NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Agata M. Lozinskaia, Evgeniy M. Ozhegov,
Alexander M. Karminsky

DISCONTINUITY IN RELATIVE CREDIT LOSSES: EVIDENCE FROM DEFAULTS ON GOVERNMENT-INSURED RESIDENTIAL MORTGAGES

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: FINANCIAL ECONOMICS
WP BRP 55/FE/2016

Agata M. Lozinskaia[1], Evgeniy M. Ozhegov[2], Alexander M. Karminsky[3]

DISCONTINUITY IN RELATIVE CREDIT LOSSES: EVIDENCE FROM DEFAULTS ON GOVERNMENT-INSURED RESIDENTIAL MORTGAGES[4]

This paper investigates the distribution of relative credit losses given mortgage default for loans provided by a major government-sponsored creditor in a local area. We use borrower's individual and loan-level data on residential mortgages originated in the period 2008–2012. Our numerical analysis indicates that mortgages bunching at certain Loan-to-Value ratios (LTV) led to a discontinuity in relative credit loss given mortgage default. Through regression analysis, we demonstrate discrete jumps in the approximated historical credit losses generated by loans with a high LTV ratios and find thresholds allowing the segmentation of loans according their credit risk. In addition, our results suggest that mortgage insurance is a potentially valuable instrument for compensation for expected loss in certain risk segments.

JEL Classification: C21; G21; G32; R20; R58.

Keywords: discontinuity; credit risk; mortgage default; government mortgage lending programs; loss evaluation.

---

[1] National Research University Higher School of Economics. Department of Economics and Finance. E-mail: AMPoroshina@gmail.com

[2] National Research University Higher School of Economics. Department of Economics and Finance. E-mail: tos600@gmail.com

[3] National Research University Higher School of Economics. Department of Finance. E-mail: karminsky@mail.ru

[4] We thank the Group for Applied Markets and Enterprises Studies for helpful comments. The work on this paper was supported by an individual grant № 14-5352 from the Economics Education and Research Consortium, Inc. (EERC), with funds provided by the Global Development Network. All opinions expressed here are those of the authors and do not necessarily represent those of the Eurasia Foundation, the US Agency for International Development, the World Bank Institution, the Global Development Network and the Government of Sweden.

Introduction

Commercial organizations which offer a residential mortgage lending programs suffer from credit risk, more than from other forms of risk. Mortgage default[5] is traditionally regarded as the worst event of credit risk. This is arguably most relevant to the mortgage default crisis in 2007–2009 which led to large-scale credit losses and related spill-over effects. The efficiency of the credit risk management system is even more crucial for creditors organized in the form of government-sponsored enterprises (GSEs). Credit losses of GSEs arising from mortgage defaults are directly related to government spending, and therefore, attract special attention from policy makers and regulators. In addition, US experience highlights the role of GSEs and alternative mortgage products in the mortgage default crisis.

In this paper, we are mainly focused on quantifying the main credit risk parameter—relative credit loss or loss given default (LGD)—for government-insured mortgage programs in order to provide a better understanding of the specific features and drivers of credit losses for GSEs. These programs are designed by GSEs and are aimed at encouraging residential mortgage lending including the promotion of credit access and homeownership opportunities for minority groups (low-income, low-documented, self-certified households, low initial payment, poor credit history etc.) and to improve the efficiency of mortgage markets (Bhutta, Dokko, Shan, 2010). Importantly, we examine the influence on LGD of the bunching of these mortgages at certain Loan-to-Value ratios (LTV). This is specific for GSE loans and comparable with mortgage bunching at the conforming limit which are purchased by GSEs (DeFusco, Paciorek, 2014) and will be discussed further. The specific features and drivers of credit losses for GSEs play an important role in the loan loss provisioning and the financial stability not only for particular creditors but for the whole national financial system.

We use borrower's individual and loan-level data from GSE branches in a local area in Russia on mortgages originated from 2008–2012 as an example of a developing mortgage market. The increase in mortgage defaults and outstanding debt during 2008–2009 was also observed for the Russian mortgage market which allows us to test whether credit losses are highly volatile. Borrowers of government-insured mortgages in local areas are heterogeneous both in terms of their preferences, personal characteristics including income level, and default behaviour corresponding to their credit risk level. The last factor raises concern policymakers about optimal loan loss provisioning, mortgage pricing, and mortgage program design. For these reasons, our analysis looks at the credit losses for GSE in a local area, which contribute to increasing the predictive power of credit risk management systems and effective credit risk

[5]Usually, mortgage default is delinquency in monthly mortgage payment more than 90 days (90 + days).

segmentation for GSE. Ultimately, this helps to develop appropriate local policy solutions and achieve government policy goals in increasing residential mortgage lending to meet housing needs and may help to avoid mortgage crises in the future.

Our results also provide insights about the presence of specific features in credit loss distribution among the borrower samples of US GSE due to the similarity of the GSE systems in Russia and the US. Specifically, in Russia a GSE or quasi-governmental specialized financial institution—the National Institute for Development of Housing Activity "the Agency for Housing Mortgage Lending" (AHML)—was set up in 1997 to encourage residential mortgage lending[6]. Recent literature (Guriev, 2015; Khmelnitskaya, 2014) mentions the similarities between AHML and Fannie Mae and Freddie Mac—a GSE that aims to develop the mortgage market and uses a two-stage refinancing system. From this point of view, AHML can be regarded as a quasi-analogue to Fannie Mae and Freddie Mac. Khmelnitskaya (2014) shows that establishing AHML and the rules for mortgage securitization were based on US housing experience. According to Mints (2000), AHML was the first in Russia Fannie Mae-type agency. Essentially, AHML is the national regulator of the mortgage market and currently holds over 20% of outstanding residential mortgage debt. Under the two-level system of lending which takes places in the Russian mortgage market (Fig. A1 in Appendix), AHML takes credit losses in the case of mortgage defaults for government-insured mortgages to AHML by its regional branches and the commercial banks operating them. Government-insured mortgages are designed by AHML and consist of two main types: special and non-special. Non-special mortgages do not differ significantly from ones offered by commercial banks. For this market segment AHML acts as price-taker because the terms of credit for such programs are mainly defined by the largest mortgage lender in the country—the state-controlled bank Sberbank. Special mortgages are offered only by AHML and are mainly focused on the above-mentioned minority groups including young researchers, young teachers, and soldiers. Our analysis focuses on loans with different LTV which are traditionally associated with higher credit losses given mortgage defaults[7] and tend to concentrate at certain LTV levels corresponding to different mortgage rates.

This study has the following structure. It starts with a literature review of recent studies of credit loss given mortgage default modelling. The second and the third parts contain the description of the data, and the methodology. Then we discuss the empirical results, robustness checks and conclude with policy implications.

[6] For a full description of AHML system see AHML website (http://rosipoteka.ru/en/).
[7] Harrison (2004) demonstrated also the positive relationship LTV ratios and mortgage default risk.

Literature

There is an extensive body of literature related to the problem of credit risk in residential mortgage lending using probability of default (PD) modelling as its measure. Another no less important measure for credit risk is loss given default (LGD). In the case of mortgage default the share of credit obligations (the value of collateral) returned to the creditor is known as recovery rate (RR), while the remaining part is LGD.

In the mortgage default literature both accounting (Frye et al., 2000; Pennington-Cross, 2003; Leow, Mues, 2012; Zhang, 2013) and economic approaches (Qi, Yang, 2009) to estimate LGD are used. The economic approach (in contrast to the accounting approach) is based on the discounted cash flows method and takes into account the time value of cash flows. Araten et al. (2004) shows the difference in mean values for accounting (27%) and economic LGD (39.8%) based on data from JPMorgan Chase for the 18-year period (1982–1999) for 3761 borrows in default. However to predict the typical difference in the estimates is difficult because it is mainly determined by the characteristics of the bank loan portfolio and macroeconomic conditions, and by the assumptions used in the calculations of these indicators. In addition, economic LGD modelling has other difficulties such as assumptions about the discounted cash flow method including assumptions about the discounting rate and measurement errors (such as the determination of the exact time of working with past-due payments, and the calculation of the process costs). In banking practice, estimates of accounting LGD and the probability of default and exposure at default (EAD) are used to calculate expected credit losses (EL) and risk-weighted assets (RWA), which play an important role in credit risk management and effective capital allocation. For these reasons, in the empirical part of the paper, we employ the accounting approach to estimate relative credit loss.

Despite important policy implications of quantifying credit losses for government-insured residential mortgages, including LGD estimates, the empirical literature is very limited. This is due in large part to data limitations—the lack of publically available borrower's individual and loan-level data and data on historical LGD. The problem of LGD has been widely empirically studied for corporate bonds by using historical bond prices[8] (Jarrow, 2001; Altman et al., 2004; Altman et al., 2005). Gürtler and Hibbeln (2011) mention that for banking loans LGD is typically less due to higher average banking loan seniority and the monitoring quality.

A review of the loss severity literature in residential mortgage lending from the early 1990s to 2015 is provided in Jones and Sirmans (2015). The academic literature mainly pays attention to the linkage between credit loss and the LTV, including the characteristics of property

[8] In this case market approach (market LGD) is used that based on the measurement prices at the moment of default (Altman et al., 2005; Gupton, Stein, 2005) or shortly after it (Dermine, Carvalho de, 2006).

as collateral. The reason is that in the case of mortgage default the outstanding debt can be covered by the selling price of the collateral. Lekkas et al. (1993) show empirically, using data on mortgage loans issued in the period 1975–1990 and refinanced at Freddie Mac, that there is high LGD associated with a high LTV at origination. The positive dependence of LGD on LTV was supported later by several empirical studies on GSEs loans (Calem, LaCour-Little, 2004; Pennington-Cross, 2003), including Qi and Yang (2009) which provides a summary of empirical findings for residential mortgage loss severity. Leow and Mues (2012) propose a two-step LGD model for mortgage loans that includes the probability of repossession with control for LTV and the haircut model. Empirical results were based on residential mortgage default data from a major UK bank and have shown that the procedure used by the authors provided a higher goodness of fit for the observed LGD values. In the case of other types of collateral, the linkage between LGD and collateral value may be nonlinear as shown by Yang and Tkachenko (2012).

Other empirical studies discuss the negative relationship between LGD with credit age and the loan amount (Calem, LaCour-Little, 2004; Pennington-Cross, 2003; Lekkas et al., 1993). More recent literature has found the effect of macroeconomic conditions on LGD mainly through its influence on the collateral value. LGD is higher in a period of housing market downturn compared with normal housing market conditions (Qi, Yang, 2009). Pennington-Cross (2003) found that LGD depended on state foreclosure laws and the type of mortgage (prime or subprime). Lekkas et al. (1993) found a positive effect of a geographical location with a high rate of defaults on LGD and Calem, LaCour-Little (2004) found a nonlinear effect of the relative median income. Gürtler, Hibbeln (2011) found a high dependence of LGD on details of the workout process. They say that write-offs and recoveries are driven by different factors and propose a two-step approach for modelling LGD taking into account those differences.

Relative credit loss distribution in consumer lending does not appear to be normally distributed. Several empirical papers have shown that it is asymmetric, bimodal and U-shaped (Schuermann, 2004; Araten et al., 2004; Dermine, Carvalho de, 2006; Li et al., 2009; Qi, Zhao, 2011; Han, 2013). The empirical evidence on the form of LGD distribution in residential mortgage lending is inconclusive. For example, Loterman et al. (2012) find that LGD distributions for mortgages provided by international banks are not all bimodal and U-shaped. The same evidence was found for the residential mortgage portfolio for the UK bank in Tong et al. (2013).

In the spirit of these studies, we construct the relative credit loss distribution in residential mortgage lending and identify the effect of the LTV and credit terms on its value. Our study differs from previous literature in several ways.

First, we pay attention to LGD at the regional credit portfolio level of GSE, not the national one as in the vast majority of the literature. For that purpose, we use a unique data set of borrower's individual and loan-level information. This focus helps us to estimate explicit credit loss and to contribute to credit risk segmentation for the creditor, and to find practical implications for the decision-making process in local policy about stimulating residential mortgage lending and providing affordable housing. In addition, recent mortgage studies demonstrate the importance of local area analysis. For example, Agarwal et al. (2015) find a default spatial contagion effect between subprime and metropolitan statistical area prime mortgages, and Bradely et al. (2015) discuss externalities from local foreclosures.

Second, we examine the effect on LGD of the bunching loans at certain LTV levels typical for loans purchased by GSE. Specifically, credit interest rates for the analysed government-sponsored mortgages depend on LTV belonging to certain intervals defined by GSE. In contrast, Ghent et at. (2015) do not find evidence for lenders changing loan pricing around discrete eligibility cut-offs defined by GSE. To the best of our knowledge, no research has focused on the LTV bunching effect for GSE credit losses. DeFusco, Paciorek (2014) show that interest rates are typically higher for mortgages above the conforming loan limit[9] and are not purchased by GSEs. This leads to loan amounts bunching at the conforming limit. Their findings indicate a discontinuity in mortgage interest rates related to the conforming loan limit. The estimated effects of the conforming limit on first mortgage demand based on 2.7 million mortgages demonstrate that the average marginal reduction for the bunching borrower is 3.8% of the borrower's loan balance. Discontinuity in GSE intervention in the residential mortgage market across the conforming loan limit is discussed in Kaufman (2014). By using a discontinuity regression design, he finds a modest effect of Fannie Mac and Freddie Mac purchases on loan terms in 2003–2007 and no significant effect on loan default or foreclosure rates.

Third, the empirical analysis of relative credit loss addresses the sample selection bias problem by using Heckman's approach. This point is discussed in the literature on the probability of mortgage default (Ross, 2000; Jacobson, 2003; Lin etc., 2016), but not in the previous empirical studies on the severity of credit loss given mortgage defaults.

Data

The data was provided by the regional branch of AHML on all applications for government-insured mortgage loans collected from 2008 to 2012. In the post-crisis period 2010–2012 the

[9] The nominal and real conforming loan limit for U.S. mortgage market increased over 1997-2007.

total numbers of mortgage applications in the dataset was 15% higher compared with the crisis period 2008–2009 (Fig. 1). The region can be regarded as a representative Russian region. It is an old industrial region with two major agglomerations concentrated on the machinery and chemical industries with the total population 2.7 million. Income level, unemployment rate, the affordability of housing, housing provision and mean property value are near the national mean. The volume of loans issued per capita is 20% higher than the national mean.

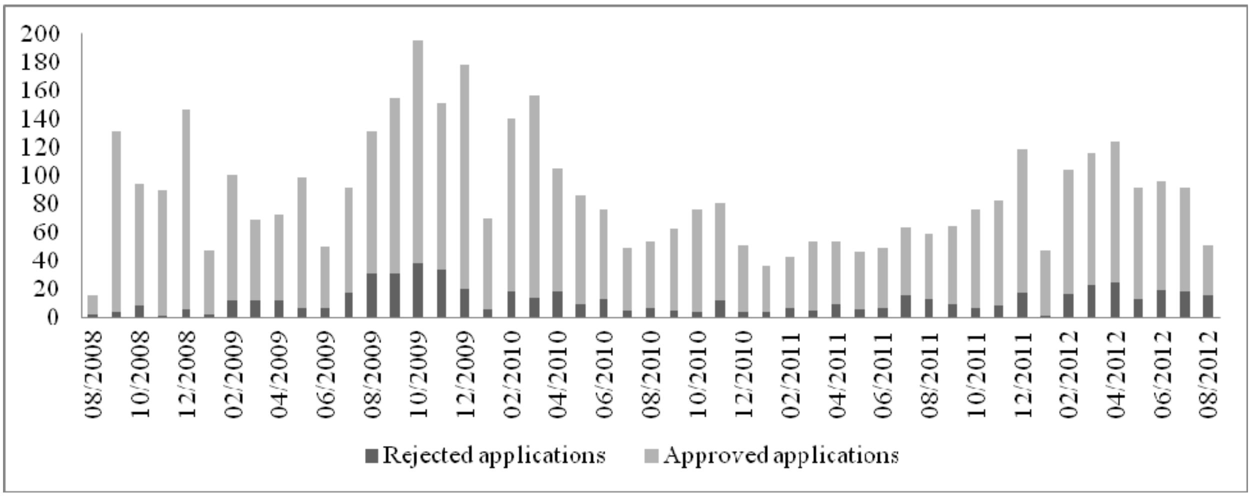

Fig 1. Number of mortgage applications from August 2008 to August 2012

The individual-level dataset contains the socio-demographic characteristics of each of the 4298[10] applicants as main potential borrowers and their co-borrowers on the date of application and flags of approval and contract agreement, but excludes information on credit history. For all 2799 signed contracts, we observe the loan limit set by the creditor, loan contract details including the assessed value of mortgaged property at the origination. The characteristics of the borrower are fully observable and the contract terms are partially observable for only the subsample of applicants who signed a contract. All issued mortgages were refinanced with the AHML regional operator; 60.3% were special mortgage programs.

166 loans defaulted which is equivalent to a 6% default rate. The Russian market for retail lending is usually characterized by a small share of delinquent loans. For an average Russian household, a property purchase is very often the single most significant purchase which can explain the responsible attitude to mortgage repayment and the small default rate in residential mortgage lending. Taking the second mortgage to help pay outstanding debt is not typical for the residential mortgage lending market. In the data set, there is a flag for mortgage

[10] Initially we had a large dataset, but after data cleaning, approximately 12.2% observations were left out as outliers. These are observations with incorrect borrower age, monthly payment, the assessed value of mortgaged property etc.

default, but not the default date, bank actions in past-due payments or historical credit losses given mortgage default.

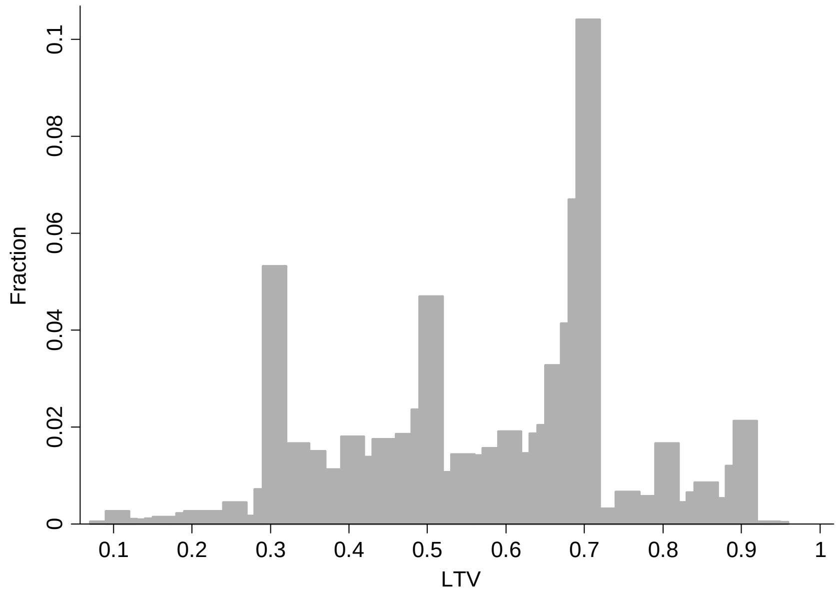

Fig 2. Empirical distribution for LTV ratio

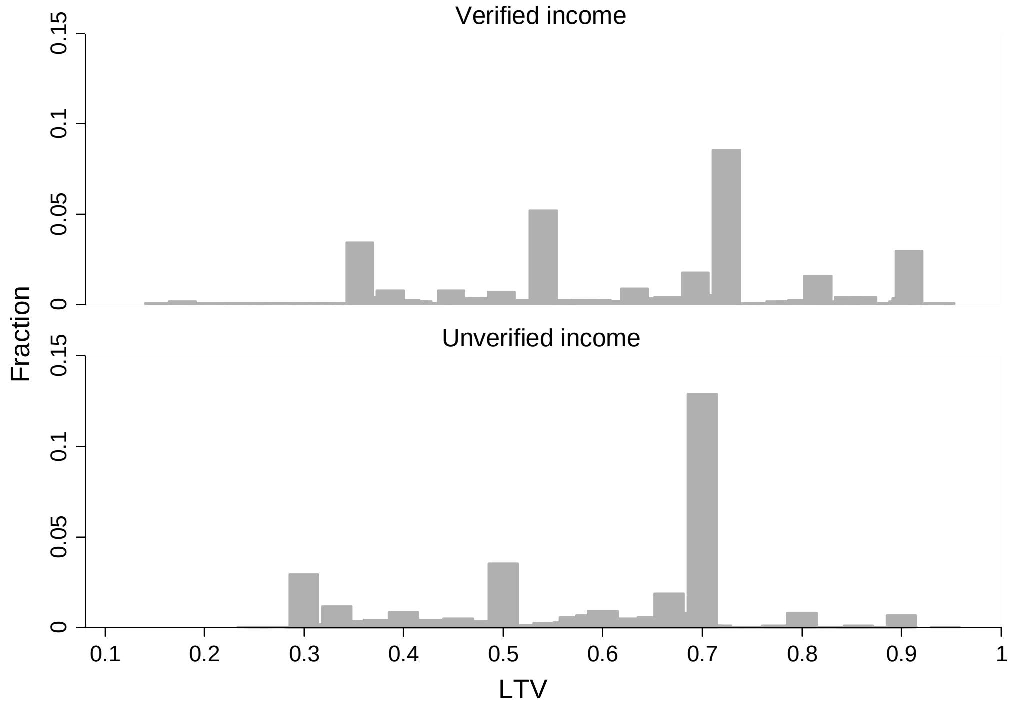

Fig 3. Empirical distribution for LTV ratio for borrowers with verified and unverified income

Some mortgage programs allow the applicants to provide confirmation of their income in “free form” which is known as “self-certified” or “low doc” programs. These programs are usually linked with a higher credit interest rate. For such applicants, income is not recorded in the data set. They are denoted as applicants and borrowers with unverified income and Debt-to-Income ratio, respectively. As shown in Table A1 in Appendix, the share of such applicants and borrowers is relatively high—67.9% and 59%, respectively. The reason for the choice of such mortgage programs may be explained by a temporary or changeable income (LaCour-Little, 2007), for instance, for entrepreneurs. Generally, income should be considered endogenous while modelling the approval of borrower or contract terms. However, we can control for employment category in credit risk estimation, which rejects inconsistency due to the possible endogeneity of income. Moreover, co-borrower income may also be endogenous and we cannot provide any proxy for co-borrower income since we do not have any characteristics for co-borrowers. This is a limitation of the research, but we may consider it as insignificant for the choice of contract terms compared to the income of the main borrower.

Government-insured mortgages are traditionally nominated in the national currency (Russian rubles). In the sample the largest share of residential mortgages consists of loans with fixed credit interest rate (86.5%), maturity 15–19.9 years (39.5%), LTV 0.5-0.7 (54.7%) which are mainly related to non-special mortgage programs (48.4%) and originated in the post-crisis period 2010–2012 years (57.6%) (see Table A1 in Appendix). Fig. 2 shows that the distribution for the LTV is asymmetric and hump-shaped with peaks at values 0.3, 0.5, and 0.7. This fact is mainly explained by the design of government-sponsored mortgages. The credit interest rate depends on the LTV, potential borrower characteristics, and contract terms while the credit interest rate is fixed when the LTV belonging to intervals up to 0.3, from 0.3 to 0.5, from 0.5 to 0.7, and over 0.7. These intervals are defined by GSE and remain the same for borrowers with verified and unverified income (see Fig. 3). Under these conditions despite the small difference in LTV, for example, 0.69 and 0.71, credit interest rates along with other contract terms differ more and shift more of a financial burden onto borrowers. Customers tend to bunch just below thresholds to avoid a higher interest rate. This is similar to Kaufman’s (2014) findings for bunching at the conforming loan limit. Intuitively, it may lead to a discontinuity in relative credit losses given mortgage default over certain LTV intervals.

The relationship between LTV and the cumulative default rates tends to be non-monotonic as is shown in Table 1, and in contrast to Order and Zorn (2000). To control for this effect we use a set of dummies for different LTVs in the probability of default model. The cumulative default rate is higher by 11.65% for borrowers with verified income and it is approximately twice as high as the default rate at the portfolio level. This result can be explained by the fact that the borrower’s income reported in the mortgage application and as a consequence in the dataset is related to the borrower’s official potential income rather than the real income. The real income in contrast to official one can be sufficient to cover mortgage payment which explains why borrowers with unverified income are not more likely to default. With the same LTV borrowers with verified income have a higher chance of default ranging from 7.37% to 45.45% for different LTV groups.

Tab. 1. Cumulative default rates (%) by verified/unverified borrower's income and LTV ratio

Borrower's income	LTV										
	0-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1	All
Verified	-	45.45	40	7.37	10.55	7.07	9.67	26.53	19.46	-	12.8
Unverified	-	-	6.67	1.24	2.33	0.44	0.55	2.70	1.69	-	1.15
All	-	45.45	30	3.94	5.68	2.46	3.58	16.28	14.42	-	

We use such macroeconomic factors at the regional level as the regional quarterly unemployment rate, refinancing rate, the mean regional property value which is publically available at Bank of Russia, Federal Bureau of Statistics Russian Federation and AHML websites and is reported in Tables A1 in Appendix.

Methodology

We start our analysis with the estimation of the probability of the mortgage default model including credit duration as an explanatory variable, and controlling for borrower standard socio-demographic characteristics, loan parameters and macroeconomic conditions at the regional level (unemployment rate, mean price value) at the origination, which demonstrates high discriminative power and correlates with the dependent variable. We take into account the methodological recommendations of Central Bank of Russia (CBR) for developing the Internal-Ratings-Based Approach (Bank of Russia, 2012), MILAN analysis—Moody's Individual Loan Analysis of Residential Mortgage-Backed Securities (Moody's, 2009), and the results of the monitoring of financial behaviour (Radaev, Kuzina, 2011). Models with and without correction for sample selection bias are estimated by simultaneously modelling the probability of endorsement and the probability of mortgage default.

Next, we approximate the historical accounting LGD—the Expected Loss Given Default (ELGD)—because the data set excluded information on the realized credit loss for mortgage loans which default. In this approach, we use the estimated probabilities of default over the analysed period of time to approximate LGD for defaulted loans due to the lack of information on the default date. We assume the sale of mortgaged properties due to foreclosure[11] at discounts relative to other properties on the market. The steps of this applied approach to estimate ELGD are:

[11] Lambrecht (2003) discussed the decision rules in possessions process and Cordell (2015) costs of foreclosure delay.

1. Generating the ordered variable t based on the explanatory variable for PD—credit duration (the difference between August 2012 and the mortgage loan issue date). The variable t represents the number of months from the mortgage loan issue date to the date of ELGD calculation (August 2012) starting from 4th month (t=4, 5, 6 etc.) (see Table 2).

2. Predicting $\hat{PD}_{it}$ for the i defaulted mortgage loan time at each time point t based on a probit PD model with correction for sample selection bias. The predicted PD based on the probit model represents cumulative PD, but we are interested in non-cumulative PD represented in column 4 in Table 2. For example, non-cumulative PD at t=5 is calculated as the difference in cumulative PD estimates at t=5 and t=4.

Tab. 2. Predicting *PD* for mortgage defaulted loans at each time point

# of defaulted loan	t	Cumulative PD based on probit-model PD	$\hat{PD}_{it}$	LGD
1	4	0.3	0.3	0.4
1	5	0.4	$0.4-0.3=0.1$	0.5
1	…	…	…	…
2	4	0.5	0.5	0.2
2	5	0.6	$0.6-0.5=0.1$	0.3
2	…	…	…	…
…	…	…	…	…
165	4	0.2	0,2	0.4
165	5	0.3	$0.3-0.2=0.1$	0.1
165	…	….	…	…

3. Approximating the market collateral value R (with a haircut), workout process costs C, EAD as current outstanding residual loan amount and accounting LGD for the i defaulted mortgage loan at each time point t. The accounting LGD is measured as:

$$LGD_{it} = \frac{EAD_{it} - R_{it} + C_{it}}{EAD_{it}} = 1 - \frac{R_{it} - C_{it}}{EAD_{it}}$$

4. Approximating ELGD for the i defaulted mortgage loan as follows:

$$ELGD_i = E(LGD_i) = \sum_{t=1}^{M} \hat{PD}_{it} \cdot LGD_{it}, \quad t = 1, \ldots, M,$$

where M is the number of months starting from 4th from the mortgage loan issue date and August 2012. In ELGD for i defaulted mortgage loan PD weighting for this loan at each time point t is used.

To obtain an empirical distribution of the relative loss rate at the credit portfolio level several assumptions are made which are mainly determined by the empirical data. First, we assume that time is discretely measured in months, the frequency mortgage payments. Second, due to the lack of information about past-due payment regulation by the creditor, we assume that creditor uses legal collection in the case of mortgage default. Third, the ratio of the assessed property value (per sq. meter) at the mortgage loan issue date to the regional market property value (per sq. meter) at the end of the workout process *T*, which is assumed constant over the time, and the total area of the flat are used in the approximation of R . The haircut is 20%, which is determined by the current legislation. The workout process *T* takes 5 months from the approximated default date based on an expert estimate of AHML. Fourth, AHML estimates workout process costs at 5–15% of the current collateral value corresponding to the share of legal costs in the accounting records of the AHML regional operator. For this reason, we regard four possible scenarios for workout process costs *C*: 0, 5%, 10%, and 15%.

EAD (Exposure at Default) is calculated as:

$$EAD = A \times (M - t_1 + 3) + P,$$

where A is the annuity payment, M is the maturity, t_1 is the number of the month from the default date to the end of the loan term, P is fees and penalties. We assume that P=0 because in the legal process the borrower can achieve a cancellation of fees and penalties.

In addition, we assume that related borrowers are absent as the correlation between the defaults of related mortgage borrowers is absent. Repeated[12] mortgage defaults are absent. The approximated accounting LGD is censored on the interval [0;1][13] as in (Li et al., 2009).

Finally, we estimate an OLS linear regression model for the approximated historical ELGD to investigate the effect of socio-demographic characteristics and credit terms including LTV. To examine the effect of the bunching of loans at certain LTV values on relative loss we include a set of LTV dummies defined by the empirical data. We include also product terms LTV and LTV dummies to test whether the relationship between relative credit loss and LTV is not constant across different LTV intervals. In addition, to correct for sample selection bias we follow Heckman's approach.

[12] This problem is discussed in the paper (Dermine, Carvalho de, 2006).
[13] Approximately 2% of defaulted loans have LGD out of the unit interval.

Empirical Results

Results of the probit models of the probability of default[14] do not signal a substantial bias in estimated coefficients with and without correction for sample selection. According to the above-mentioned approach to approximate historical losses for 165[15] defaulted loans we use the estimated probabilities of default from the specification with credit duration as an explanatory variable, which demonstrates similar predictive power to other specifications.

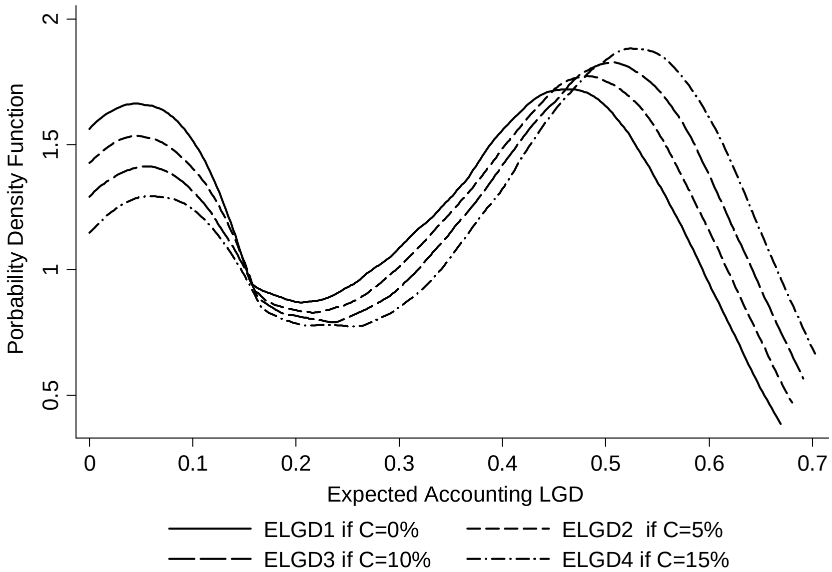

Fig 4. Empirical distribution of expected loss given default (*ELGD*)

Fig. 4 shows ELGD. For the government-insured defaulted mortgages, it has hump-shaped distribution with peaks close to 0.1 and 0.5 and with a standard deviation of 0.2. These results are robust under different scenarios for workout process costs and match Araten et al. (2004) in that LGD has a bimodal distribution. This result is explained by the loss heterogeneity due to different LTV. The first and the second modes are related to low and high LTV, respectively. The first two charts in Fig. 5 show that with LTV up to 0.5 the hump in relative loss is close to the first mode and over 0.5, it is close to the second mode. This is also found in previous empirical studies (Lekkas et al., 1993; Pennington-Cross, 2003; Araten et al., 2004; Calem, LaCour-Little, 2004; Qi, Yang, 2009).

[14] Results are available upon request.
[15] 1 observation without information on the net floor area of the property was excluded.

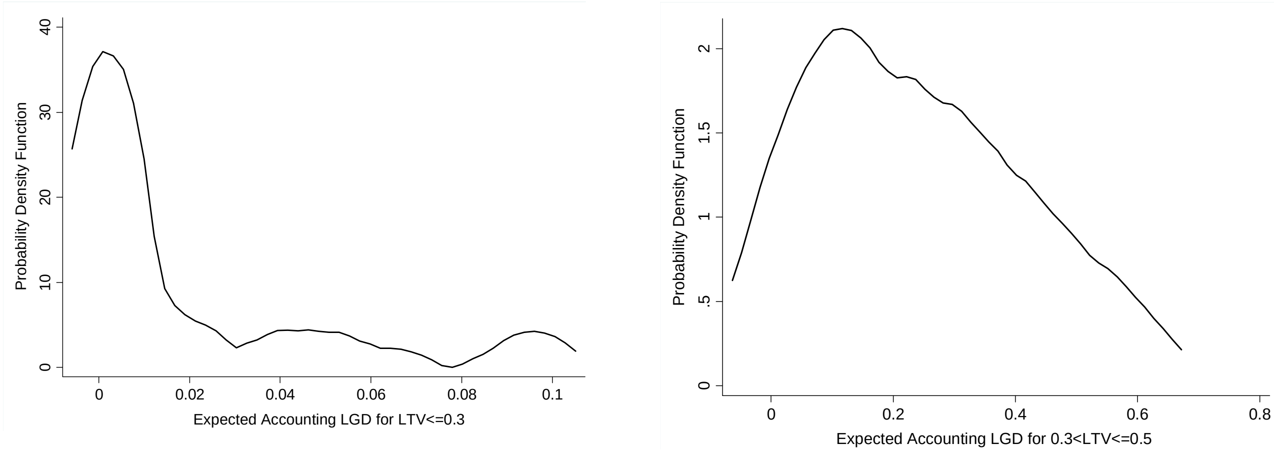

Fig 5. Empirical distribution of expected loss given default (ELGD) under *C*=15% for different LTV ratios at origination

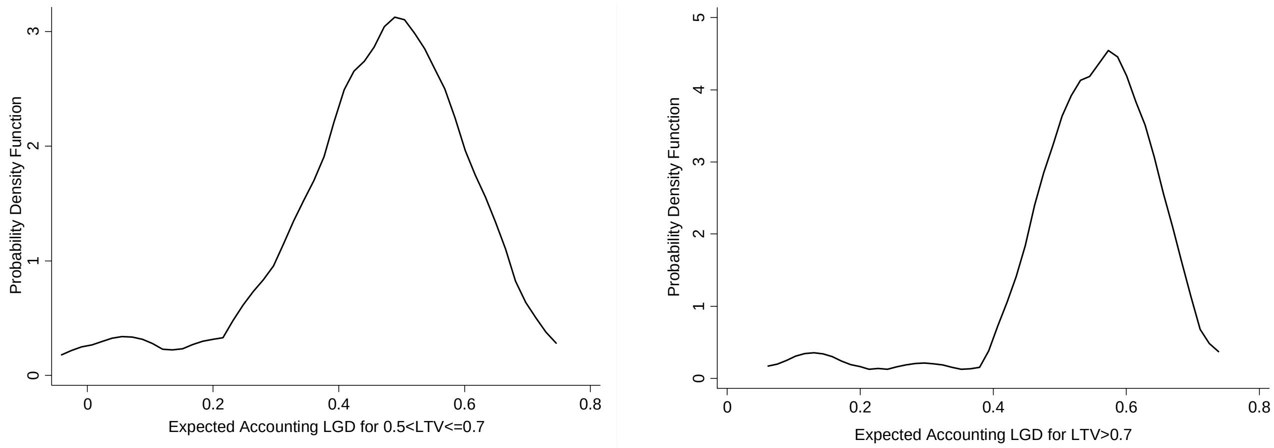

Fig 5. Empirical distribution of expected loss given default (ELGD) under *C*=15% for different LTV ratios at origination (cont.)

These results correspond to the regression analysis results in Table 3. They demonstrate that LGD under different scenarios for workout process costs is positively related to LTV. This result is robust to specifications with and without correction for sample selection bias. Despite the linear monotonic relationship between relative credit loss and LTV at origination, we found a statistically significant impact for dummies on LTV from 0.3 to 0.5 and above 0.7. This finding provides empirical evidence for a LGD discontinuity that depends on LTV. The statistical significance of the LTV dummies interaction effects with the LTV signal, which slopes in the regression for relative credit loss for certain LTV intervals are different. For example, to the left of the threshold 0.7, LGD increase rapidly approaching LTV=0.7. Beyond the threshold, they rise gradually. Significant discrete jumps in LGD for contracts with LTV above certain thresholds are shown in Fig. 6. This finding is robust under different workout process costs shown in Fig. 7.

Tab. 3. Estimated parameters for loss given mortgage default

	With correction for sample selection bias				*Without correction for sample selection bias*			*Bias*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	
Variable	ELGD1 (*C* =0)	ELGD2 (*C* =5%)	ELGD3 (*C*=10%)	ELGD4 (*C*=15%)	ELGD4 (*C*=15%)	ELGD4 (*C*=15%)	ELGD3 (*C*=10%)	
LTV	0.906***	0.877***	0.828***	0.765***	0.770***	0.680***	0.738***	0.005
	(0.172)	(0.170)	(0.164)	(0.156)	(0.158)	(0.104)	(0.111)	(0.314)
LTV [0;0.3]	0.143	0.084	0.010	-0.076	-0.073	-0.129	-0.048	0.003
	(0.131)	(0.132)	(0.129)	(0.125)	(0.126)	(0.100)	(0.103)	(0.251)
LTV (0.3;0.5]	-0.051	-0.112	-0.176	-0.238*	-0.234*	-0.290***	-0.236**	0.004
	(0.138)	(0.136)	(0.132)	(0.127)	(0.129)	(0.098)	(0.102)	(0.256)
LTV (0.7;1]	0.302**	0.276*	0.237*	0.194	0.193	0.036	0.077	-0.001
	(0.148)	(0.141)	(0.131)	(0.122)	(0.124)	(0.107)	(0.112)	(0.246)
LTV [0;0.3]×LTV	-0.474	-0.372	-0.239	-0.074	-0.063	0.054	-0.101	0.011
	(0.330)	(0.333)	(0.331)	(0.325)	(0.324)	(0.323)	(0.329)	(0.649)
LTV (0.3;0.5]×LTV	0.108	0.232	0.358	0.476**	0.478**	0.570***	0.470**	0.002
	(0.242)	(0.239)	(0.231)	(0.223)	(0.224)	(0.186)	(0.191)	(0.447)
LTV (0.7;1]×LTV	-0.442**	-0.408**	-0.356*	-0.296*	-0.294*	-0.086	-0.143	0.002
	(0.207)	(0.197)	(0.185)	(0.172)	(0.175)	(0.142)	(0.149)	(0.347)
Constant	-0.317	-0.494	-0.674*	-0.790**	-1.092***	-0.888***	-0.920***	-0.302
	(0.400)	(0.379)	(0.360)	(0.356)	(0.141)	(0.081)	(0.086)	(0.497)
Number of parameters	34	34	34	34	33	15	15	
N	166	166	166	166	166	166	166	
Adjusted R^2	0.947	0.952	0.956	0.958	0.958	0.951	0.949	

Note: 1. Regressions (1)-(4) include Heckman's lambda.

2. Only statistically significant main contract variables reported. All regressions include dummies for maturity, rate, credit duration, DTI, regional property value. In regressions (1)-(5) we also control for borrower's socio-demographic characteristics such as the age of the main borrower and its square, education level, activity category. In regressions (6)-(7) we include only dummies for maturity, rate, credit duration, and regional property value.

3. Robust standard errors in the parenthesis.

4. Bias is calculated as the difference in the estimated parameters for ELGD4 without (5) and with sample selection correction (4).

5. Base categories: LTV – (0.5; 0.7], DTI – [0.2-0.4), maturity≥25 years, incomplete higher education, married, unemployed.

6. *, ** and *** indicate significance at 10%, 5% and 1% levels, respectively.

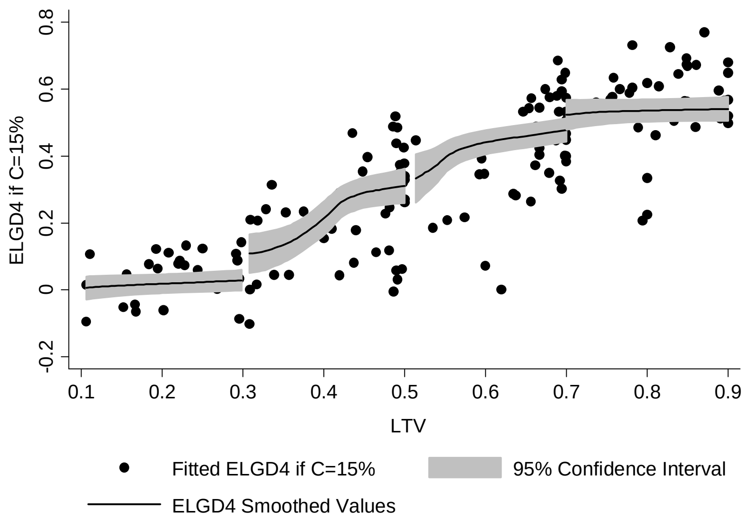

Fig 6. Relationship between predicted relative credit loss given default under *C*=15% and LTV at origination

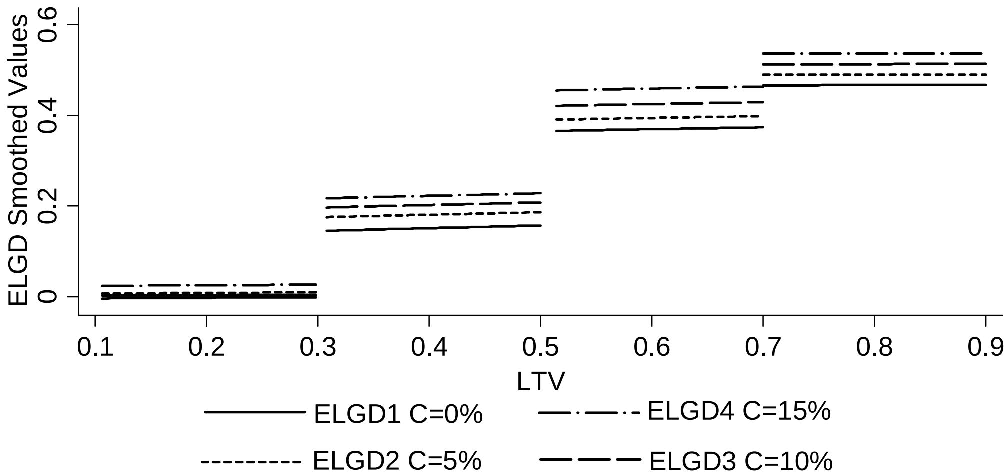

Fig 7. Relationship between predicted relative credit loss given default for different workout process costs and LTV at origination

Fig. 6 and Fig. 7 show that for LTV up to 0.3 LGD are close to zero. The almost zero slope in the LGD regression for this LTV interval indicates that an increase in LTV leads to a negligible increase in LGD. For LTV over 0.7, we also observe a moderate increase in LGD (on average between 0.4 to 0.6) when LTV increases. Currently, AHML uses liability insurance for borrowers with mortgage loans having LTV over 0.7. It means that LGDs can be partly compensated for by the insurance company. The credit loss is more sensitive to an increase in LTV in the middle intervals. The slopes for LTV from 0.3 to 0.5 and from 0.5 and to 0.7 have higher positive slopes in ELGD compared to other LTV intervals. For such mortgage loans LGD is compensated through an increase in initial credit interest rate. We also find a statistically significant impact on LGD credit terms which leads to an increase in the model's explanatory power which corresponds to previous literature (Qi, Yang, 2009).

Robustness Checks

First, due to the lack of historical information on workout process costs we check how our LGD analysis depends on the workout process cost value ranging from 0% to 15%. The results remain the same for empirical distributions for the expected relative credit losses given mortgage default (Fig. 3) and OLS results for LGD (Tab. 3, Fig. 7).

Secondly, we replicate our kernel density function estimates for LGD using a kernel bandwidth that differs from the optimal half-width. For example, for empirical distribution for LGD in Fig. 3 an optimal half-width is 0.07. The discussion and formula for the optimal window width can be found in Salgado-Ugarte et al. (1993). Under the change of bandwidth from 0.02 to 0.09, the shape of the distribution is similar to those reported in Fig. 3.

Finally, we estimate specifications for LGD with and without correction for sample selection bias by using Heckman's approach. Since we approximate losses only for borrowers who have defaulted, it is necessary to check whether the sample of defaulted borrowers is biased in terms of expected losses. The results in Table 3 demonstrate that the estimated parameters and statistical significance for main contract term parameters (LTV, DTI, maturity and credit interest rate) are not remarkably different. In addition, the selection correction term (Heckman's lambda) and the absolute bias in the estimated parameters for explanatory variables are not statistically significant. This result indicates that there is no significant correlation between error terms in equations of PD and LGD conditional on the explanatory variables. Our finding is similar to that discussed by Ross (2000) in that sample selection bias disappears when all the significant variables to explain both selection and outcome equations in the sample selection model are included.

Conclusion

This paper examines LGD for a regional credit portfolio of government-insured mortgage loans originating from 2008–2012 in Russia. Given the lack of information on historical LGD, we employ an approach to approximate it. Our results confirm that LGD has a bimodal distribution that is consistent with the previous literature. The relative loss distribution contains spikes around 0.1 and 0.5. The nature of the modes is mainly determined by the difference in LTV for defaulted mortgage loans.

Relative LGD on government-insured loans increases monotonically with an increase in LTV. Closer analysis of credit pools shows that loans tend to bunch at certain LTV. We directly compare relative credit loss for loans above and below certain thresholds by using regression analysis. We find that relative credit loss corresponds to LTV in statistically significant discrete jumps. This finding is supported by regression results showing different slopes in relative credit loss regression beyond the LTV thresholds 0.3, 0.5, 0.7. Creditors should pay special attention to these thresholds which will allow credit risk segmentation from the point of potential LGD and optimal loan loss provisioning. AHML are currently using mortgage insurance to cover credit losses for contracts with LTV above threshold 0.7.

To conclude, we find that a relative LGD analysis at the regional level allows the segmentation of contracts according to their credit risk which may contribute to a socially and economically efficient decision-making process by local policymakers on a design mortgage programs and have applications for mortgage pricing, state legislation, taxation, and regulation.

References

Agarwal, S., B.W. Ambrose and Y. Yildirim. 2015. The Subprime Virus. *Real Estate Economics* 43(4): 891–915.

Altman, E.I., B. Brady, A. Resti and A. Sironi. 2005. The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business-Chicago* 78(6): 2203–2228.

———, A. Resti and A. Sironi. 2004. Default recovery rates in credit risk modelling: a review of the literature and empirical evidence. *Economic Notes* 33(2): 183–208.

Araten, M., M. Jacobs and P. Varshney. 2014. Measuring LGD on commercial loans: an 18-year internal study. *RMA Journal* 86(8): 96–103.

Bank of Russia. 2012. Pis'mo Banka Rossii ot 29.12.2012 N 192-T «O Metodicheskikh Rekomendatsiyakh po Realizatsii Podkhoda k Raschetu Kreditnogo Riska na Osnove Vnutrennikh Reytingov Bankov» [Letter of Bank of Russia from 29.12.2012 N 192-T «The guidance on implementation of the approach to the calculation of credit risk based on internal ratings of banks»].

Bhutta, N., J.K. Dokko and H. Shan. 2010. The Depth of Negative Equity and Mortgage Default Decisions. Finance and Economics Discussion Series, No. 2010-35. Washington, DC: Board of Governors of the Federal Reserve.

Bradley, M.G., A.C. Cutts and W. Liu. 2015. Strategic Mortgage Default: The Effect of Neighborhood Factors. *Real Estate Economics* 43(2): 271–299.

Calem P.S. and M. LaCour-Little. 2004. Risk-based capital requirements for mortgage loans. *Journal of Banking and Finance* 28(3): 647–672.

DeFusco, A.A. and A. Paciorek. 2014. The interest rate elasticity of mortgage demand: Evidence from bunching at the conforming loan limit. Finance and Economics Discussion Series, No. 2014-11. Washington, DC: Federal Reserve Board.

Dermine J. and C.N. de Carvalho. 2006. Bank loan losses-given-default: A case study. *Journal of Banking and Finance* 30(4): 1219–1243.

Frye J., L. Ashley, R. Bliss, R. Cahill, P. Calem, M. Foss and J. Nelson. 2000. Collateral damage: A source of systematic credit risk. *Risk* 13(4): 91–94.

Cordell, L., L. Geng, L.S. Goodman and L. Yang. 2015. The Cost of Foreclosure Delay. *Real Estate Economics* 43(4): 916–956.

Ghent, A.C., R. Hernández-Murillo and M.T. Owyang. 2015. Did affordable housing legislation contribute to the subprime securities boom? *Real Estate Economics* 43(4): 820–854.

Gupton G.M. and R.M. Stein. 2005. LossCalc V2: Dynamic Prediction of LGD Modeling Methodology. Moody's KMV Working Paper.

Guriev, S. 2015. Mify ekonomiki. Zabluzhdeniya i stereotipy, kotoryye rasprostranyayut SMI i politiki [Myths of the economy. Misconceptions and stereotypes that are distributed by mass-media and politicians]. Litres.

Gürtler M. and M. Hibbeln. 2011. Pitfalls in modeling loss given default of bank loans. Technische Universität Braunschweig Working Paper.

Han, C. and Y. Jang. 2013. Effects of debt collection practices on loss given default. *Journal of Banking and Finance* 37(1): 21–31.

Harrison, D.M., T.G. Noordewier and A. Yavas. 2004. Do riskier borrowers borrow more? *Real Estate Economics* 32(3): 385–411.

Jacobson, T. and K. Roszbach. 2003. Bank lending policy, credit scoring and value-at-risk. *Journal of Banking and Finance* 27(4): 615–633.

Jarrow R. 2001. Default parameter estimation using market prices. *Financial Analysts Journal* 57(5): 75–92.

Jones, T. and G.S. Sirmans. 2015. The Underlying Determinants of Residential Mortgage Default. *Journal of Real Estate Literature* 23(2): 167–205.

Kaufman, A. 2014. The influence of Fannie and Freddie on mortgage loan terms. *Real Estate Economics* 42(2): 472–496.
Khmelnitskaya, M. 2014. Russian housing finance policy: state-led institutional evolution. *Post-Communist Economies* 26(2): 149–175.
LaCour-Little, M. 2007. The Home Purchase Mortgage Preferences of Low- and Moderate Income Households. *Real Estate Economics* 35: 265–290.
Lambrecht, B., W.R.M. Perraudin and S. Satchell. 2003. Mortgage default and possession under recourse: A competing hazards approach. *Journal of Money, Credit, and Banking* 35(3): 425–442.
Lekkas V., J.M. Quigley and R. Order. 1993. Loan Loss Severity and Optimal Mortgage Default. *Real Estate Economics* 21(4): 353–371.
Leow M. and C. Mues. 2012. Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. *International Journal of Forecasting* 28(1): 183–195.
Li D., R. Bhariok, S. Keenan and S. Santilli. 2009. Validation techniques and performance metrics for loss given default models. *Journal of Risk Model Validation* 3(3): 3–26.
Lin, Z., Y. Liu and J. Xie. 2016. Immigrants and Mortgage Delinquency. *Real Estate Economics* 44(1):198–235.
Loterman, G., I. Brown, D. Martens, C. Mues and B. Baesens. 2012. Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting* 28(1): 161–170.
Mints, V. 2000. Selecting a Housing Finance System for Russia. *Housing Finance International* 15(2): 49–57.
Moody's MILAN Methodology for Rating Russian RMBS. 2008. Moody's.
Order R.V. and P. Zorn. 2000. Income, location and default: some implications for community lending. *Real Estate Economics* 28(3): 385–404.
Pennington-Cross, A. 2003. Subprime and prime mortgages: Loss distributions. Working Papers WP 03-1. Washington, DC: Office of Federal Housing Enterprises Oversight (OFHEO).
Qi M. and X. Yang. 2009. Loss given default of high loan-to-value residential mortgages. *Journal of Banking and Finance* 33(5): 788–799.
——— and X. Zhao. 2011. Comparison of modeling methods for Loss Given Default. *Journal of Banking and Finance* 35(1): 2842–2855.
Radaev, V.V. and O.E. Kuzina. 2011. Monitoring of the financial behavior of the population. Available at http://www.hse.ru/org/projects/47265530.
Ross, S.L. 2000. Mortgage lending, sample selection and default. *Real Estate Economics* 28(4): 581–621.
Salgado-Ugarte, I.H., M. Shimizu and T. Taniuchi. 1993. Exploring the shape of univariate data using kernel density estimators. *Stata Technical Bulletin* 3(16): 8–19.
Schuermann T. 2004. What do we know about Loss Given Default? Working Paper, No. 04-01.
Tong, E.N., C. Mues and L. Thomas. 2013. A zero-adjusted gamma model for mortgage loan loss given default. *International Journal of Forecasting* 29(4): 548–562.
Zhang, Y. 2013. Fair Lending Analysis of Mortgage Pricing: Does Underwriting Matter? *The Journal of Real Estate Finance and Economics* 46(1): 131–151.
Yang B.H. and M. Tkachenko. 2012. Modeling exposure at default and loss given default: empirical approaches and technical implementation. *Journal of Credit Risk* 2(8): 81–102.

Appendix

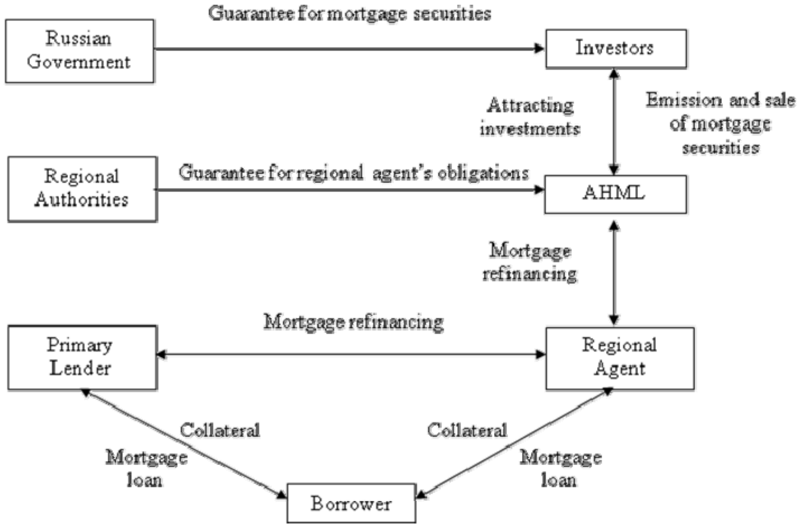

Fig A1. AHML's lending system

Tab. A1 Descriptive statistics

Variable	Description	Mean	Std. Dev.	Min	Max
Probability of endorsement	1 - credit organization approves mortgage application; 0 - otherwise	—	—	—	—
Probability of contract agreement	1 - contract agreement; 0 - otherwise	—	—	—	—
Probability of default (*PD*)	1 - mortgage default (90+ days delayed); 0 - otherwise	—	—	—	—
	Socio-demographic characteristics (4298 applications)				
Age of client	Years	34	7.6	21	61
Gender	=1, male	—	—	—	—
Borrower's income	Monthly declared borrower's income, thou. Rus. rub.	30.7	26.2	1.7	385.6
Income of co-borrowers	Monthly declared co-borrowers's income, thou. Rus. rub.	17.7	11.6	38.3	72.8
	Contract terms (2799 contracts)				
Credit limit	Credit limit, mln Rus. rub.	1.1	0.6	0.12	12.7
Loan amount	Mortgage loan amount, mln Rus. rub.	1.0	0.6	0.12	10.0
Credit interest rate	Fixed credit interest rate,%	11.59	1.64	9.55	19
Type of credit interest rate	=1, fixed credit interest rate	—	—	—	—
Maturity	Years	15.75	5.18	2.17	30
Downpayment	Downpayment, mln Rus. rub.	0.9	0.7	0.04	13.8
Flat value	Assessed mortgaged property value at origination, mln Rus. rub.	1.9	1.1	0.3	15.3
Monthly payment	Monthly mortgage payment , thou. Rus. rub.	12.6	7.3	1.9	140.0
LTV	Mortgage loan amount to assessed mortgaged property value ratio at origination (loan-to-value ratio), [0;1]	0.56	0.17	0.11	0.94
DTI	Monthly payment to monthly borrower's income ratio (debt-to-income ratio), [0;1]	0.45	0.18	0.06	0.99
Credit duration	Months	28.9	13.99	0.6	49.57
	Macroeconomic variables (49 months)				
Unemployment rate	Regional quarterly unemployment rate, %	8.4	1.5	6.3	10.9
Regional property value	Average property value per 1 sq. meter in region, thou. Rus. rub.	38.6	6.2	28.8	51.3
Refinancing rate	Refinancing rate of Central Bank of Russia, %	9.44	1.78	7.75	13

Variable[16]	Total	%
Socio-demographic characteristics (4298 applications)		
Gender		
female	1879	43.7
male	2419	56.3
Marital status		
not declared	46	1.1
single	1220	28.4
married	2358	54.9
widowed	56	1.3
divorced	618	14.4
Category of employment		
not declared	138	3.2
unemployed	1	0.0
soldier	13	0.3
hired employee	3963	92.2
entrepreneur	39	0.9
state-owned employee	144	3.4
Education level		
not declared education level	205	4.8
primary education	65	1.5
secondary education	1748	40.7
not complete higher education	138	3.2
complete higher education	2142	49.8
Monthly income of borrower		
unverified	2918	67.9
0-9999 руб.	118	2.8
10 000-19999 Rus. rub.	376	8.8
20000-39999 Rus. rub.	597	13.9
≥40000 Rus. rub.	289	6.7
Income of co-borrowers		
unverified	3724	86.6
0-9999 Rus. rub.	159	3.7
10000-19999 Rus. rub.	225	5.2
≥20000 Rus. rub.	190	4.4
Contract terms (2799 contracts)		
Type of credit interest rate		
fixed	2421	86.5
adjustable	378	13.5
Maturity		
< 10 years	181	6.5
10-14.9 years	595	21.3
15-19.9 years	1106	39.5
20-24.9 years	690	24.6
≥25 years	227	8.1
LTV		
<0.5	968	34.6
0.5-0.7	1531	54.7
≥0.7	300	10.7
DTI		
unverified	1651	59.0
<0.2	41	1.5
0.2-0.4	505	18.0
0.4-0.6	379	13.5
0.6-0.8	160	5.7
≥0.8	63	2.3

[16] All categorical variables are included in the regression model as a set of dummies.

Variable[17]	Total	%
Additional variables (4298 applications)		
Time of mortgage application		
2008-2009 years	1821	42.4
2010-2012 years	2477	57.6
Type of creditor		
regional operator AHML	1856	43.2
primary lenders	2442	56.8
Region of mortgage loan		
not declared	1370	31.9
base region	2849	66.3
other regions	79	1.8
Type of credit program		
not declared	1532	31.3
non special	2372	48.4
special	993	20.3

[17] All categorical variables are included in the regression model as a set of dummies.

Corresponding author:

Agata M. Lozinskaia
National Research University Higher School of Economics (Perm, Russia).
Senior Lecturer at the Department of Economics and Finance
Junior Research Fellow at the Group for Applied Markets and Enterprises Studies
E-mail: AMPoroshina@gmail.com