzany, 2011	Default and recovery data on a retail portfolio obtained from a large Czech bank over the period 2002-2008	PD x LGD correlation (0.0775) is positive with low significance (probably due to the too short observed time period)

© Penikas, Ermolova 2015

Correlation (PD, LGD) in regulatory requirements

- The recommendations of Basel II do not describe explicitly the requirements for the accounting component of the PLC in the process of calculation of capital requirements, pricing, etc.
- The proposed approach by regulators is to use stress LGD value for taking into account the relationship between PD and LGD to calculating capital adequacy is not a conservative approach.
  - Source (figure 3,4): *Meng Q. et al. Implications of PD-LGD correlation in a portfolio setting//Modeling Methodology, Moody's KMV, 2010.*

Figure 3: Unexpected losses of loan portfolio: 3 scenarios how to take into account correlation (PD, LGD)

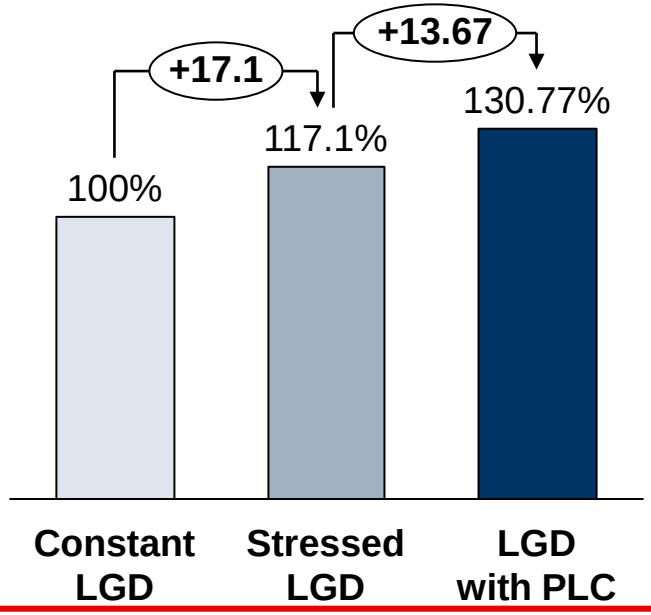
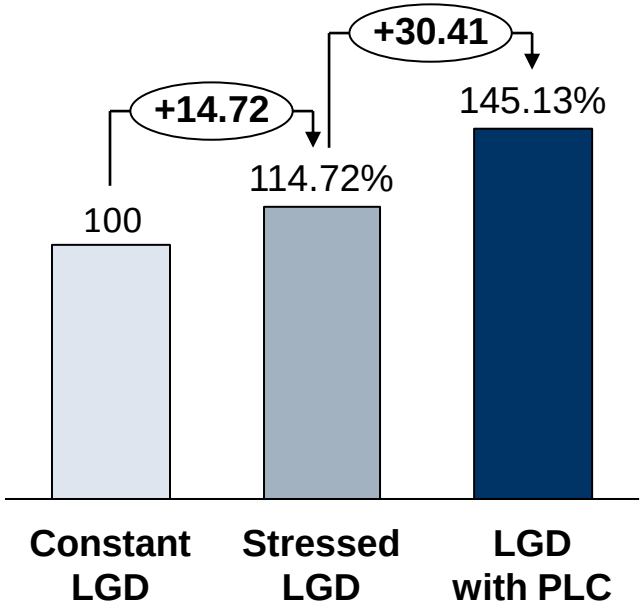


Figure 4: Economical capital of loan portfolio: 3 scenarios how to take into account correlation (PD, LGD)



- The influence of the relationship of PD and LGD measures measure significant risk that requires attention from the Bank, otherwise the Bank is exposed to risk to underestimate the level of risk of its loan portfolio.

© Penikas, Ermolova 2015

Research objectives

- There were no studies working with the problem of relationship between PD and LGD on data of Russian Corporate Bond Market. **Our research is the first based on Russian data**!

- Today the Basel regulations are being implemented within the Russian Banking Industry
  - The question of accurate estimation of economic capital is topical issue.
  - Not to take into account PD-LGD correlation leads to underestimating economic capital of a bank.

- Research objectives

  - to estimate the relationship between PD and LGD on Russian data

- Thus the today's presentation focuses **on the first stage of the study**:

  - Is there the correlation between PD and LGD on the Russian Corporate Bond Market?
  - Whether Russian banks should overcharge their economic capital due to correlation between PD and LGD?

© Penikas, Ermolova 2015

Data

Data about default cases was extracted from Financial Information Cbonds

- The database of Cbonds provides information about all emissions of debt securities.
- The annual default rate can be calculated only since 2000 (due to data unavailability). Since 2000 there are years without defaults (especially before 2007).
- The majority of defaults occurred in 2008 – 2010.
- Now there is a tendency of increase of default rate

- Financial statements were extracted from SPARK

  - SPARK (Professional market and company analysis system) is the largest database of Russian, Ukrainian and Kazakh companies. The database contains official information from over 20 different sources, including federal departments, ministries and government agencies, key mass media and companies themselves.

- **Russian data** from Cbonds and SPARK is more representative as comparison with other information systems (Moody's database and etc.)

© Penikas, Ermolova 2015

Sample to estimate corr (PD, LGD)

1. Sample from research Antonova E., 2013

59 default cases

(with available estimates of LGD)

2. Our sample

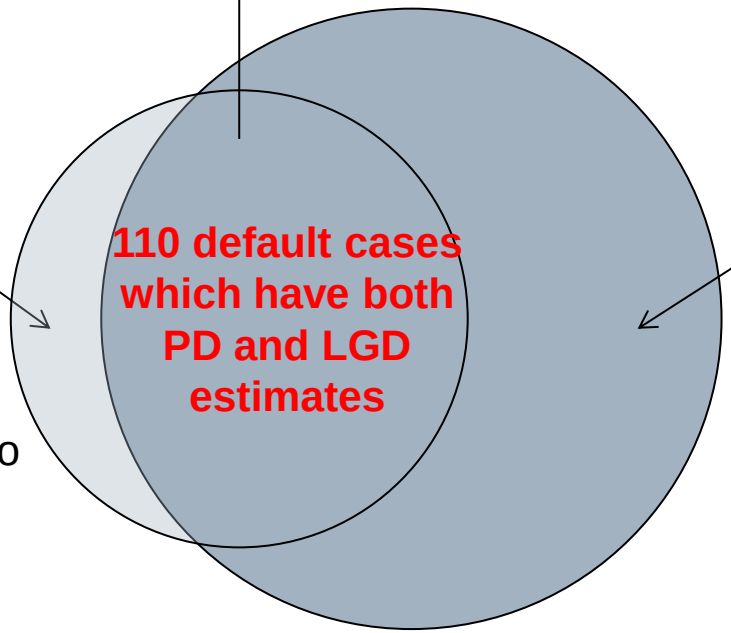
201 default cases

(with available estimates of PD)

3. Sample with 110 default observations with PD and estimated LGD
=>
Can estimate correlation (PD, LGD) on these observations

3. Our LGD sample

+ 88 default observations with available LGD



37 default cases which do not have financial statements that is needed to obtain PD estimates =>

- LGD estimates – YES,
- PD estimates - NO

54 default cases which do not have LGD estimates =>

- LGD estimates – NO,
- PD estimates - YES

© Penikas, Ermolova 2015

Methodology

	Status	PD		LGD	
		Estimated	Observed	Estimated	Observed
All years in sample	No default	$\widehat{PD_1}^{ND}$	0	$\widehat{LGD_1}^{ND}$	N/A
		$\widehat{PD_2}^{ND}$	0	$\widehat{LGD_2}^{ND}$	N/A
		…	0	…	N/A
		$\widehat{PD_n}^{ND}$	0	$\widehat{LGD_n}^{ND}$	N/A
	Default	$\widehat{PD_1}^{D}$	1	$\widehat{LGD_1}^{D}$	LGD_1
		$\widehat{PD_2}^{D}$	1	$\widehat{LGD_2}^{D}$	LGD_2
		…	1	…	…
		$\widehat{PD_k}^{D}$	1	$\widehat{LGD_k}^{D}$	LGD_n

Note:

ND = non-defaulted bonds
D = defaulted bonds
n = number of non-defaulted bonds
K – number of defaulted bonds

- **Cannot estimate** the correlation between observed PD and LGD for non-defaulted bonds
- **Cannot estimate** the correlation between observed PD and LGD for defaulted bonds
- **Can evaluate** the correlation between estimates of PD and LGD
- How to obtain estimates of PD and LGD?

© Penikas, Ermolova 2015

Methodology | LGD approaches

- **Workout LGD**

  - Discounted flow of expected cash, discounted at the date of default
  - **Inapplicability of the method to our study**: the lack of data of actual workout.

- **Market LGD**

  - The market value of defaulted bonds immediately after the moment of default.
    - B_d - market value of the bonds in 30 days after default; F - nominal value of the bonds;
    - RR_d - rate recovery after the occurrence of default.

$$LGD = 1 - RR_d = 1 - \frac{B_{d\ (+ndays)}}{F}$$

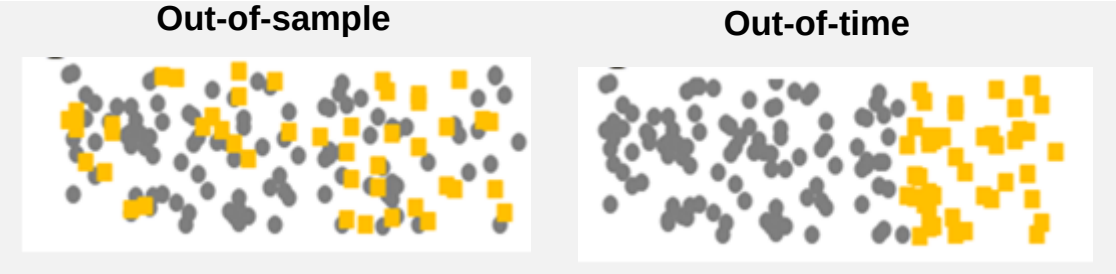
  - **We have obtained data of RR+30** for 59 companies from research Antonova, 2013

    - Source: Antonova E. The estimation of the rate of recovery in corporate bonds of Russian issuers, NRU Higher School of Economics, 2013 (http://www.hse.ru/data/2013/06/24/1299828576/Dis_Antonova.pdf)

© Penikas, Ermolova 2015

Methodology | PD model

- PD model
  - Splitting the sample in 2 ways:



Grey dots – observations from training sample
Amber dots – observations from testing sample

  - To build 3 models: Logit model.

  - Selection criteria to choose the best model:
    - Statistical significance of coefficients (independently and jointly)
    - The same direction of both sign of coefficient and sign of correlation
    - Discriminatory ability (ROC-curve) on training sample and testing sample

Step 1	Step 2	Step 3	Step 4	Step 5
Splitting the sample	*Dividing the sample*	*Building models*	*Choose the best models*	*Choose the best model*
Out-of-sample	Development sample Testing sample	Logit	Best model 1	Final model used to obtain PD estimates to calculate corr (PD, LGD)
Out-of-time	Development sample Testing sample	Logit	Best model 2	

© Penikas, Ermolova 2015

PD model: out-of-sample (logit)

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Basic	Basic + macro	Basic + macro + ln	Basic + macro + sq	Basic + macro + sq + ln
cap_ap	-0.579***	-0.574***	-0.885*	-0.761***	-1.227***
clca					0.0688***
sq_fix_ass_liab				-2.169**	
ln_cur_ass					0.173*
salta	-0.555***	-0.506***	-0.816***	-0.525***	-0.989***
cpi		0.163***	0.199***	0.169***	0.196***
ln_ap			0.257***		0.121**
intwo	1.326***	1.309***	1.545***	1.369***	1.428***
ln_cfta			-0.209***		-0.250***
gdpgain		-5.749***	-6.110***	-5.994***	-5.616***
cur_ass_liab	-0.589***	-0.425*	-0.506*		-0.879***
wcta	0.371***	0.341***	0.756***	0.430***	0.852***
ln_ar_debt					0.417***
fix_ass_liab		-1.221**	-1.566**		
cfta	-7.828*	-8.665**		-9.184**	
Constant	-1.548***	-18.53***	-28.53***	-19.22***	-28.81***
Observations	774	774	744	766	743
Pseudo R-Squared	0.229	0.273	0.364	0.262	0.408
Area Under ROC Curve (training data)	0.829	0.852	0.898	0.844	0.912
Area Under ROC Curve (testing data)	0.757	0.781	0.802	0.765	0.787
Sensitivity	20	24.35	39.25	24.35	49.53
Specificity	97.27	97.42	96.86	97.70	97.64
Correctly classified	85.79	86.56	88.58	86.68	90.71
Standard errors in parentheses					
*** p<0.01, ** p<0.05, * p<0.1					

© Penikas, Ermolova 2015

PD model: out-of-time (logit)

	(1)	(2)	(3)
VARIABLES	Basic	Basic + macro	Basic + macro + ln
cur_ass_liab			-1.029***
cfta	-10.06***	-6.116**	-6.355*
fix_ass_liab	-1.256**	-1.498***	-1.431*
salta	-0.560**	-0.571**	-1.221***
intwo	1.080***	1.114***	1.245***
cpi			0.140**
ln_ar_debt			0.472***
clca			0.0192*
ln_cur_ass			0.254***
gdpgain		-3.022***	-4.327***
cap_ap	-0.677***		
Constant	-1.328***	-1.056***	-20.05***
Observations	722	722	562
Pseudo R-Squared	0.199	0.212	0.354
Area Under ROC Curve (training data)	0.804	0.819	0.885
Area Under ROC Curve (testing data)	0.716	0.621	0.734
Sensitivity	3.810	22.86	48.51
Specificity	99.84	96.60	95.88
Correctly classified	85.87	85.87	87.37
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

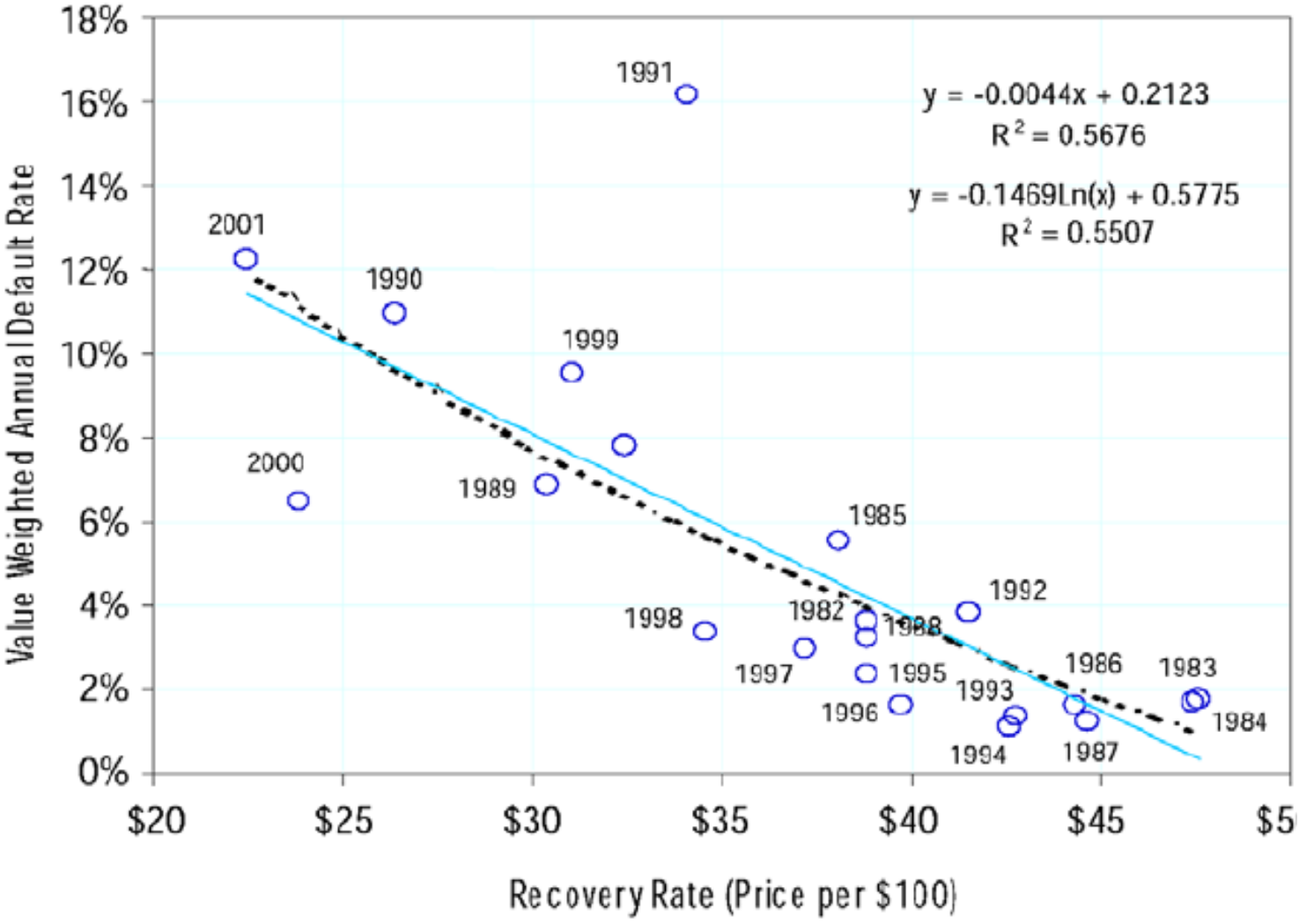
© Penikas, Ermolova 2015

PD model : using panel data

		Logit - regression			Probit - regression		
lag_default	Corr (lag_default, X)	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
salta	- 0.0051	- 0.4916	0.1832	0.7%	- 0.1726	0.0637	0.7%
intwo	0.1182	1.3060	0.2006	0.0%	0.5998	0.0885	0.0%
wcta	0.0049	0.3050	0.1304	1.9%	0.1620	0.0621	0.9%
cfta	- 0.0481	- 11.7277	4.4264	0.8%	- 4.1233	1.4524	0.5%
fix_ass_liab	- 0.0029	- 2.2917	0.5763	0.0%	- 1.0219	0.2401	0.0%
ebitta	- 0.0001	- 0.3102	0.1274	1.5%	- 0.1645	0.0620	0.8%
cpi	0.1480	0.1112	0.0297	0.0%	0.0517	0.0139	0.0%
gdpgain	- 0.0592	- 4.0334	0.8862	0.0%	- 1.9051	0.4255	0.0%
_cons		- 15.1129	3.1804	0.0%	- 7.3006	1.4791	0.0%
AUC (training sample)		85 % (standard error 0.0158)			85 % (standard error 0.0155)		
AUC (testing sample)		77 % (standard error 0.0224)			77 % (standard error 0.0220)		

© Penikas, Ermolova 2015

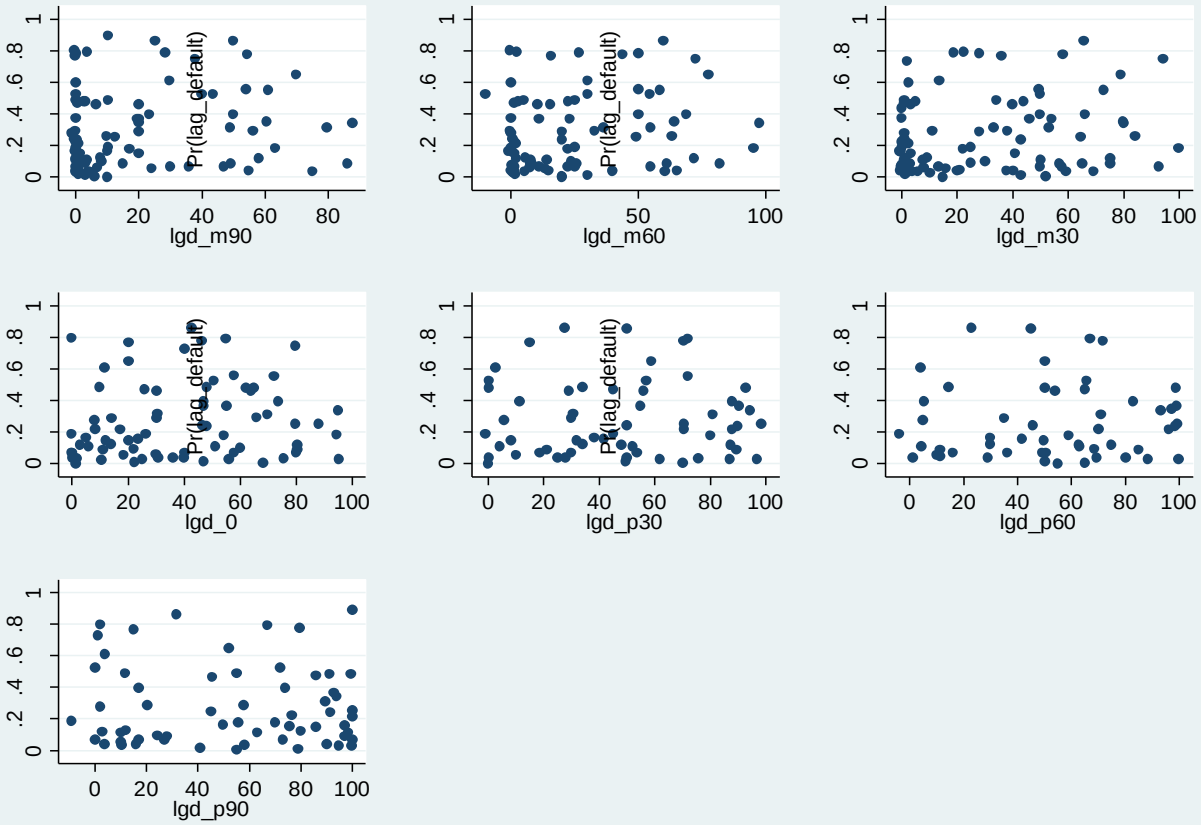
Moody's results [2010]

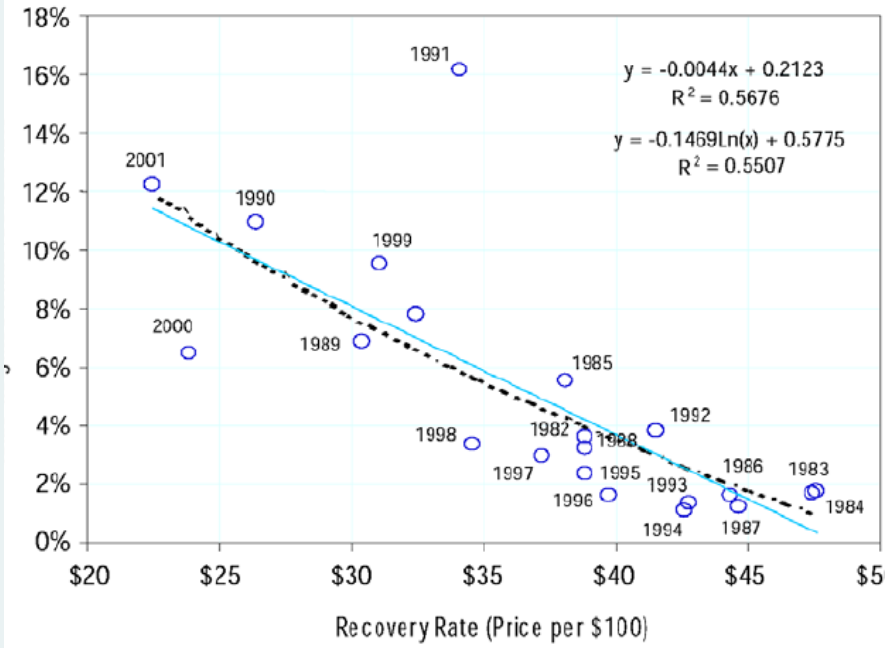


Source : *Meng Q. et al. Implications of PD-LGD correlation in a portfolio setting//Modeling Methodology, Moody's KMV, 2010.*

© Penikas, Ermolova 2015

Correlation (PD, LGD) : results (out-of-sample)



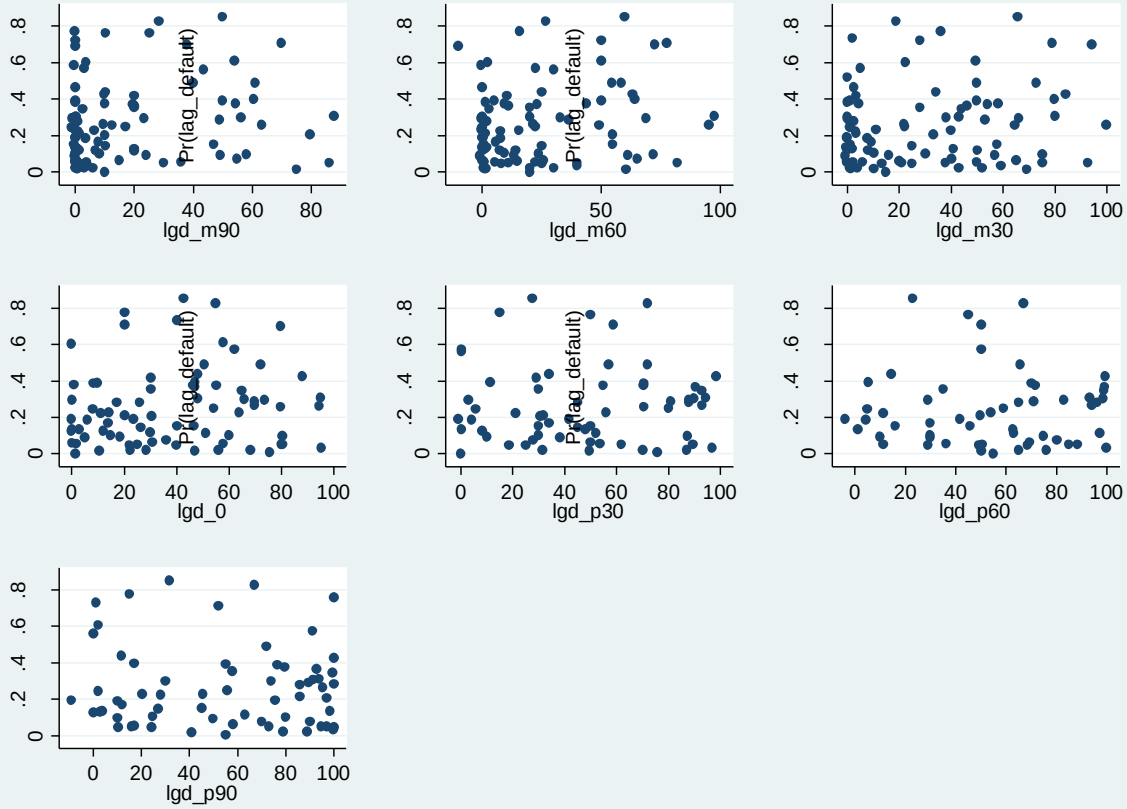


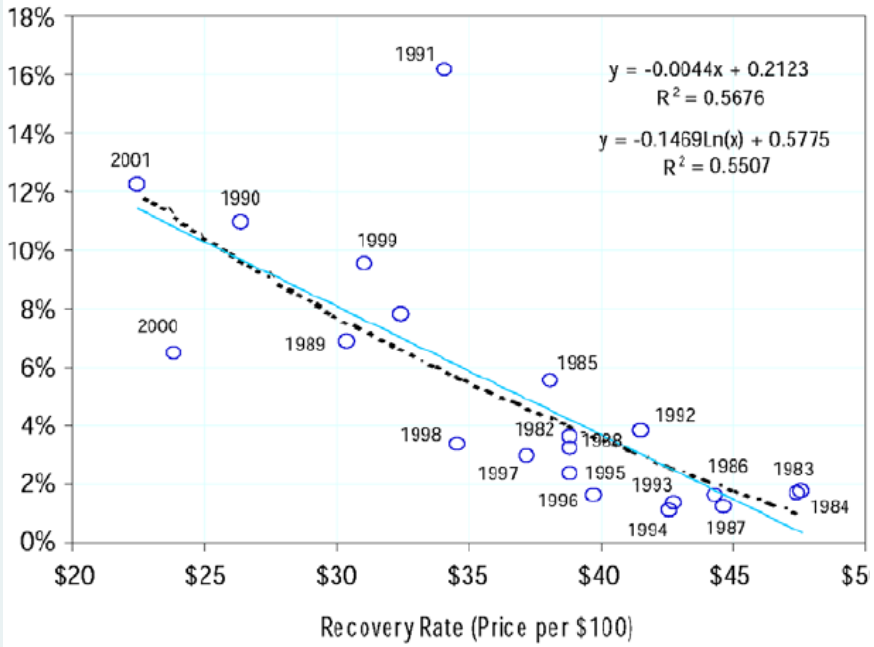
- Visually link between PD and LGD were difficult to determine

- There is no a uniform rule that the ratio of the probability of default and loss given default cannot be proved.

© Penikas, Ermolova 2015

Correlation (PD, LGD) : results (out-of-time)





- Visually link between PD and LGD were difficult to determine

- There is no a uniform rule that the ratio of the probability of default and loss given default cannot be proved.

© Penikas, Ermolova 2015

Correlation (PD, LGD) : results

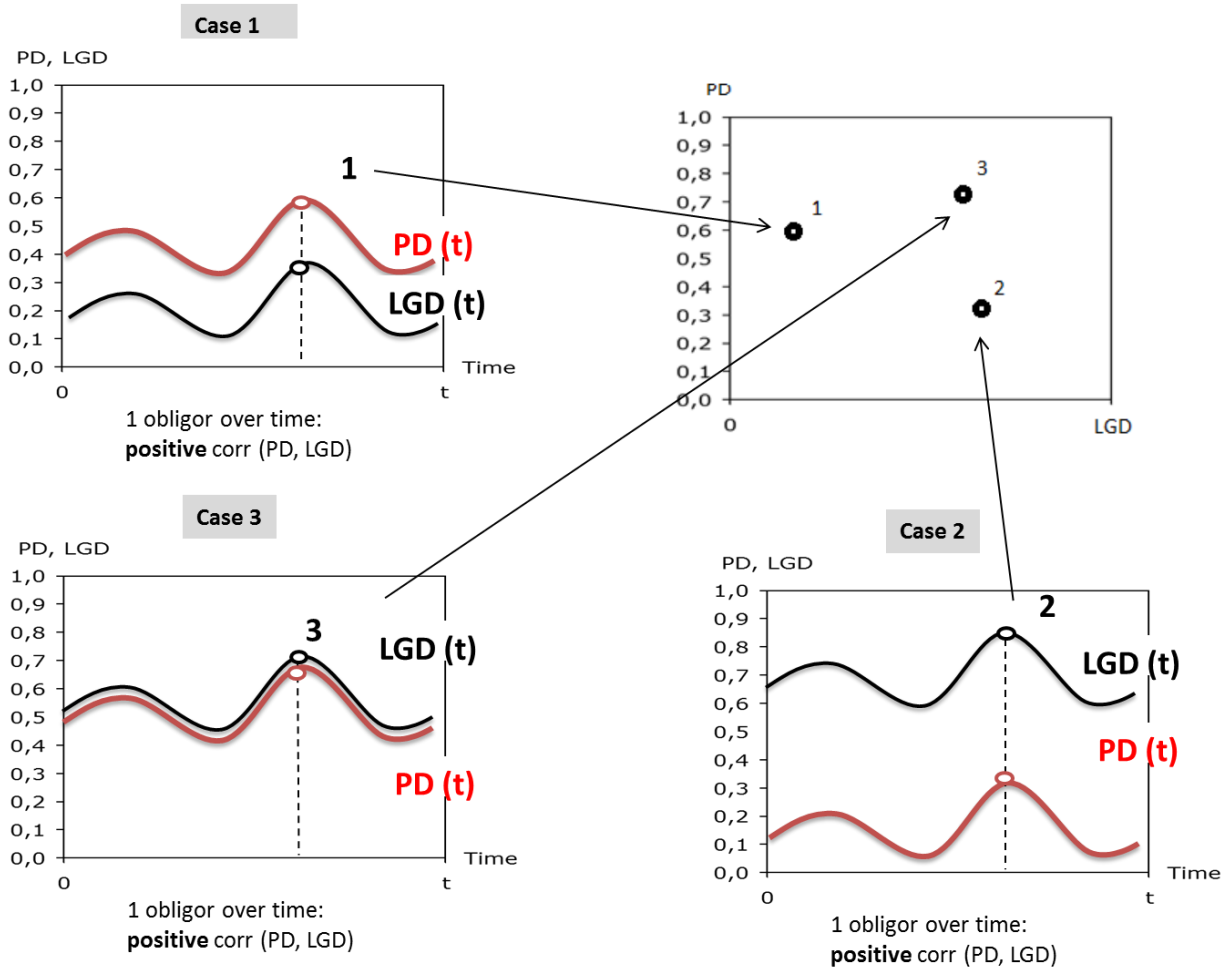
	PD: Logit, out-of-sample	PD: Logit, out-of-time	PD: Logit, RE	PD: Probit, RE	PD: Pooled logit	PD: Pooled probit
lgd_m90	0.1362	0.2254	0.2318	0.2313	0.1839	0.1859
lgd_m60	0.1402	0.2854	0.1716	0.1746	0.2082	0.2033
lgd_m30	0.1685	0.2301	0.1392	0.1378	0.1361	0.1294
lgd_0	0.1187	0.1571	0.1041	0.0983	0.1261	0.1094
lgd_p30	0.0032	0.0279	0.0628	0.0453	0.0051	-0.0234
lgd_p60	0.0216	-0.0189	0.0869	0.0732	0.0693	0.0415
lgd_p90	-0.0211	-0.0032	0.0831	0.075	0.0568	0.0406

- Visually link between PD and LGD were difficult to determine

- There is no a uniform rule that the ratio of the probability of default and loss given default cannot be proved.

© Penikas, Ermolova 2015

PD-LGD Correlation Study: the other side of the coin (1)

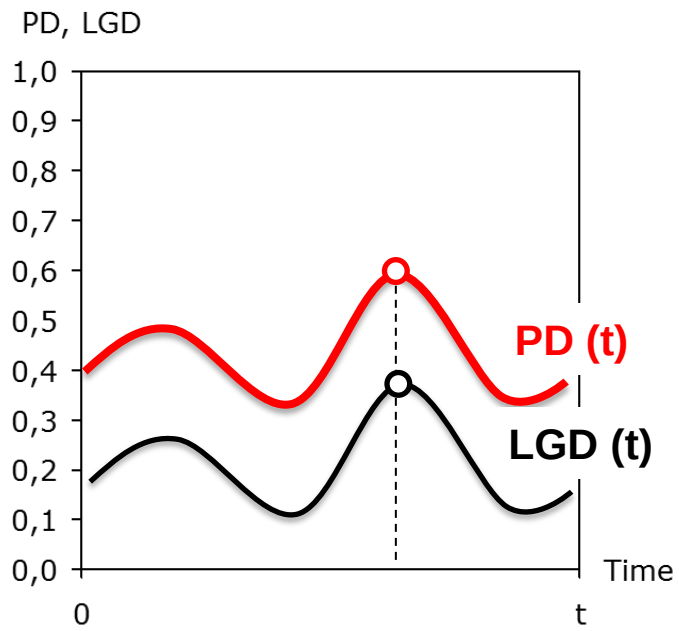


- The lack of correlation in the first coordinates does not mean that there is no correlation in the other coordinates.
- The relationship of two time series (PD and LGD) can exist for a hypothetical borrower, however, the relationship cannot be seen, if we evaluate the correlation of default for all borrowers with different years of default.

© Penikas, Ermolova 2015

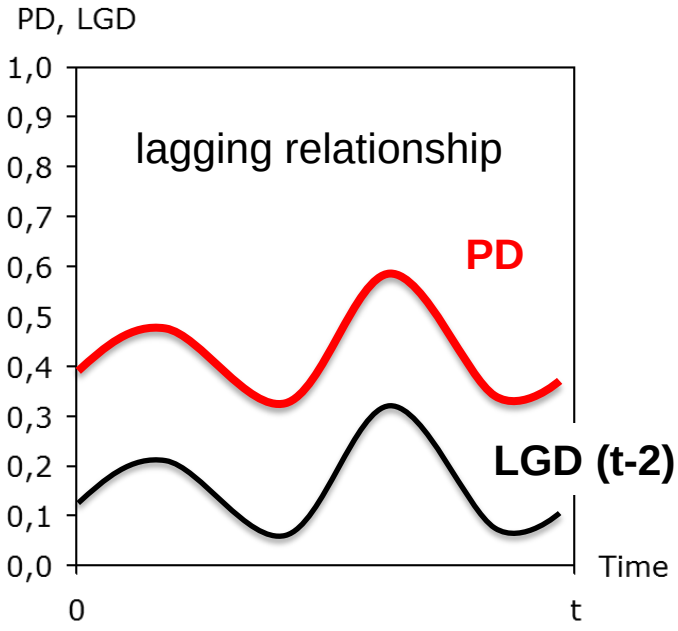
PD-LGD Correlation Study: the other side of the coin (2)

Case 1



1 obligor over time:
positive corr (PD, LGD)

Case 2

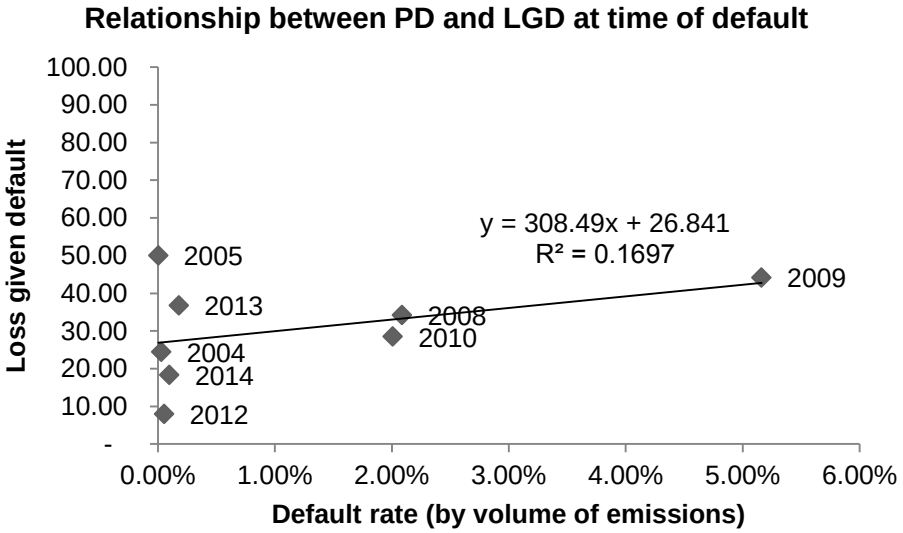


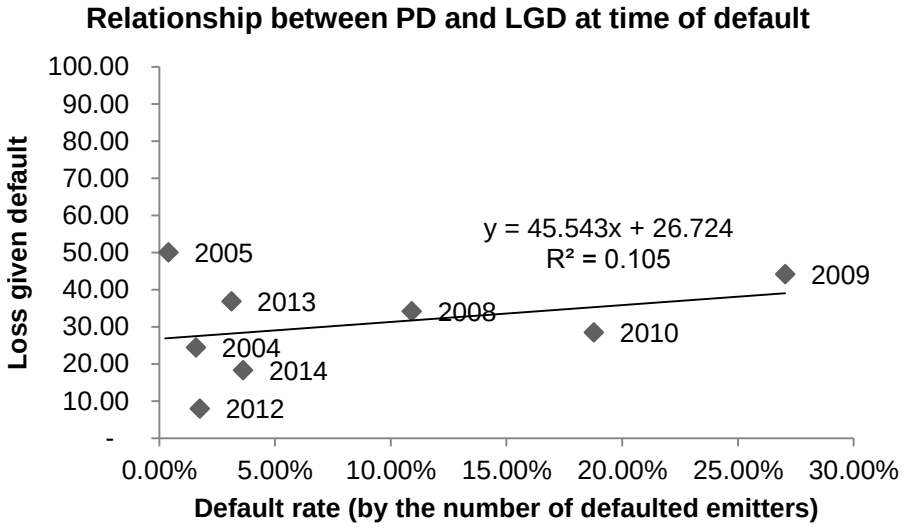
1 obligor over time:
corr (PD, LGD) **= 0 or < 0**

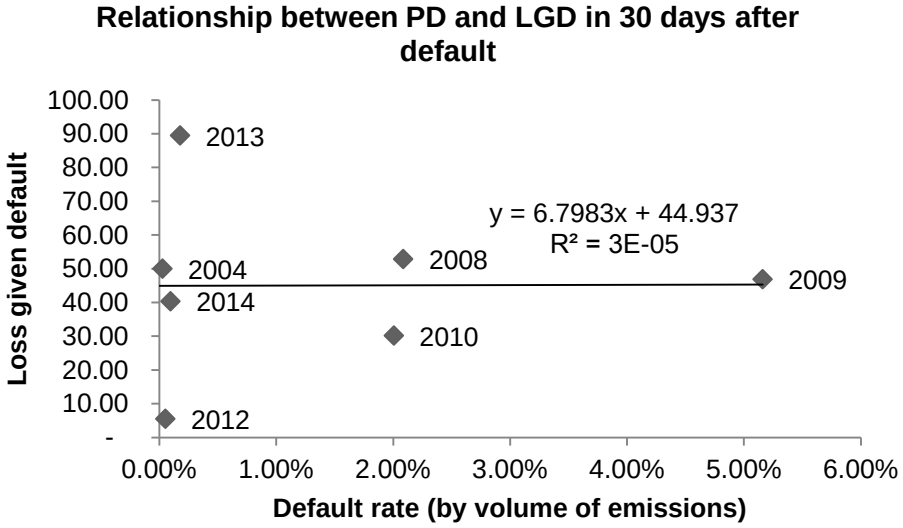
- A lagging relationship depending upon a bank's workout practices

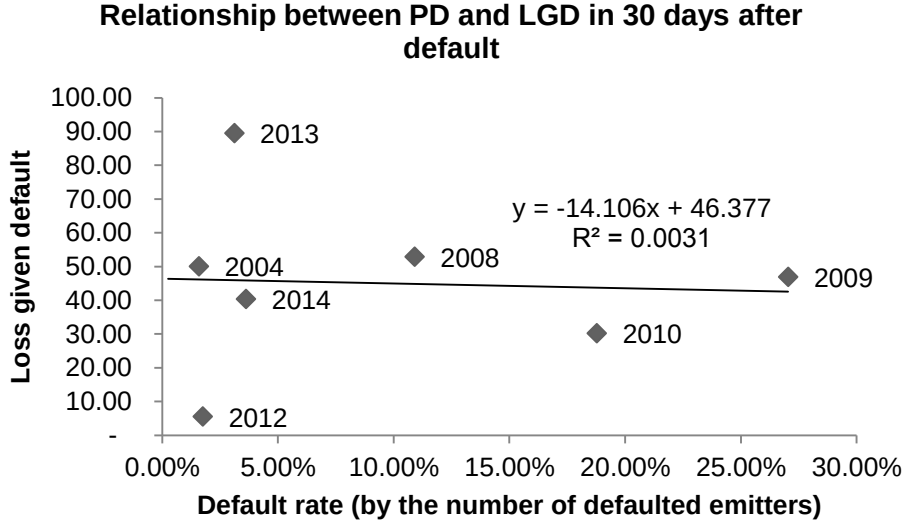
© Penikas, Ermolova 2015

PD-LGD correlation: annual default rate and LGD observed



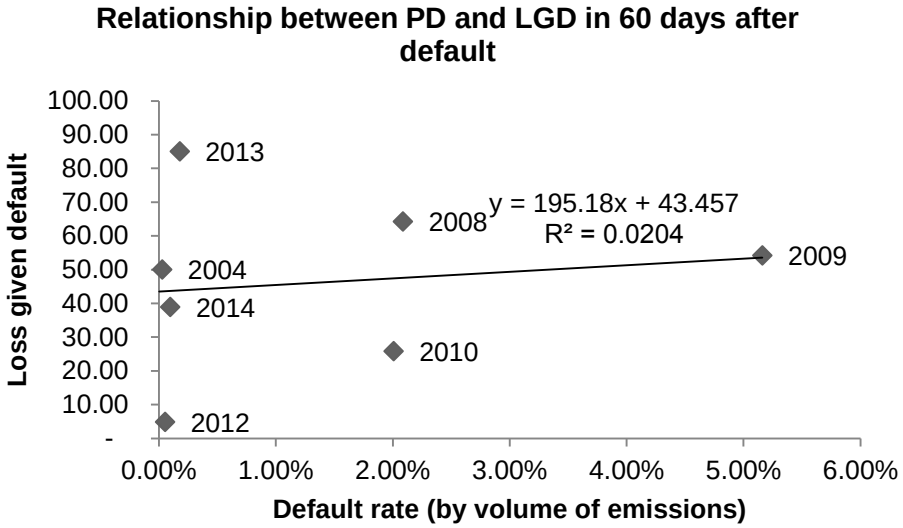


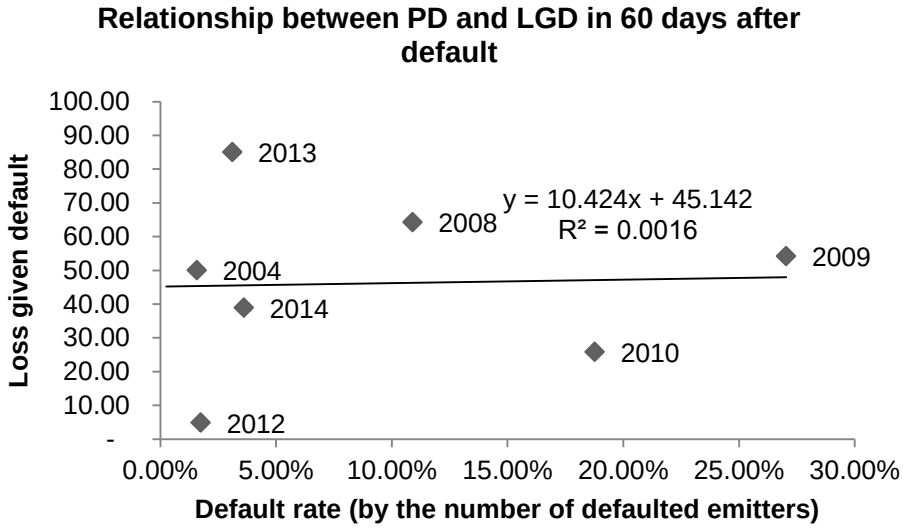


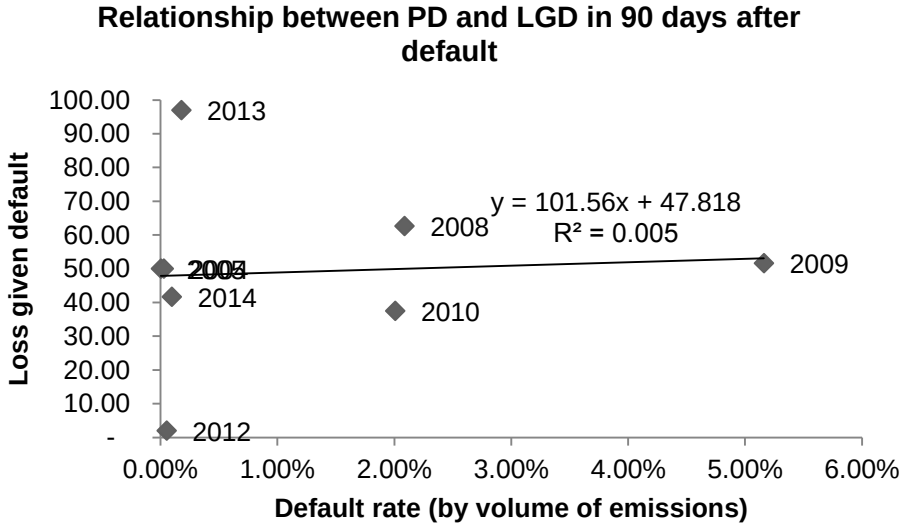


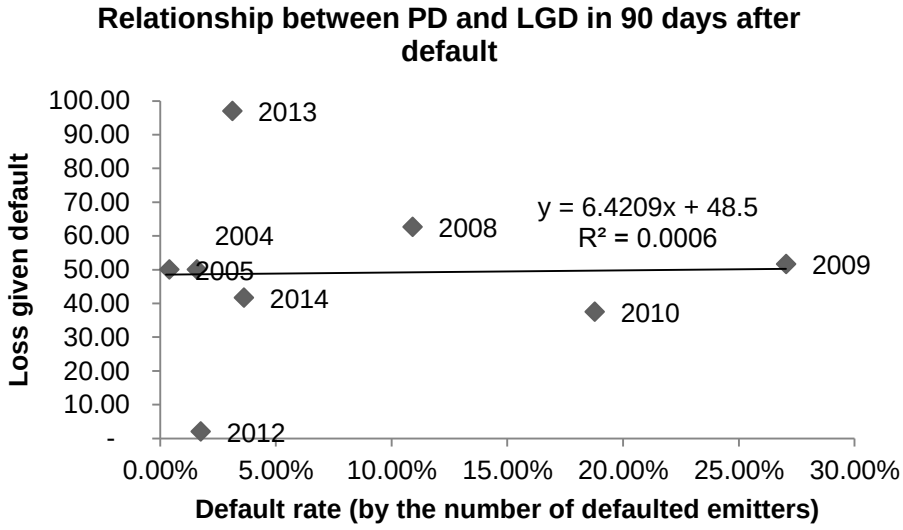
© Penikas, Ermolova 2015

PD-LGD correlation: annual default rate and LGD observed









© Penikas, Ermolova 2015

Conclusions

- Thus the today's presentation focuses on the question

  - ❑ Is there the correlation between PD and LGD on the Russian Corporate Bond Market?

  - ❑ *Answer:* the correlation between PD and LGD is weak that it can be argued that it is almost not observed on Russian corporate data.

  - ❑ Positive correlation between PD and LGD is the most frequent result of previous research. **However**, this research cannot argue that there is positive PD-LGD dependence for Russian corporate borrowers.

  - ❑ The results is different due to
    - The relationship is estimated on a sample of different borrowers with different years of default, i.e. on obligor-level but not on year-level where each observation presents a ratio of PD and LGD for year.

  - ❑ The implication would be that there is no need to overcharge capital for Russian banks compared to their counterparts from developed countries.

© Penikas, Ermolova 2015

Thank you
for your attention!

Maria Ermolova
Bachelor, Higher School of Economics
Email: m.d.ermolova@gmail.com

Henry Penikas, PhD.
Project Manager, Alfa-Bank
Associate Professor, Higher School of Economics
Email: penikas@hse.ru
URL: http://www.hse.ru/staff/penikas

© Penikas, Ermolova 2015

Appendix 1: Review of Literature in details (continued)

Research	Data	Method	Result
Dullmann, Trapp, 2004	Time series of default rate and recovery rates from Standard and Poor's Credit Pro database: bond and loan information from 1982 to 1999	Extend the single risk factor model by assuming that the recovery rates also depend on this risk factor and follow a logit–normal distribution; compared the result with two related models of Frye (2000) and Pykhtin (2003), which pose the assumption of a normal and a log–normal distribution of recovery rates.	Found that estimates of recovery rates based on market prices at default are sensitive to changes of the systematic risk factor than are recovery rates determined at restructuring from default. Showed that choice between these 2 methods has a stronger impact on economic capital
Altman et al., 2004	Defaulted bonds' data for the period 1982-2000, which includes the relatively high-default years of 1999 and 2000,	Implemented supply of defaulted bonds as a factor that explain a substantial proportion of the variance in bond recovery rates aggregated across all seniority and collateral levels. In high default years, the supply of defaulted securities tends to exceed demand, thereby driving secondary market prices down. This in turn negatively affects RR estimates, as these are generally measured using bond prices shortly after default.	A negative correlation between default rates and RRs Showed that both the expected loss and the unexpected loss are vastly understated if one assumes that PDs and RRs are uncorrelated
Giese, 2005	American bonds over the period from 1982 to 2001	Proposed non-linear model of dependency of PD and LGD	Positive relationship between annual default frequency and average LGD
Miu, Ozdemir, 2006	Portfolio consisting of 2000 loans made 2000 different borrowers is used. The size of each loan is the same and equal to 1$.	Systematic and specific risk parameters PD and LGD for each borrower were calculated for each scenario of economy. Then it was determined individual probability of default and loss given default for each borrower and the correlation between them. After that they checked whether a borrower is defaulted. This algorithm was repeated 20 000 times. In result the authors obtained the values LGD and PD which correspond to the different scenarios of systematic and idiosyncratic risk. Simulation creates the values of economic capital in monetary units under various scenarios systematic and idiosyncratic sensitivity parameters.	Correlation between systematic PD and LGD was estimated on data of large corporations, therefore it was overestimated (53%). If the average market portfolios are examined where loans are secured by collateral, and that means it less connected with systemic risk, the correlation will be much weaker.
Meng, et al., 2010	1,000 long-term loans. It is assumed that the maturity of each loan is 3 years, PD=1%, the correlation between the value of assets$R_A^2 = 0{,}25$, the correlation between the value of assets and the level of the recovery is $\rho_{A,RR} = 0{,}4$; $R_{RR}^2 = 0{,}34$, average LGD=0,58.	perform numerical experiment on homogeneous portfolio, Regular and stressed LGD have been obtained from the LossCalc model. To estimate the parameters of the correlation $\rho_{A,RR}$ and R_{RR}it was used the data of recovery money of 1,424 default public and private companies for the period from 1981 to 2004. The size of the companies which was determined by the total value of sales was ranged from 0 to 48 billion $.	It has been concluded the following. MTM will be lower with non-zero correlation between the parameters PD and LGD. Order price of asset's purchase decreases to reduce systemic risk return. As a consequence, spread of a credit instrument increases.
Witzany, 2011	Default and recovery data on a retail portfolio obtained from a large Czech bank over the period 2002-2008	proposes a two-factor model to capture retail portfolio probability of default (PD) and loss given default (LGD) parameters : LGD systematic factor is now composed of the PD systematic factor and an additional systematic factor characterizing possibly changing economic conditions during the recovery period, changes in the organization and efficiency of the workout process, etc.	PD x LGD correlation (0.0775) is positive with low significance (probably due to the too short observed time period)

Appendix 1: Review of Literature in details

Research	Data	Method	Result
Vasicek, 1987, 1991, 2002	Actual bank portfolio, 479 loans: maturities 6 months - 6 years; default probabilities from 0.0002 to 0.064; LGD averaged 0.54. The asset returns were generated with 14 common factors.	In theoretical work a closed form Loss Rate Distribution for a Large Homogeneous Portfolio was built. Loss Rate Distribution depends only on two parameters: default probability (p) and assets' correlation (ρ).	Model does not account for systematic correlation between PD & LGD
Frye, 2000	Moody's Default Risk Service database: data on defaults and recoveries from U.S. corporate bond data for the 1982-1997 period	Propose a linear model that accounts for dependency of recoveries on systematic risk: where the same economic conditions are assumed to cause defaults to rise, for example, and RRs to decline. The correlation between these two variables therefore derives from their common dependence on the systematic factor.	• Suggested quantifying estimates of collateral damage on the example of 2 loans • Showed that bond recoveries might decline 20–25 percentage points from their normal-year average. Loan recoveries may decline by a similar amount but from a higher level. • The result is inverse relationship between realized default rates and recovery rates
Hu, Perraudin, 2002	Moody's historical bond market data	standardize the quarterly recovery data in order to filter out the volatility of recovery rates due to changes over time in the pool of rated borrowers find correlations between quarterly recovery rates and default rates for bonds issued by US-domiciled obligors	Corr (quartely RR, default rates) = -0.22 for 1983-2000 Corr (quartely RR, default rates) = -0.19 for 1971-2000
Carey, Gordy, 2003	• Moody's Default Risk Service database of bond • defaults and LGDs • Society of Actuaries database of private placement defaults and • LGDs • Standard & Poor's database of bond defaults and LGDs • Portfolio • Management Data's database of LGDs • *Time period: 1970 - 1999*	examined the aggregate recovery at emergence on the debt of each firm in a sample of bankrupt firms	Their preliminary results: Estimates of simple default rate-LGD correlation are close to zero. Limiting the sample period to 1988-1998, estimated correlations are more in line with main results about negative correlation between PD and RR (0.45 for senior debt and 0.8 for subordinated debt). Concluded that a weak or asymmetric relationship suggests that default rates and LGDs may be influenced by different components of the economic cycle.

© Penikas, Ermolova 2015

Appendix 2: Data

Data from research [Antonova, 2013], defaults 2006-2010 (the source: Cbonds.info)		Additional defaults from 2002 to 2010		New defaults 2012 – 2Q 2015		Total	
Russian bonds	Eurobonds	Russian bonds	Eurobonds	Russian bonds	Eurobonds	Russian bonds	Eurobonds
59 defaults	0 defaults	67 defaults	5 defaults	14 defaults	2 defaults	140 defaults	7 defaults

Days before / after default	The numbers of observations of bond price	The percent of missing values
90 days before default	114 out of 140 defaults	19 %
60 days before default	109 out of 140 defaults	22 %
30 days before default	108 out of 140 defaults	23 %
at moment of default	93 out of 140 defaults	34 %
30 days after default	74 out of 140 defaults	47 %
60 days after default	67 out of 140 defaults	52 %
90 days after default	78 out of 140 defaults	44 %

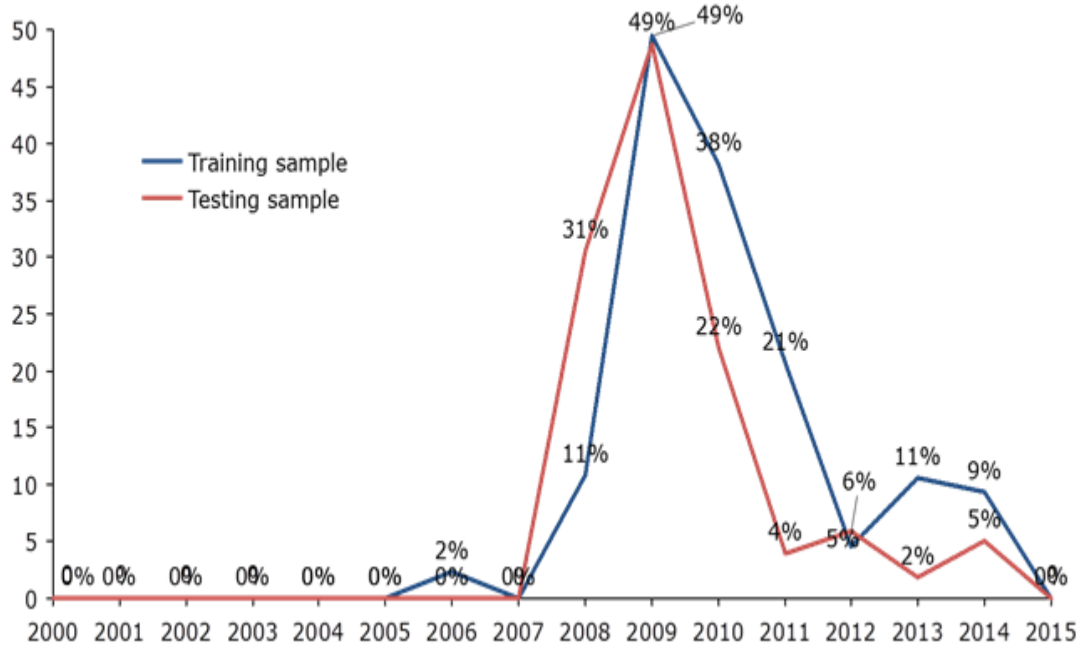
- **It can be noted bond price tends to decrease to default moment. However, for some observations such rule is not executed.**

© Penikas, Ermolova 2015

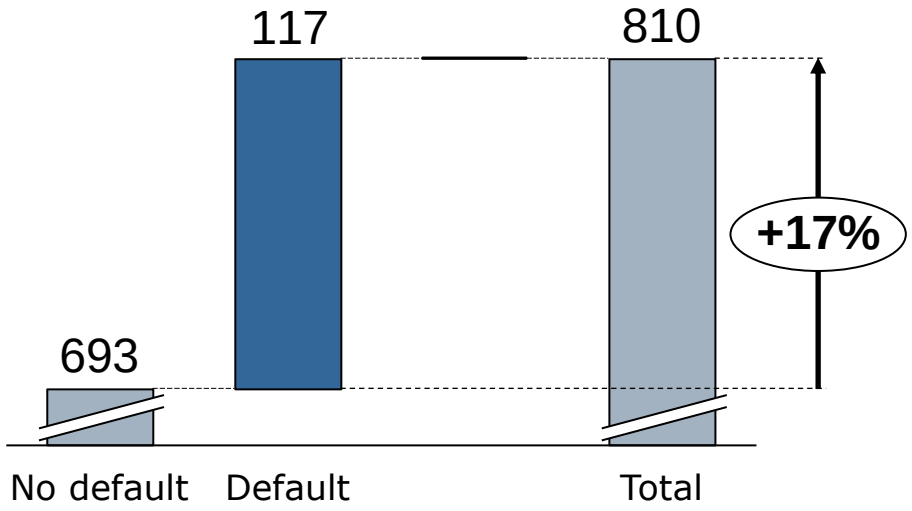
Appendix 3: PD model: sample splitting (out-of-sample)

- Splitting the sample into two groups to build the model and then validate its ability to predict the probability of default.

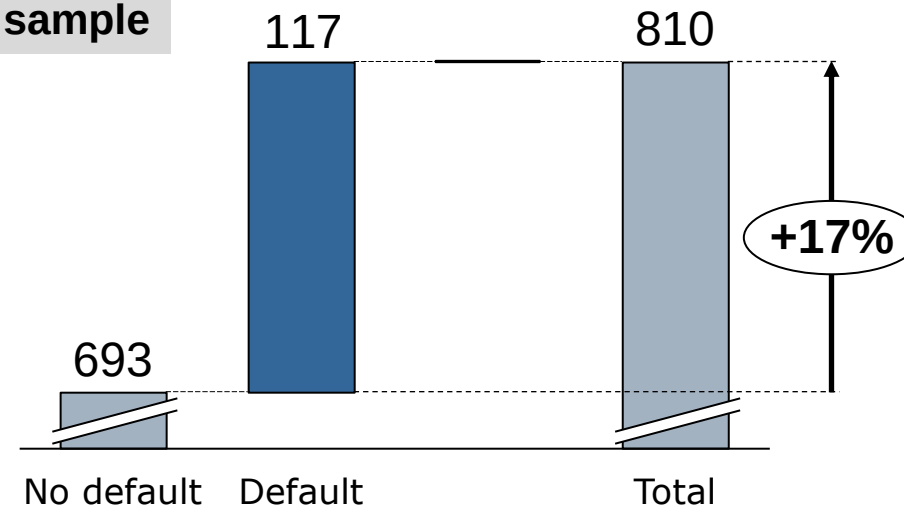
Default rate over the period 2000 – 2014 (%)



Development sample



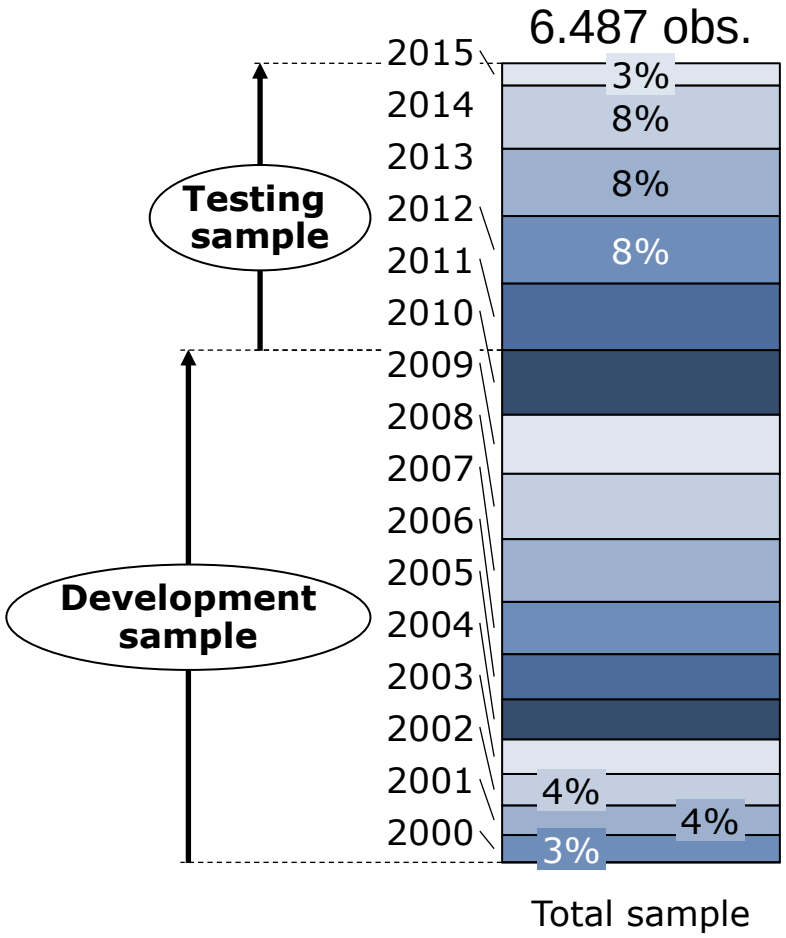
Testing sample

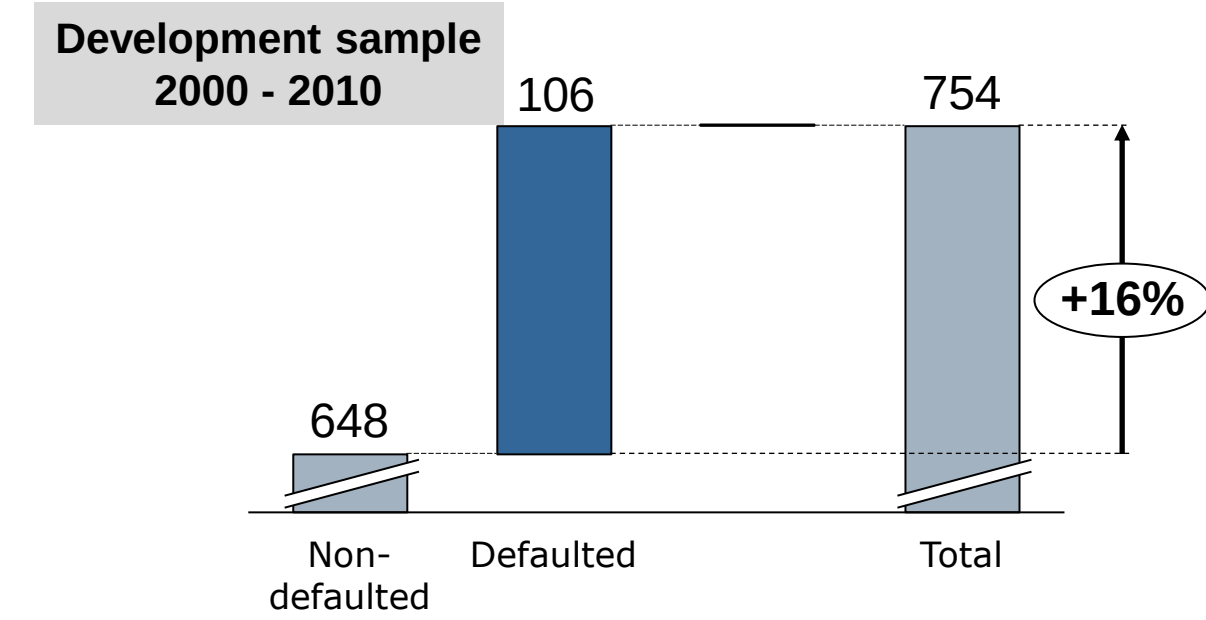


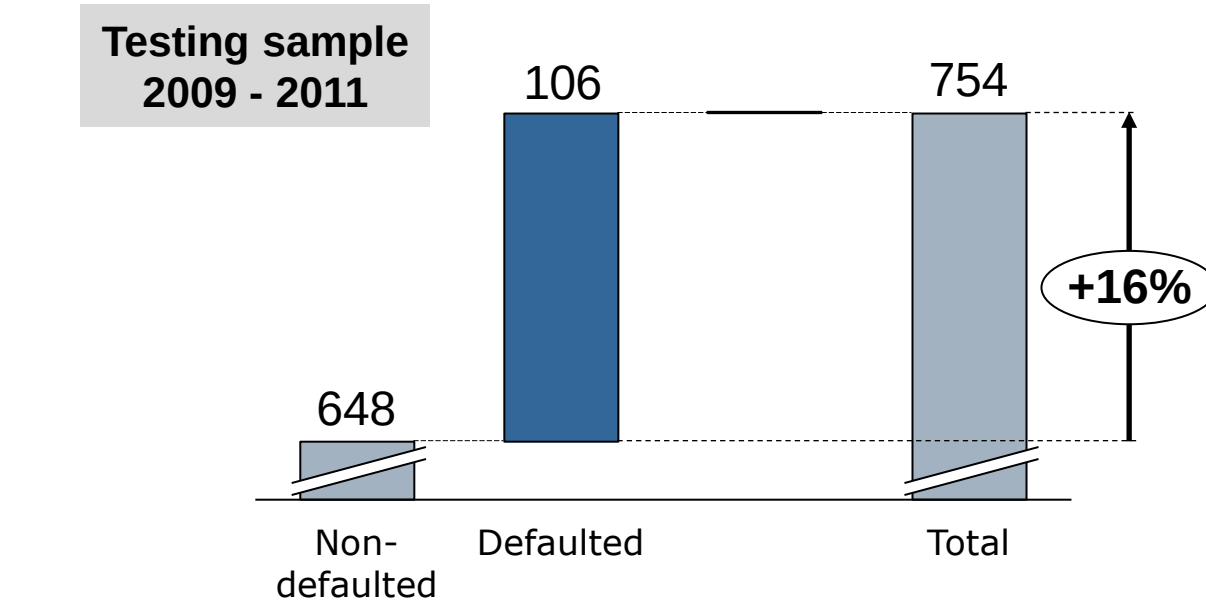
© Penikas, Ermolova 2015

PD model: sample splitting (out-of-time)

- Splitting the sample into two groups to build the model and then validate its ability to predict the probability of default.



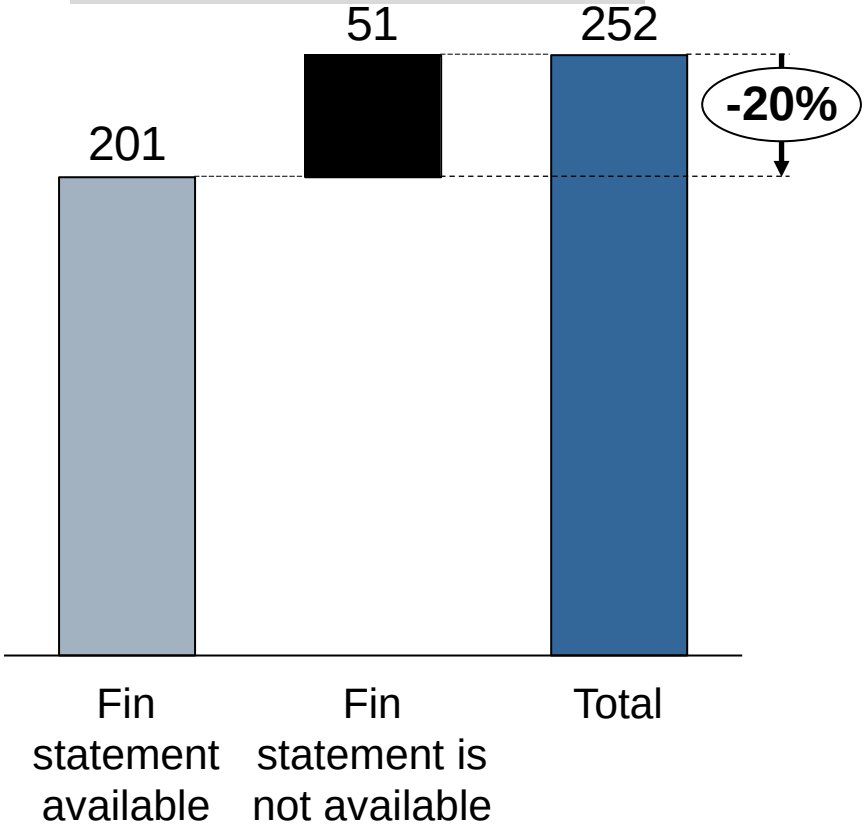




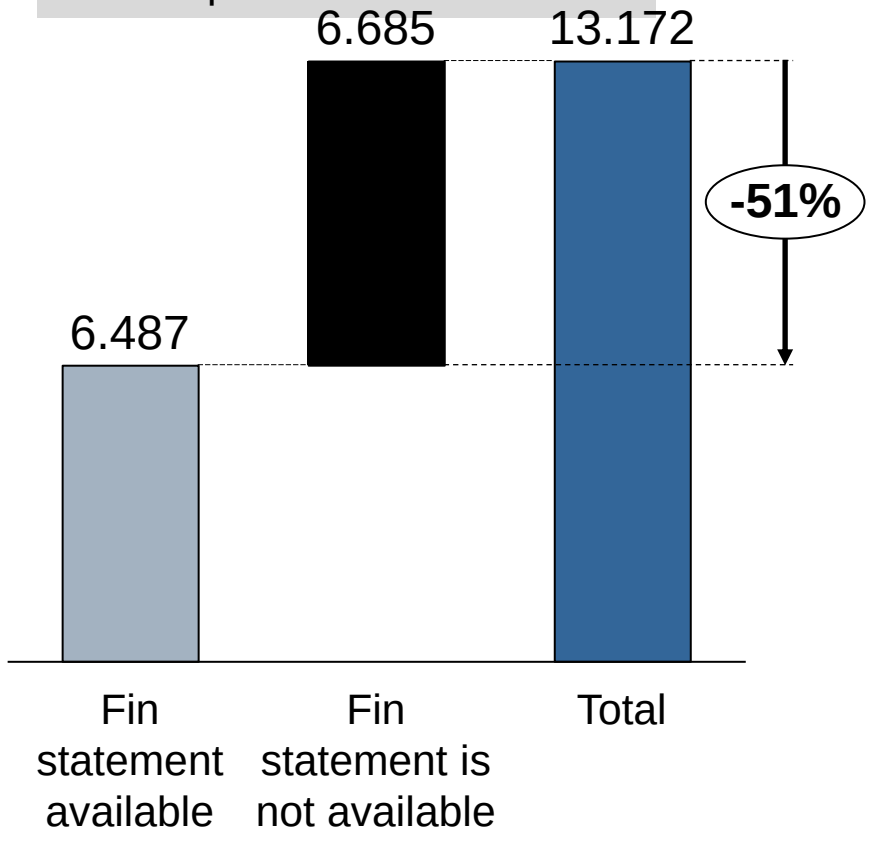
© Penikas, Ermolova 2015

Appendix 3: Data for PD model

 # Defaulted companies
in sample to build PD model



 # Non-Defaulted companies
in sample to build PD model



- Data of the balance sheet, statement of profit and loss and financial ratios taken from SPARK system
- Data on defaults taken from a resource Cbonds.info
- The control line 300 of the balance sheet
- The stability analysis variables by borrowers (number of missing values is not more than 10%)

© Penikas, Ermolova 2015

Appendix 4: LGD model [Antonova, 2013]

- Source: Antonova E. The estimation of the rate of recovery in corporate bonds of Russian issuers, NRU Higher School of Economics, 2013 (http://www.hse.ru/data/2013/06/24/1299828576/Dis_Antonova.pdf)

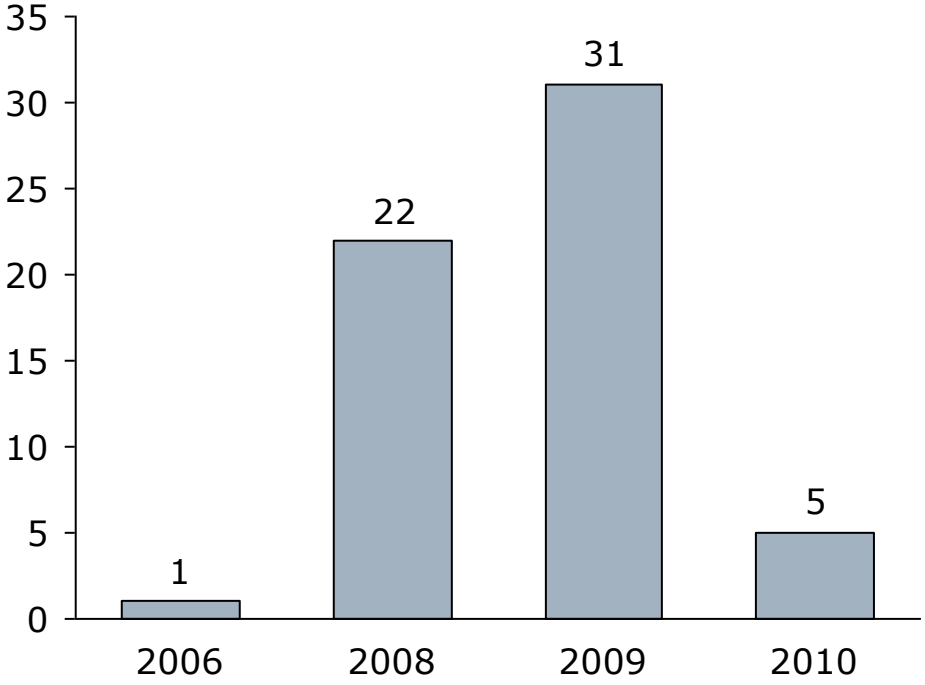
- Model is based on market LGD
  - $RR_d = \frac{B_d}{F}$
    - B_d - market value of the bonds at the time of a default; F - nominal value of the bonds;
    - RR_d - rate recovery after the occurrence of default.
- $RR = 55 + 37 * S - 18 * R - 20 * B - 22 * FR - 13 * L + 33 * C + 27 * D + 25 * G + 29 * F$
  - RR – recovery rate (continuous variable)
  - S – the government support (binary variable)
  - R – debt restructuring (binary variable)
  - B – bankruptcy (binary variable)
  - FR – illegal actions in the corporate governance (binary variable)
  - L – credit from state banks (binary variable);
  - Dummy variables:
    - C – mining operation;
    - D – manufacturing;
    - G – trade;
    - F – construction.

- The model was built on 59 defaulted observations
- The coefficient of determination = 43%
- The statistical significance of the factor L (credit from state banks) = 10%
- The statistical significance of other factors is at 5% significance level
- The absence of multicollinearity, heteroscedasticity and autocorrelation.

© Penikas, Ermolova 2015

Appendix 5: Defaulted bonds with known LGD values

The number of defaulted companies from reseach [Antonova, 2013] with estimates of LGD calculated



Characteristics of data in sample [Antonova, 2013]

- The total number of observations – 59 emissions, where defaults were occurred in
  - Coupon – 21 observations
  - Offer – 22 observations
  - Repayment – 16 observations
- Type of bond – defaulted bonds
- Mean of LGD – 48.8 %
- St. deviation of LGD – 29.2 %
- Mediana of LGD – 50 %
- Years of defaults – 2006 – 2010

Share of outstanding commitments, %	Mean, %	St. deviation, %	Min	Max	Kurtosis	Skewness
Coupon	4.25%	3.81%	0%	45%	51.9	4.97
Offer	50.61%	41.48%	0%	100%	- 1.78	- 0.01
Repayment	49.94%	46.02%	0%	100%	- 1.92	0.08

© Penikas, Ermolova 2015