КОНСОРЦИУМ ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ И ОБРАЗОВАНИЯ
ECONOMICS EDUCATION AND RESEARCH CONSORTIUM
Working Paper Series

ISSN 1561-2422

KEY DETERMINANTS OF DEMAND, CREDIT UNDERWRITING, AND PERFORMANCE ON GOVERNMENT-INSURED MORTGAGE LOANS IN RUSSIA

Agata Lozinskaia, Evgeniy Ozhegov

Working paper No E16/03

This project (No 14-5352) was supported
by the Economics Education and Research Consortium
and funded by GDN

Agata Lozinskaia

National Research University Higher School of Economics, Department of Economics and Finance, Laboratory of Interdisciplinary Empirical Studies Research Group for Applied Markets and Enterprises Studies; Senior Lecturer, Junior Research Fellow

27, Lebedeva Street, 614000, Perm, Russia

Tel.: +7 (342) 2009552, email: AMPoroshina@gmail.com

Evgeniy Ozhegov

National Research University Higher School of Economics, Department of Economics and Finance, Laboratory of Interdisciplinary Empirical Studies Research Group for Applied Markets and Enterprises Studies; Senior Lecturer, Junior Research Fellow

27, Lebedeva Street, 614000, Perm, Russia

Tel.: +7 (342) 2009552, email: tos600@gmail.com

JEL Classification: C36; D12; R20

Keywords: mortgage, demand, credit risk.

Abstract

This research analyses the process of lending from Russian state-owned mortgage provider. Two-level lending and insurance of mortgage system lead to substantially higher default rates for insured loans. This means that underwriting incentives for regional operators of government mortgage loans perform poorly. We use loan-level data of issued mortgage by one regional government mortgage provided in order to understand the interdependence between underwriting, choice of contract terms including loan insurance by borrower and loan performance. We found an evidence of a difference in credit risk measures for insured and uninsured loans and interest income.

Contents

1. Introduction

Key issues of government policy include providing of affordable housing, identifying the main drivers of mortgage borrowing and performance of mortgage loans. Therefore, the problem of developing optimal credit contracts and effective risk management systems, especially on the residential mortgage market, is becoming crucial.

The mortgage lending system in Russia is based on the principles of the two-level system (Fig. 1). National institute for development of housing activity - Agency of Home Mortgage Lending (AHML) was set up in 1997 as an analog of Fannie Mae and the first steps were taken towards the introduction in Russia mortgage securities. AHML is a quasi-governmental specialized financial institution or, in other words, a government-sponsored enterprise (GSE) (Khmelnitskaya, 2014). It helps to implement strong government housing policy and anti-recessionary measures to support mortgage lending in Russia. In essence, AHML is the national regulator of the mortgage market.

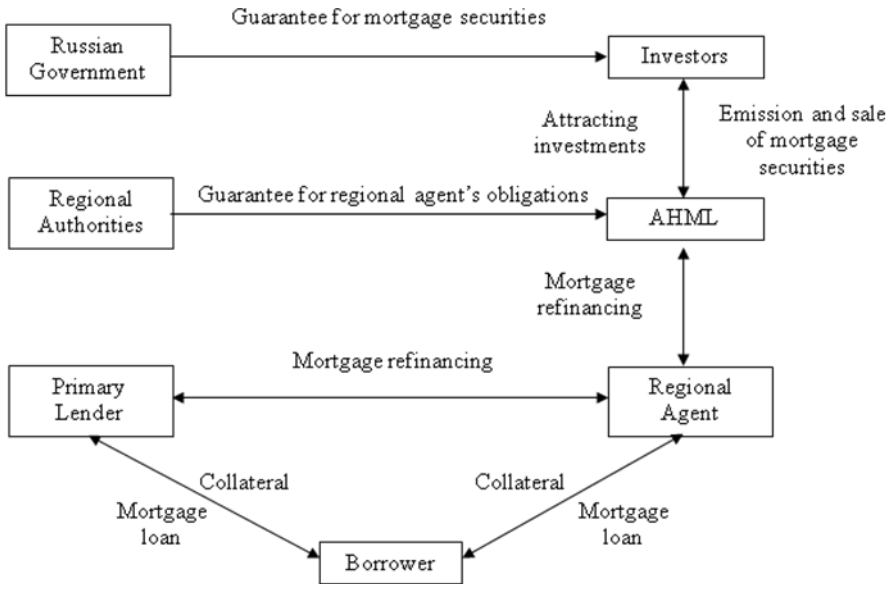

Figure 1. AHML's lending system

AHML uses the two-level system of lending (Fig. 1) where in the first step banks and non-credit organizations provide mortgage loans to households according to the common standards of AHML. The second step is refinancing (redemption) of mortgage receivables by AHML. AHML develops both nonspecial and special mortgage programs and refinances risks from its regional branches and commercial banks, which operates such programs. Terms of nonspecial mortgage programs do not very differ from ones offered by other credit organizations (primary lenders). For this market segment AHML is acting as price-taker because terms of credit for such programs are mainly defined by the largest mortgage lender in the country - state-controlled bank Sberbank. However, special mortgage programs in some sense are unique credit products offered only by

AHML. The list of special programs contains «Young researchers", «Young teachers», «Mortgage for Soldiers», «Mothers' capital» and other social and subprime programs including «self-certified». Special programs are usually linked with the lower downpayment. If downpayment is less than 30% then borrower's third party liability must be insured in «AHML insurance company». While the Russian residential mortgage market has been stable over the past 8 years from the point of the mean probability of default varied from 4 to 5%[1], government-insured AHML loans performed substantially worse and showed a 16% probability of default. This means that government insurance covers potential losses from such loans and may affect its approval process. We are interested in the conditions leading to having a government-insured loan, its performance and the underwriting process of such loans. As a part of the research we study the distribution of credit risk between special and nonspecial programs and discusses the methodology for estimation one of the crucial credit risk parameter, especially from the government policy – Loss Given Default (*LGD*).

Obtained results can help to the tradeoff between achieving social goals and credit risk losses for the government. Also, it may help to revise the underwriting process and incentives for regional AHML operators.

This study has the following structure. It starts with literature review and some generalization of recent studies of the probability of mortgage default and loss given default modeling. The second and the third part contain the description of data and identification strategy. Then we discuss empirical results and conclude with its policy and with further work.

2. Literature Review

Different concepts are used to measure credit risks, such as the probability of default (PD), loss given default (LGD), exposure at default (EAD), maturity (M) and correlated defaults (CD). The default is regarded as the worst event of credit risk and it is arguably most relevant to the recent Russian mortgage crisis (2008-2009) and related spillover effects.

The notion of mortgage default has not yet been incorporated in the Russian legislation. According to BIS (2006) and the Bank of Russia (2012) a borrower is in default if any of the following credit events happen:

- a borrower cannot repay a loan without selling collateral (for mortgage loans, collateral is real property);
- monthly payments are not met for 90 days or more.

[1] Agency of Housing Mortgage Lending data, www.ahml.ru

Credit organizations are used different actions in working with past-due payments and cases of mortgage defaults. Hard (1-20 arrears days) and soft collections (30-90 arrears days) include negotiation process with the borrower. In Russia usually after 90 arrears days, bank starts a legal collection that can lead to the disposal of pledged property with a haircut. The last action is connected with a higher cost for a bank. For this reason, bank is interested in a nonlegal collection of past-due mortgage payments.

From the beginning of the 1960s and extending to the present, an important stream of literature has addressed to the default problem. Two theories are used to explain borrower's incentives for mortgage default: *the ability-to-pay theory* and *the put-option theory*. The ability-to-pay theory predicts that the mortgage default decision would be expected to be mainly driven by the borrower's debt-to-income ratio (DTI). Under the put-option theory, the borrower's mortgage default decision is determined by the ratio of financial gains and losses or, in other words, the ratio of the market loan value to property value. In this case, default is regarded as put-option (Deng et al., 2000). The importance of the last theory was firstly demonstrated in the paper (Vandell, 1978). One of the first attempts to examine empirically these approaches was made by Jackson and Kaserman (1980). They have shown based on American mortgage data, the higher explanatory power of the put-option theory. This finding was supported by the later studies (Bhutta et al., 2010; Bajari et al., 2008; Foote et al., 2008; Ambrose et al., 2005). At the same time, an excess of the outstanding balance of the debt collateral value of property is a necessary, but not sufficient condition for the mortgage default (Vandell, 1995; Archer et al., 1996; Clapp et al., 2001; Pavlov, 2001; Deng et al., 2005).

Classical binary choice models (probit- and logit-) of mortgage default are widely used parametric approaches to constructing regression functions for PD. The main issue in such models is sample selection bias that arises with self-selection of borrowers not to participate in some steps of borrowing process. Moreover, self-selection generates partial observability of contract terms and loan performance data. Thus, we only have this data for all approved borrowers and for those who signed the mortgage contract. Then the magnitude of sample selection bias depends on the strength of correlation between underwriting process, choice of credit terms and loan performance (Ross, 2000).

The second issue when default modeling is simultaneity bias. Mortgage borrowing as a sequence of consumer and bank decisions firstly introduced by Follain (1990). He defines the borrowing process as a choice of how much to borrow (the Loan-To-Value ratio, LTV decision), if and when to refinance or default (the termination decision), and the choice of mortgage instrument itself (the contract decision). Later, Rachlis and Yezer (1993) suggested a theoretical model of mortgage lending process, which consists of a system of four simultaneous equations:

(1) borrower's application, (2) borrower's selection of mortgage terms, (3) lender's endorsement, and (4) borrower's default. They showed that all of four equations (and decisions) should be considered as interdependent and if it is not so then the estimated would be inconsistent. Later, Rachlis and Yezer (1993) suggested a theoretical model of mortgage lending process, which consists of a system of four simultaneous equations: (1) borrower's application, (2) borrower's selection of mortgage terms, (3) lender's endorsement, and (4) borrower's default. They showed that all of four equations (and decisions) should be considered as interdependent and if it is not so then the estimated would be inconsistent. Both papers discussed estimation techniques that can be used for such kind of simultaneous equation systems. They also explored the necessity for a better understanding of mortgage choices to answer important policy questions, but without any empirical framework.

From the mid-1990s, such data as American mortgage datasets from the Federal Housing Authority (FHA) foreclosure, The Boston Fed Study, The Home Mortgage Disclosure Act (HMDA) became publicly available. Then several empirical studies analyzed mortgage lending process and studied the interdependency of bank endorsement decision and borrower's decisions modeled by bivariate probit model using this sort of data.

As an extension of study (Rachlis, Yezer, 1993), Yezer, Phillips, Trost, (1994) applied Monte-Carlo experiment to estimate the above-listed theoretical model. They empirically showed that isolated modeling processes of the credit underwriting and default lead to the biased parameter estimates. Later on Phillips and Yezer (1996) and Ross (2000) supported these findings.

Phillips and Yezer (1996) compared the estimation results of the single equation approach with those of the bivariate probit model. They showed that discrimination estimation is biased if the lender's rejection decision is decoupled from the borrower's self-selection of loan programs, or if the lender's underwriting decision is decoupled from the borrower's refusal decision.

Ross (2000) studied the link between loan approval and loan default by bivariate probit model and found that most of the approval equation parameters have the opposite sign compared with the same from the default equation after correction for the sample selection. It was explicitly shown that not taking into account nonrandom selection of borrowers by the bank during the underwriting process leads to 15-30% biased estimates in default equation. Studying default decisions we need to take into account that some variables which determine default also explain underwriting decision including unobserved in data borrowers' characteristics. This makes a sample of approved borrowers biased comparing with the general population. Obtained biased estimates may lead to incorrect interpretation of causal relationships between the probability of default or other credit risk measures with its true determinants. Hopefully, as more information on borrower's characteristics is available, including credit history and other risk metrics, as less the sample selection bias will be, but it may only be measured after applying a correction for it.

As key determinants of default on mortgage contract usually considered socio-demographic and financial characteristics of borrowers and contract terms. When data on characteristics of borrowers is unavailable, some papers, for ex. (Bajari et al., 2008), deal with aggregated demographics and unemployment rate as proxies for individual demographics.

In addition, Central Bank of Russia develops a methodological recommendation for developing and implementation Internal-Ratings-Based Approach (IRB-approach) (Central Bank of Russia, 2012) and Moody's developed MILAN analysis - Moody's Individual Loan Analysis of Residential Mortgage-Backed Securities (Moody's, 2009).

Relatively few papers focus on *LGD* modelling. For mortgage lending specifically, the literature is much more limited. It is mainly explained by the lack of publically available data and historical losses given default. Two approaches are widely used to estimate and model LGD in mortgage lending: *accounting (accounting LGD)* and *economic (economic LGD).* The main difference is that economic approach is based on the discounted cash flows method and takes into account the time value of cash flows. In mortgage default literature accounting (Frye et al., 2000; Pennington-Cross, 2003; Leow, Mues, 2012; Zhang, 2013) and economic approaches (Qi, Yang, 2009) are used. The paper (Araten et al., 2004) has shown the difference in mean values for accounting (27%) and economic *LGD* (39.8%) based on data from a universal bank JPMorgan Chase for the 18-year period (1982-1999) in 3761 defaulted borrowers. However to predict the typical difference in the estimates is difficult because it is mainly determined not only by the characteristics of the loan portfolio of a particular bank, macroeconomic conditions but also assumptions used in calculations of these indicators. In addition, economic *LGD* modelling is attended with some difficulties such as assumptions of on discounted cash flows method including assumptions about discounting rate and measurement errors (determination of exact time of working with past-due payments, calculation cost of workout process etc.). In bank practice, the accounting *LGD* is more frequently used especially to calculate expected losses (*EL*) and risk-weighted assets (RWA) that are important from the point of bank credit risk management and effective capital allocation.

In *LGD* modeling classical linear regression model and *R*-squared as a measure of the model are widely used (0,04–0,06 (LaCour-Little, Zhang, 2014), 0,06–0,17 (Lekas et al., 1993), 0,15 (Qi, Yang, 2009), 0,2 (Araten et al., 2004), 0,95 (Pennington-Cross, 2003). Several papers have shown that *LGD* has censored distribution with higher values in the period of economic recessions and application of classical linear regression model could lead to incorrect inference (Qi, Zhao, 2011; Dermine, Carvalho de, 2006; Araten et al., 2004; Schuermann, 2004; Felsovalyi, Hurt, 1998). In a variety of papers transformation of dependent variable is used (log-log function

(Dermine, Carvalho de, 2006); beta-distribution (Qi, Zhao, 2011; Bellotti, Crook, 2012; Huang, Oosterlee, 2012; Gupton, Stein, 2005) or gamma-distribution (Yashkir, Yashkir, 2013;Sigrist, Stahel, 2011). However as mentioned in the paper (Yang, Tkachenko, 2012), using the transformation of dependent variable could lead to high measurement errors. An alternative way assumes special classes of econometric models (tobit model, fractional response regression). Empirical results for the various debt instruments from the S&P LossStat (1990-1991, 2001-2002. 2008-2009) of Yashkir, Yashkir (2013) confirmed that the classical linear regression and beta regression have the highest predictive power.

In early paper (Lekkas et al., 1993) for the data on mortgage loans issued in the period 1975-1990 refinanced at Freddie Mac has shown empirically that the high *LGD* associated with a high LTV ratio of on the mortgage loan issue date, the geographical location of the territory with a high rate of defaults, as well as short-lived age credit. The positive dependence *LGD* and LTV were supported later by several empirical studies (Qi, Yang, 2009; Calem, LaCour-Little, 2004; Pennington-Cross, 2003). Empirical studies discussed negative relationship *LGD* not only with credit age but the loan amount (Calem, LaCour-Little, 2004; Pennington-Cross, 2003).

Special attention in modeling LGD should be paid to the characteristics of the collateral. In the paper (Leow, Mues, 2012) proposed a two-step LGD model for mortgage loans that includes the probability of repossession model and haircut model. Empirical results showed that the procedure used by the authors provided the higher goodness of fit for the observed LGD values. However, the authors did not discuss the effects of macroeconomic conditions (Bellotti, Crook, 2012; Qi, Yang, 2009) and comparative results.

Summarizing findings of recent research it should be mentioned that 1) When model borrowers’ default decisions we should consider simultaneity and interdependency of choice in all stages of borrowing process; 2) Errors in contract terms, credit risk and demand equations may be biased by sample selection; 3) The nature of error terms correlations and regression functions can be nonlinear and is much complicated to specify parametrically; 4) *LGD* has censored and even bimodal distribution.

3. Data Description

The data used for this research is provided by the regional AHML operator on all applications for mortgage loans collected from 2008 to 2012. In post-crisis period 2010-2012 the total number of mortgage applications was 15% higher comparing with crisis period 2008-2009 (Fig. 2).

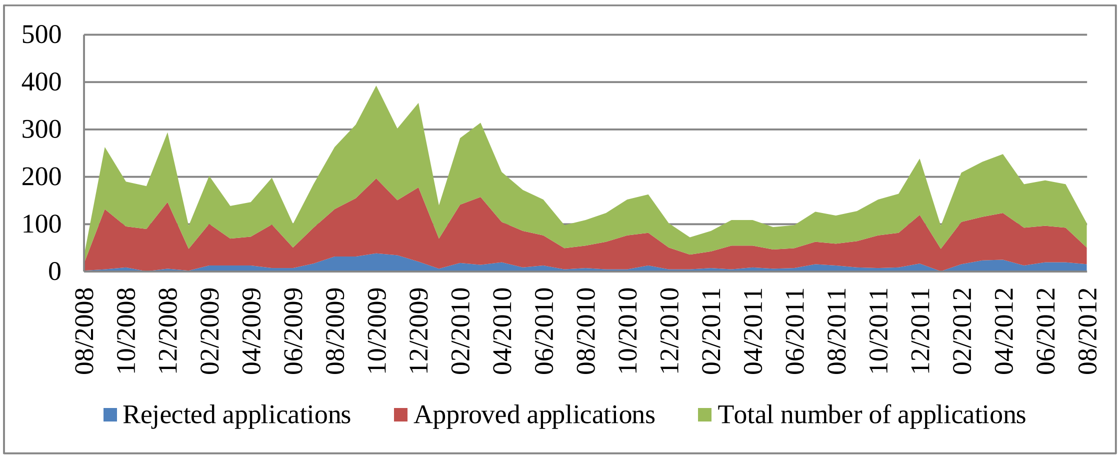

Figure 2. Number of mortgage applications from August 2008 to August 2012

The individual-level dataset that contains the socio-demographic characteristics of each of the 4298[2] applicants as main potential borrowers and their co-borrowers on the date of application and flags of approval and contract agreement. Unfortunately, the sample excluded 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 on the date of application. The characteristics of the borrower are fully observable and the contract terms are partially observable for only the subsample of applicants who signed the contract. Default occurs in 166 mortgage loans that are equivalent to 6% default rate. In data set, the flag of mortgage default is observed, but not the default date, bank actions in past-due payments and default losses.

Some mortgage programs allow the applicants to provide confirmation of their income in the «free form» that are known as «self-certified» or «low doc». These programs are usually linked with a higher contract rate. The reason for this choice may be explained by a temporary or changeable income (LaCour-Little, 2007), for instance, for entrepreneurs. Generally, income should be considered endogenous while modeling the approval of borrower or contract terms. However, we can control for employment category, which rejects the inconsistency due to 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 of co-borrowers. This is a limitation of the research. But we may consider it as insignificant for the

[2]Initially we had 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.

choice of contract terms compared to the income of the main borrower.

In addition, we use some macroeconomic factors at the regional level that are publically available at Bank of Russia, Federal Bureau of Statistics Russian Federation and AHML websites. These are regional quarterly unemployment rate and refinancing rate. Descriptive statistics are presented in Table A1-2.

The region itself may be described as old industry region with two major agglomerations concentrating with machinery and chemical industries. The total population is 2.7 mln people. Income level as well unemployment rate, affordability of housing, housing provision and mean price of housing is near the mean Russian level. The volume of issued loans *per capita* comparing with national level is 20% higher.

We perform a set of tests for equal means, medians and variances in different groups of mortgage applicants and borrowers. The corresponding results are reported in the Table A3. Approved applicants have statistical differences comparing with rejected ones in terms of their socio-demographic characteristics such as age, education level, workplace, income, coborrowers income. They are traditionally included in credit underwriting systems. Approved and rejected applicants also statistically differ in terms of macroeconomic conditions under which they apply for mortgage loans, especially in our data set they are characterized by differences in regional property value and refinancing rate. In addition, we find statistical differences in terms of types of creditor and credit program, a region of mortgage loan that may signal a difference in credit underwriting policy of regional operator AHML in deferent regions and primary lenders and their incentives to support some credit program, for example, socially oriented according to AHML strategy.

The total number of approved applications (3698) is not equal to all issued loans (2799) for different reasons that are mainly consisted of a better bids of other creditors, issues with finding and buying appropriate property in line with credit limit and mortgage contract conditions, dramatic change in personal income, macroeconomic conditions etc. But in our data set, we do not observe these reasons. The results of the discriminate analysis for these groups of clients allow us to identify statistical difference not only in the above-mentioned socio-demographic, the most of macroeconomic and additional factors for approved and rejected applicants but also in their gender, marital status, unemployment time of mortgage application and region of the mortgage loan. These findings indicate the need for further investigation and may be important mainly for strategic and marketing policy design of creditors.

Obviously, in the results of the discriminate analysis for defaulted (166) and non-defaulted (2633) mortgage borrowers (see Table A3) we find some of the same variables with high discrimination power as in results for approved and rejected applicants. They include age, workplace, educational level, income, coborrowers income from socio-demographic characteristics, regional mean property value, refinancing rate from macrovariables and types of creditor and credit program as a region of a mortgage loan from additional variables. It is not surprising because the probability of default estimation is included in the credit underwriting and decision-making process. We find also a statistical difference for defaulted and non-defaulted groups in terms of their credit parameters, including categorical debt-to-income ratio and loan-to-value ratio, according to with the above-mentioned the ability-to-pay and the put-option theories to explain borrower's incentives for mortgage default. In addition, we cannot reject the assumption that the means, medians and variances for «credit duration» are homogeneous for defaulted and non-defaulted loans. This variable will play an important role in the developed methodology to approximate historical accounting loss given default that will be discussed further. The variables with high discrimination power are used in the probability of endorsement and the probability of default equations as explanatory variables.

4. Methodology

4.1. Identification strategy

Mortgage borrowing process can be represented by following sequence of decisions:

1. Application of borrower. A potential borrower realizes the necessity of borrowing, chooses the credit organization and credit program that reflects her preferences, fills an application form with demographic and financial characteristics.

2. Approval of borrower. Considering application form and recent credit history, credit organization endorses the application or not, inquires the form data (some banks also set the loan amount limit when the borrower is endorsed).

3. Choice of credit terms. The approved borrower makes a choice on the property to buy and credit terms: loan amount (not more than the limit), down payment, the presence of insurance and maturity. All of this determines the interest rate and annual payment.

4. Loan performance. Borrower chooses the strategy of loan performance: to pay in respect to contract terms, to default or prepay.

Econometric model partly repeats steps of the structural one. Then it may be represented by following equations:

$$d_i = \begin{cases} 1, g_0(w_{0i}, x_i) + e_{0i} \geq 0 \\ 0, g_0(w_{0i}, x_i) + e_{0i} < 0 \end{cases}$$
$$\begin{cases} y_{1i}^* = g_1(x_i, w_{1i}, y_{-1i}^*) + e_{1i} \\ \dots \\ y_{ki}^* = g_k(x_i, w_{ki}, y_{-ki}^*) + e_{ki} \end{cases} \tag{1}$$

$$def_i^* = \begin{cases} 1, g_{def}(y_i^*, x_i) + e_{def,i} \geq 0 \\ 0, g_{def}(y_i^*, x_i) + e_{def,i} < 0 \end{cases}$$

$$(y_i, def_i, x_i) = \begin{cases} (y_i^*, def_i^*, x_i) \; is \; observed, if \; d_i = 1 \\ \quad x_i \; is \; only \; observed, otherwise \end{cases}$$

where d_i is a binary indicator of contract signing (both bank's and borrower's decision), x_i is a set of demographic and financial characteristics of the borrower and co-borrowers, y_i is a set of credit terms (contains loan limit, LTV, logarithm of rate, logarithm of maturity and probability of having the government insurance). def_i is a binary indicator of default. $(w_{0i}, w_{1i}, \dots, w_{ki})$ is a set of excluded instruments for contract signing decision, credit terms and loan limit respectively. Discussion on the instruments used is provided in next section.

The paper of Ozhegov (2015) extends methods for identification and estimation of a non-triangular system of simultaneous equations with sample selection, endogenous regressors and arbitrary joint error distribution and functional form of regression and control functions in reduced and structural forms. We may apply this method to estimate model (1) with the following steps.

1. Firstly, we need to estimate the propensity score for the contract agreement equation:

$$p = E[d|x_0, w_0] = g_0(w_0, x_0) \tag{2}$$

2. Then we will estimate each contract term equation in the reduced form corrected for sample selection using estimates of propensity score:

$$E\left[y_j \middle| x, w, w_0, d = 1\right] = \gamma_j(x, w) + \mu(\hat{p}) \tag{3}$$

3. On the next step we will estimate the structural form contract terms equations corrected for sample selection and simultaneity using estimates of propensity score and reduced form contract terms residuals:

$$E\left[y_j \middle| x, y_{-j}, w_j, w_{-j}, w_0, d = 1\right] = g_j\left(x, w_j, y_{-j}\right) + \varphi(\hat{p}, \hat{e}_{-j}) \tag{4}$$

4. On the last step we will estimate the probability of default equation corrected for sample selection and endogeneity of contract terms using propensity score and structural form residuals:

$$E[def|x, y, w, d = 1] = g_{def}(x, y) + \varphi_{def}(\hat{p}, \hat{e}) \tag{5}$$

In Ozhegov (2015) it was show that if all regression and correction functions are continuously differentiable and we have at least one excluded variable for selection equation and matrix of instrument's marginal effects in reduced form contract terms equations has full rank then equations (2)-(5) is identified up to additive constant.

An estimation procedure is based on approximation by series of power functions which depend on an initial set of regressors.

4.2. Estimation of Loss Given Default

The data set excluded information on the default date and realized loss on loans that default. Our general approach is to approximate historical accounting *LGD* (Expected Loss Given Default, *ELGD*) by using the probability of default estimation from the previous step and assuming the sale of mortgaged properties due to foreclosure at discounts relative to other flats on the market in case of mortgage default. The following methodology is applied to estimate *ELGD*:

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 a 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 1).

2. Predicting $\hat{PD}_{it}$ for the *i* defaulted mortgage loan time at each time point *t* based on probit *PD* model with correction for sample selection (see Tables 2-3). We should mention that predicted *PD* based on probit-model represents cumulative *PD*, but we are interested in noncumulative *PD* represented in column 3 in Table 1. For example, noncumulative *PD* at *t*=5 is calculated as the difference in cumulative *PD* estimates at *t*=5 and *t*=4.

Table 1

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 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 loss occurs given default is measured as:

$$LGD_{it} = \frac{EAD_{it} - R_{it} + C_{it}}{EAD_{it}} = 1 - \frac{R_{it} - C_{it}}{EAD_{it}} \qquad (7)$$

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

$$ELGD_i = E(LGD_i) = \sum_{t=1}^{M} \hat{PD}_{it} \cdot LGD_{it},\ t = 1, \ldots, M, \tag{8}$$

where M – 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.

5. Repeating steps 1-4 N times that corresponds to number of defaulted mortgage loans in the credit portfolio and approximating portfolio expected losses (EL) as follows:

$$EL_p = \sum_{i=1}^{N} EL_i = \sum_{i=1}^{N} \hat{PD}_i \cdot EAD_i \cdot LGD_i,\ i = 1, \ldots, N, \tag{9}$$

To obtain an empirical distribution of losses rate in the credit portfolio several assumptions are made and mainly determined by the empirical data. Firstly, we assume that time is discrete (month) that corresponds to the frequency of making the monthly payment. Secondly, due to the lack of information about ways of past-due payments regulation by the creditor, we assume that creditor uses the legal collection in case of mortgage default. Thirdly, in approximation of R the ratio the assessed property value (for 1 sq. meter) at the mortgage loan issue date to regional market property value (for 1 sq. meter) at the end of workout process T, that is assumed constant over the time, and total square of flat (see Table A4) are used. Haircut equals to 20% that is determined by the current legislation. Workout process T takes 5 months from the approximated default date that is based on an expert estimate of AHML. Fourthly, AHML estimates workout process costs 5-15% from the current collateral value. 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:

$$EAD = A \times (M - t_1 + 3) + P, \tag{10}$$

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

In addition, we assume that credit portfolio consists of homogeneous mortgage loans and number of mortgage borrowers and signed contracts are equal. Related borrowers are absent as result correlation between defaults of related mortgage borrowers is absent. «Repeated»[3] mortgage defaults («default» — «no default» — «default») are absent. Finally, the approximated accounting LGD is censored on the interval [0;1][4] as in (Li et al., 2009).

---

[3] This problem is discussed in the paper (Dermine, Carvalho de, 2006).

[4] Approximately 2% of defaulted loans have LGD out of the unit interval.

5. Empirical Results

5.1.Contract Terms

To estimate the model we need to find a set of relevant excluded instruments for the probability of signing a contract, the loan limit and each credit term.

Bajari et al. (2008) discussed the possibility of using aggregated district-level variables as proxies for unavailable data. We will use the same strategy to find the set of instruments. Since we have data without spatial variation we can use time variation in applications. We have data from July 2008 to August 2012 and we know the application date for each applicant. Each application was matched with the set of aggregated mortgage and housing market characteristics for the same month. On average, the process takes two months from the date of application to the date of contract agreement. Then we need to fix the aggregated market characteristics for each application not only in the month of application, but also the 1-2 months prior the application, and use these as instruments.

Table A3 represents the descriptive statistics of aggregated mortgage and housing market characteristics for the period from July 2008 to August 2012 (50 months).

About 15% of issued loans were refinanced by AHML, but not all of them were issued by the bank supplying the data. Generally, the number of applications to the bank is fewer than the number refinanced by AHML by all the regional banks.

The difference between the number of loans refinanced by AHML and the number of applications to the bank within a particular month may be the excluded variable which explains the probability of contract agreement, but it does not affect the contract term choice. Since every commercial bank operates with the same AHML programs, the difference in the approval process does not affect the terms choice. But an increase in the number of refinanced loans shows the changes in the underwriting process in other banks and may correlate with the probability of a contract agreement with the bank. This variable should be considered as exogenous since each individual decision explains a negligible variation of the aggregated market characteristic (less than 1%).

As an excluded instrument for the loan limit, we use the mean Debt-to-Income ratio (DTI). The positive dependence of these two variables is expected because the mean DTI for all issued loans reflects the evaluation of the mean credit risk (the higher the DTI of issued loans, the less risk). It positively correlates with the loan limit, which reflects the willingness to issue a larger loan for a particular borrower. This variable is also valid since individual shocks of loan limit do not affect the aggregated characteristic of issued loans.

As excluded instruments for credit terms, LTV, interest rate, maturity and the probability of having insurance, we used mean LTV, median rate, median maturity for issued loans and the housing affordability coefficient. The relevance of the first three instruments is implied by the interdependence of mortgage market characteristics and the AHML credit programs conditions. Validity is implied by the exogeneity of the program terms for each particular borrower.

The affordability coefficient is relevant for the probability of insurance because the increase of affordability should lead to the choice of a lower LTV and consequently to a lower probability of loan insurance. Validity is also implied by the independence of individual preference on insurance shocks and the aggregated affordability of housing. All the variables are may be considered as valid and may be used as instruments.

First, we estimated the model of the probability of endorsement based on the characteristics of the borrower and co-borrowers and the difference between the number of AHML refinanced loans and the number of applications. The last variable which was taken as an excluded instrument is significant at the 1% level. The sign and significance of borrower characteristics are consistent with recent research. The demographic characteristics, such as age, sex, marital status of the borrower are insignificant, which supports the absence of discrimination. However, borrowers are discriminated by the level of education. The probability of a contract agreement is positively correlated with the income of the main borrower and, on the contrary, negatively correlates with the failure to provide income details. Entrepreneurs have a higher probability of a contract agreement *ceteris paribus*. The propensity score $\widehat{p_i} = E[d_i|x_i, w_{0i}]$ was obtained from the probit model.

For each credit term, we estimated the reduced form equation. The control function was approximated by the polynomial with power M on the estimate of the propensity score and the loan limit equation residuals. The regression function was estimated as partially polynomial. It was linear for the characteristics of the borrower and polynomial for the excluded instruments for contract terms with power M. We test three set of instruments described earlier. First, we fix market-level variables on the month of application. For the second and third sets, we used market-level data for month one and two months before the month of application respectively. The proof of the relevance of excluded instruments based on conditional F-test (Sanderson, Windmeijer, 2014) is provided in Table A4. All sets of excluded instruments are relevant on 5% level. The set of market-level variables fixed in the month before the month of application (models (II)) is significant at 1% level. We use then the reduced form residuals obtained from models (II).

We estimated the contract term equations in the structural form using a polynomial approximation with power ξ_2 for the control function on $\hat{p}$, residuals from the loan limit equation and the reduced form of the contract term equations. The regression function was partially polynomial, linear for the characteristics of the borrower and polynomial with power ξ_1 for the credit terms and loan limit. Estimation results are provided in Table 2.

The sign and significance for the majority of marginal effects remain the same with the increase of the polynomial power. This supports the robustness of the results. The significance of propensity score marginal effects in contract terms equations supports the hypothesis of nonrandom sampling of applicants who took a loan. Marginal effects of terms interdependence in models without correction ($\xi_2 = 0$) for sample selection, the endogeneity of the loan limit and the simultaneity of contract term choice are significantly different from the corrected ones (comparing models (3) and (4) for each equation). This results in evidence inconsistency of the estimates without correction and the necessity of using the proposed estimation procedure.

The result of estimates in LTV equation is not counterintuitive and supports recent results. LTV is higher when the rate is lower and maturity, the probability of having insurance and loan limit are higher. The interest rate will increase with higher LTV which is consistent with mortgage programs design. However, we expected positive dependence of rate on maturity and no on insurance but it is rejected by estimation results. The choice of longer maturity is linked with higher LTV, rate and lower loan limit that support the recent results that maturity is a very flexible instrument of monthly payments adjustments for borrowers with credit constrained borrowers.

Previous literature on mortgage credit risk found that riskier contracts are with higher LTV, interest rate, and low maturity. By our estimation results, there are the same terms that lead to the choice of insurance which means that insurance is more likely to be chosen for contracts that are riskier in common sense.

Marginal effects of selection propensity score reflect the covariance between error terms in selection and terms equations. Since the exact terms of the contract are unobservable in data on the stage of borrower's selection we may interpret these results as dependence between the probability of selection of borrower and contract terms that bank expects for particular application during the underwriting. Thus, the application is more likely to be underwritten if the bank expects lower LTV and long maturity which is usually linked with potential good loan performance. The probability of being underwritten is higher with a higher rate. On the one hand, it looks counterintuitive because higher rate usually leads to higher credit risk. But the increase of rate leads to reduced LTV, longer maturity and a higher probability of insurance which means that credit risk of increased rate will be compensated by the choice of different terms. On the other hand, higher rate makes higher bank's profit with all other fixed. This leads to the preference for loans with higher interest rate by the bank.

Better underwriting is also linked with a higher probability of loan to be insured despite the fact of higher default rates for insured loans. But it is clear that expected loss given default of insured loan is less than if it is not insured. So we only may conclude that if bank and borrower expect high credit risk they tend to agree on the government insurance of borrower's liability which leads to a better underwriting process because the risk is relocated from bank to insurance company. The borrower pays extra for insurance but this is the way to be selected to mortgage lending process. From the government point of view, it should be the case that large risk taken by AHML insurance company is the cost of providing housing for higher risk borrowers who cannot participate in mortgage lending provided by commercial banks.

Table 2. Estimates for the contract terms equations in structural form.

	Eq 1. LTV				Eq 2. Log. of rate				Eq. 3. Log. of maturity				Eq 4. Probability of insurance			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
LTV	-	-	-	-	0.211***	0.215***	0.213***	0.101	0.409***	0.314**	0.311	-0.013	0.941***	0.931***	1.125***	1.142***
					(0.017)	(0.039)	(0.072)	(0.067)	(0.069)	(0.152)	(0.287)	(0.089)	(0.047)	(0.036)	(0.041)	(0.043)
Log. of rate	-0.041*	-0.053**	-0.158***	-0.029	-	-	-	-	0.117*	0.224**	0.228*	0.030	0.759***	0.192***	0.163***	0.180**
	(0.032)	(0.033)	(0.053)	(0.045)					(0.080)	(0.113)	(0.156)	(0.199)	(0.053)	(0.028)	(0.013)	(0.102)
Log. of maturity	0.052***	0.039***	0.034***	0.021***	0.001	0.010*	0.006	0.112***	-	-	-	-	-0.036***	-0.006*	-0.013***	0.120***
	(0.009)	(0.009)	(0.007)	(0.004)	(0.006)	(0.006)	(0.006)	(0.022)					(0.013)	(0.005)	(0.002)	(0.021)
Probability of insurance	0.339***	0.381***	0.342***	0.228***	0.154***	0.277***	0.136**	-0.032	-0.096***	0.333	0.071	-0.111	-	-	-	-
	(0.010)	(0.014)	(0.157)	(0.069)	(0.008)	(0.082)	(0.096)	(0.154)	(0.038)	(0.580)	(0.471)	(0.332)				
Log. of loan limit	0.118**	-0.017	-0.083	0.122***	-0.022	-0.118**	-0.158**	-0.025***	-0.283**	-0.196*	-0.312*	0.181***	0.081	0.171**	0.227***	0.001
	(0.056)	(0.072)	(0.084)	(0.006)	(0.047)	(0.065)	(0.073)	(0.007)	(0.159)	(0.124)	(0.190)	(0.031)	(0.112)	(0.058)	(0.018)	(0.006)
Prop. score	-0.123***	-0.184***	-0.214***	-	0.595***	0.587***	0.541***	-	0.154***	0.131**	0.112*	-	0.121*	0.222**	0.180*	-
	(0.012)	(0.066)	(0.086)		(0.033)	(0.055)	(0.114)		(0.032)	(0.562)	(0.981)		(0.102)	(0.093)	(0.113)	
k	24	60	132	49	24	60	132	49	24	60	132	49	24	60	132	49
N	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019	2019

Note: In the table cells there are mean marginal effects of changing of dependent variable on a change of another endogenous variable and selection propensity score. Bootstrap standard errors for 100 replications clustered on the month of application are in the parenthesis.
Significance level obtained from bootstrap distribution,
* - 10%, ** - 5%, *** - 1%.
k – number of estimated parameters, N – number of observations.
For each equation, model (1) was estimated for $\xi_1 = 1$, $\xi_2 = 1$, model (2) for $\xi_1 = 2$, $\xi_2 = 2$, model (3) for $\xi_1 = 3$, $\xi_2 = 3$, model (4) for $\xi_1 = 3$, $\xi_2 = 0$.

5.2.Probability of Default Model

The coefficients estimates from the probit-model *PD* and bivariate probit-model *PD* with correction for sample selection bias are presented in Tables 3. In order to avoid multicollinearity problem due to the high correlation of «credit duration» with the unemployment rate and contract rate, we estimate two separate specifications (1)-(2). We control for additional variables that control for application time, type of creditor (regional operator AHML or other creditors), type of credit program in other specifications (3), (4), (6) and find that they demonstrate statistically significant influence on probability of mortgage default, but do not lead to increase the predictive power.

The correlation between default and approval unobservables is negative (see Table 3 specifications (5)-(6)) and tends to decrease when control for additional explanatory variables. On the contrary to (Ross, 2000) the correlation is not statistically significant and the magnitude of the bias is not substantial. For not declared activity category, entrepreneur and unverified DTI the absolute bias is the highest and ranges from 0.115 to 0.146. However the bias does not exceed standard errors of parameters and the sign of the coefficients do not change. In the probit- model for the probability of approval positive statistical significant effects demonstrates dummy for the complete higher education of applicant and refinancing rate. The probability of approval decreases for applicants with unverified income, activity category, mortgage program, the region and applicants from regional operator of AHML. The last result is mainly explained by quarterly limits of regional operators of AHML on the volume of refinanced mortgage loans.

Most of the coefficient estimates in the probability of default models have the opposite sign to the ones in the probability of approval model that similar to (Ross, 2000) and signal to the validity of current credit underwriting techniques of creditors. In contrast to results in (Ross, 2000), this finding is valid not only with selection bias correction, but without it.

Results show that probability of mortgage default does not increase for mortgage contracts with unverified *DTI* ratio. This finding is mainly explained by the difference in the level of verified income and real income of borrowers and the large share of such borrowers in the dataset. In MILAN-analysis of Moody's (2009) «penalty» for this category of loans is used that depends on LTV ratio. According to (Bajari et al., 2008; Ross, 2000) the influence of income-related factors may be not robust and have joint effect with FICO score. An important feature of the data is a lack of credit history.

Table 3. Estimated parameters for probability of mortgage default equation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Bias
Male	0.331***	0.387***	0.401***	0.400***	0.391***	0.400***	0.382***	-0.004
	(0.111)	(0.117)	(0.117)	(0.117)	(0.117)	(0.117)	(0.117)	
Family status: Not declared	0.634**	0.761***	0.869***	0.859***	0.789***	0.860***	0.782***	-0.028
	(0.261)	(0.255)	(0.258)	(0.257)	(0.256)	(0.257)	(0.253)	
Family status: Single	0.281**	0.337**	0.342**	0.338**	0.345***	0.339**	0.343**	-0.008
	(0.126)	(0.133)	(0.134)	(0.135)	(0.133)	(0.135)	(0.133)	
LTV<0.5	0.128	0.141	0.107	0.106	0.174	0.110	0.168	-0.033
	(0.113)	(0.130)	(0.131)	(0.132)	(0.134)	(0.135)	(0.134)	
LTV>0.7	0.201	-0.014	-0.106	-0.113	-0.047	-0.114	-0.035	0.033
	(0.142)	(0.134)	(0.133)	(0.135)	(0.135)	(0.135)	(0.137)	
Contract rate		0.387***	0.382***	0.383***	0.390***	0.384***	0.382***	-0.003
		(0.042)	(0.042)	(0.041)	(0.042)	(0.041)	(0.042)	
Maturity <10 years	0.708**	0.570*	0.437	0.462	0.532	0.460	0.557	0.038
	(0.320)	(0.342)	(0.346)	(0.353)	(0.342)	(0.352)	(0.348)	
Maturity 10-14.9 years	0.550*	0.495*	0.374	0.384	0.459	0.382	0.483*	0.036
	(0.281)	(0.291)	(0.294)	(0.300)	(0.290)	(0.299)	(0.293)	
Maturity 15-19.9 years	0.191	0.422	0.339	0.353	0.375	0.350	0.381	0.047
	(0.270)	(0.272)	(0.272)	(0.278)	(0.272)	(0.276)	(0.276)	
Maturity 20-24.9 years	0.264	0.363	0.317	0.335	0.329	0.333	0.343	0.034
	(0.286)	(0.290)	(0.287)	(0.293)	(0.287)	(0.291)	(0.293)	
DTI is unverified	-0.464***	-0.291**	-0.238	-0.229	-0.437***	-0.243	-0.378**	0.146
	(0.132)	(0.147)	(0.150)	(0.153)	(0.169)	(0.183)	(0.173)	
DTI [0;0.2)	0.114	0.131	0.213	0.201	0.122	0.199	0.161	0.009
	(0.300)	(0.329)	(0.336)	(0.336)	(0.332)	(0.338)	(0.334)	
DTI [0.4-0.6)	0.091	-0.001	0.001	-0.000	0.005	0.000	0.004	-0.006
	(0.129)	(0.137)	(0.136)	(0.138)	(0.138)	(0.138)	(0.138)	
DTI [0.6-0.8)	0.317*	0.236	0.284	0.277	0.278	0.281	0.304*	-0.042
	(0.172)	(0.176)	(0.179)	(0.180)	(0.180)	(0.183)	(0.181)	
DTI [0.8-1]	0.138	0.103	0.053	0.058	0.151	0.063	0.169	-0.048
	(0.262)	(0.297)	(0.296)	(0.299)	(0.304)	(0.303)	(0.303)	
Unemployment rate		0.301***		0.052	0.245***	0.049	0.032	0.056
		(0.063)		(0.105)	(0.076)	(0.108)	(0.136)	
Mean m2 value	0.000**	0.000***	0.000***	0.000**	0.000***	0.000**	0.000	0
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
«Credit duration»	0.002***						0.033**	
	(0.000)						(0.017)	
Application time 2008-2009 years			0.660***					
			(0.203)					
Regional operator AHML			0.471***	0.448***		0.443***		
			(0.121)	(0.124)		(0.128)		
Special mortgage program				-1.000***		-0.998***		
				(0.343)		(0.346)		
Fitted prob. of approval					-0.229	-0.020	-0.236	
					(0.166)	(0.173)	(0.168)	
Constant	-8.4***	-13.9***	-10.3***	-10.4***	-12.78***	-10.3***	-10.7***	
	(1.108)	(1.612)	(1.36)	(2.02)	(1.83)	(2.10)	(2.17)	
N	2799	2799	2799	2799	2799	2799	2799	
Pseudo R^2	0.346	0.432	0.440	0.447	0.433	0.447	0.438	
AIC	878.5	771.9	763.0	757.0	772.1	759.0	768.5	
BIC	1038.8	938.1	935.2	935.1	944.2	943.1	946.6	
log likelihood	-412.3	-358.0	-352.5	-348.5	-357.0	-348.5	-354.2	
% Correct predictions	93.86	94.68	94.61	94.75	94.75	94.57	94.75	
AUC	0.9129	0.9406	0.9432	0.9455	0.9455	0.9412	0.9417	
Gini coefficient	0.8258	0.8812	0.8864	0.891	0.891	0.8824	0.8834	

Note: Robust standard errors in the parenthesis. Only statistically significant and main contract variables remained. We also controlled for other sociodemographic characteristics such as age of main borrower, education level, category of activity. Base categories: female, married, maturity>=25 years, LTV - 0,5-0,7, DTI - [0.2-0.4), non-special credit program.

***, **, * — significance level 1, 5 and 10% correspondingly. Fitted probability of approval is calculated based on selection equation – the probit-model for probability of endorsement in Tab. A4. Bias is the coefficient difference between probit- (2) and probit- with correction for sample selection bias (5). $Gini\ coefficient = (AUC - 0,5) \times 2$.

We find that for males, single borrowers or borrowers with not declared family status the probability of default is higher. These findings correspond to previous studies (Radaev, Kuzina, 2011; Bajari et al., 2008; Ross, 2000) and are explained by worse payment discipline, short life span, and higher risks of divorce, illness and job loss for the above-mentioned categories of borrowers. Unlike (Radaev, Kuzina, 2011), borrower's age and age squared do not have a statistical influence on the probability of mortgage default.

Note that dummies for LTV at the origination do not have statistically significant effect on the probability of default in contrast to (Ambrose et al., 2005; Pennington-Cross, Ho, 2010; Quercia et al., 2007). This finding is unexpected but explained the special features of borrowers applied for AHML mortgage programs. Ross (2000) mentioned the unusual lower default rate for mortgage loans with LTV ratio above 0.6 connected with a small set of mortgage loans with low LTV. In addition, the probability of default is positively related with «credit age», but in contrast to (Bajari et al., 2008; Ross, 2000), we do not find evidence for «hump-shaped» profile. We also find that contract rate and short credit maturity reduces the PD while borrowers who refused to provide an income information and higher incomed borrowers performs better then low and middle incomed borrowers.

Finally, to approximate expected losses given default we have to develop a binary choice model for *PD* with high predictive power and «credit age» as an explanatory variable. Two specifications in Table 3 (1), (7) contain this variable. However, specification (7) demonstrates higher predictive power in terms of area under curve and percent of right predictions.

In order to control for possible endogeneity of credit terms in the probability of default equation we also compared the *PD* model corrected for endogeneity by including structural form residuals from contract terms equations. Unwise models (1-7) in order to control for both endogeneity and sample selection using a paper of Ozhegov (2015) we regress *PD* on LTV, interest rate, maturity and probability of insurance by smooth functions up to power 3. We found no evidence of control function significance that evidence no correlation between *PD* and choice of credit terms unobservables. For this reason, *LGD* results presented below are based on bivariate probit-model *PD* (7) in Table 3.

Bivariate probit-model *PD* (7) in Table 3 is slightly better at predicting *PD* (94.75% correct prediction) compared to the naïve prediction (94.06%) that is mainly explained by low default rate in the data set that is typical for residential mortgage lending. Of the 166 defaults, the model correctly predicted 39. Of the 2633 nondefaults, the model was right on 2613 of them. The number of misclassified cases is 147. The probability of a Type I error is 33.9%, but it is not a major concern. In this case, model would reject nondefaulted mortgage applicant and the creditor would meet the opportunity cost. Type II error predictions resulting in a mistaken credit underwriting decision to approve defaulted mortgage applicant only occurred 4.64% times. Based on the results of discrimination analysis in Table A5 we can conclude that the probability of Type II error increases with lower contract rate, downpayment, credit duration, regional property value and higher maturity, LTV ratio, and unemployment rate. This is an additional area for further work.

5.3.Loss Given Default

We approximated historical losses for 165[5] defaulted loans by using the above-described methodology. Descriptive statistics for mortgaged property for defaulted loans are presents in Table A6. Graphical illustration of *ELGD* distribution is presented in Fig. 3.

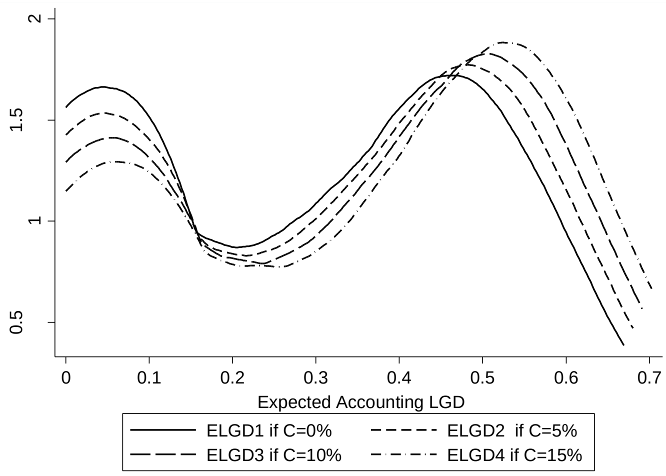

Figure 3. Empirical distribution of expected loss given default (*ELGD*)

For our dataset, the *ELGD* has hump-shaped distribution with concentration at values close to 0.1 and 0.5 with standard deviation equals to 0.2. These results are robust under different scenarios for workout process costs values. These results match the finding (Araten et al., 2004) that *ELGD* has a bimodal distribution. It is mainly explained by the heterogeneity of losses due to different LTV ratio of mortgage loans. As we can see in Fig. A2, LTV ratio is positively related with *ELGD*. The first and the second modes are related to low and high LTV ratio, correspondingly. This evidence found in previous empirical studies (Lekas et al., 1993; Pennington-Cross, 2003; Araten et al., 2004; Calem, LaCour-Little, 2004; Qi, Yang, 2009). Based on *ELGD* estimates we calculate absolute losses given default (*EL*) and expected losses for the credit portfolio (Table 4). As we can see in Fig. 4, *EL* has leptokurtic left-skewed distribution. Absolute losses at the credit portfolio level range from 116.8 mln Rus. rub to 140.4 mln Rus. rub (Table 4) under different process costs scenarios that can be used for efficient loan loss provisions calculation.

[5] 1 observation without information on net floor area of apartment was excluded.

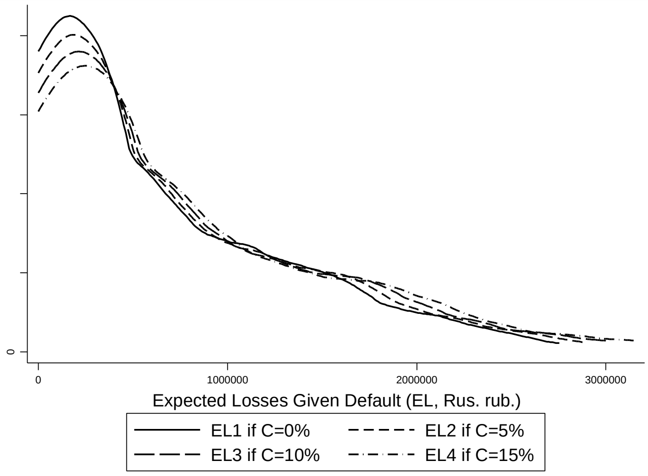

Figure 4. Empirical distribution of expected absolute losses given default (*EL*)

Table 4. Estimation of expected absolute losses

Variables	Description	All defaulted loans	Verified income	Unverified income	LTV								LTV ≤0.7	LTV >0.7	Regional operator AHML	Other creditors
					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				
Estimation of expected absolute losses at the portfolio level (mln Rus. rub.)																
ELp_1	C=0	116.8	110	6.8	0	0.06	0.61	13.2	3.59	52.8	11.4	35.5	70.3	46.5	93.1	23.7
ELp_2	C=5% of the current collateral value	124.2	117	7.2	12×10^{-6}	0.11	0.81	14.6	4.03	55.9	11.9	37.2	75.5	48.7	99.1	25.1
ELp_3	C=10% of the current collateral value	131.8	124	7.8	21×10^{-5}	0.17	1.07	16.2	4.50	59.0	12.4	38.8	80.9	50.9	105	26.8
ELp_4	C=15% of the current collateral value	140.4	132	8.4	12×10^{-4}	0.33	1.45	17.8	4.98	62.0	12.9	40.5	86.6	53.8	112.2	28.2
Estimation of expected absolute losses per mortgage contract (thou Rus. rub.)																
ELp_1	C=0	42	96	4	0	1	1	23	11	49	133	171	28	155	59	19
ELp_2	C=5% of the current collateral value	44	102	4	6×10^{-4}	2	2	25	12	51	138	179	30	162	63	20
ELp_3	C=10% of the current collateral value	44	108	5	0.6	3	2	28	14	54	144	187	32	170	67	22
ELp_4	C=15% of the current collateral value	47	115	5	0.1	7	3	31	15	57	150	195	35	179	71	21
Estimation of expected absolute losses per 1 mln Rus. rub. of loan amount (thou Rus. rub.)																
ELp_1	C=0	40	101	4	0	2	2	24	10	41	109	139	28	127	78	14
ELp_2	C=5% of the current collateral value	43	107	4	13×10^{-4}	4	3	27	12	43	113	146	30	133	83	15
ELp_3	C=10% of the current collateral value	45	114	4	0.02	5	4	30	13	45	118	152	32	139	88	16
ELp_4	C=15% of the current collateral value	48	121	5	0.1	11	5	33	14	48	123	159	34	147	94	16

Then we compare losses given default for different mortgage loans categories.

Firstly, 6% of total absolute losses given default for the credit portfolio consist of loans with unverified borrower's income (Table 4). They do not differ substantially from mortgage loans with verified income from the point of mean *ELGD* and its variance (Table A7, Fig. A3). Absolute losses for such loans are lower because of lower *EAD* (Table A7, Fig. A4) and they demonstrate lower absolute losses per mortgage contract (Table 4) and per 1 mln Rus. rub. of the loan amount.

Secondly, the results in Fig. A2 show that *LGD* is positively related with LTV at the origination. Defaulted mortgage loans with LTV>0.7 (the government-insured loans) demonstrate higher mean *ELGD*, *EAD* and *EL* (Table A7, Fig. A3-A4) with 2 times fewer variances of these parameters. In order to emphasize the difference in losses of mortgage loans with different LTV, we calculate absolute losses per mortgage contract and per 1 mln Rus. rub. of loan amount in Table 4. Despite the lower proportion of expected absolute losses for mortgages with higher LTV in the credit portfolio, these loans are characterized by higher absolute losses per mortgage contract and per 1 mln Rus. rub. of the loan amount that is approximately 5 times higher comparing with those that have LTV≤0.7. However, for these mortgage loans, AHML is used liability insurance of borrowers. It means that losses given default can be compensated by the insurance company. At the same time, loans with LTV>0.7 demonstrate higher expected interest income per mortgage contract and per 1 mln Rus. rub. of the loan amount (Table A8).

Additionally, we compare losses given default on loans provided by regional operator AHML and other creditors that operate AHML programs and refinance them to regional operator AHML. 80% of total absolute losses given default for the credit portfolio consist of loans provided by regional operator AHML (Table 4). Despite a slight difference in mean *ELGD*, *EAD* and *EL* (Table A8, Fig. A3-A4), mortgage contracts provided by regional operator AHML demonstrate 3 times greater absolute losses per contract and 5 times greater per 1 mln Rus. rub. of the loan amount.

Conclusion

This paper focus on demand, credit underwriting, and performance of mortgage the government-insured loans provided by the regional operator of Agency of Home Mortgage Lending (AHML) in Russia.

We found that presence of the government insurance is linked with commonly riskier credits such as credits with higher LTV, rate, and maturity. Despite the fact that these contracts are riskier they are better underwrote. We also found the negative correlation between the probability of default and approval, but the bias in probit estimates without correction for sample selection bias is not substantial. Under the lack of information on historical losses given default, we employ an approach to approximate them. Our results confirm that loss given default has a bimodal distribution that is consistent with previous literature.

Finally, we find the evidence for negative correlation of loss given default and loan-to value-ratio (LTV) at the origination, but not on the probability of mortgage default. We show that in defaulted loans with high LTV demonstrate not only higher absolute losses per originated loan and per 1 mln Rus. rub. of the loan amount, but also higher corresponding expected interest income. The same evidence is found for mortgage loans provided by the regional operator of AHML comparing with other creditors that operate AHML mortgage programs.

Appendices

Table A1. Descriptive statistics

Variables	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 coborrowers	Monthly declared coborrowers'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
Contract rate	Fixed contract rate,%	11.59	1.64	9.55	19
Type of contract rate	=1, fixed contract 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

Variables[6]	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

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

Table A1. Descriptive statistics (cont.)

Variables[7]	Total	%
Socio-demographic characteristics (4298 applications)		
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 coborrowers		
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 contract 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
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

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

Table A2. Aggregated mortgage and housing markets characteristics

Variable	Mean	St. dev.	Min	Max
Volume of issued mortgage in region, mln. $	23.0	14.1	2.9	54.8
Volume of issued mortgage in region, number	894.4	529.2	134	2112
Mean loan amount, $	28 814.2	6299.8	22 482.7	47 705.0
Median maturity, months	200.79	12.81	173	222.2
Median rate, %	12.97	0.80	12	14.3
Mean LTV	0.58	0.03	0.48	0.65
Mean DTI[8]	0.35	0.01	0.33	0.37
Mean ft^2 value, $	89.7	14.3	66.9	119.2
Affordability of housing coefficient[9]	0.287	0.055	0.215	0.389
Number of refinanced in AHML loans	129.1	83.7	30	310
Number of application to the bank	121.4	51.9	43	222

Table A3. Results of discrimination analysis

Variables	Approved applicants(3698)			Rejected applicants (600)		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Age of client	34	33	7.7	33	32	6.9
Borrower's income	30.4	24.8	26.2	42.2	37.3	26.1
Income of coborrowers	17.4	14.6	11.3	25.1	20.0	16.6
Unemployment rate	8.4	8.5	1.5	8.4	8.0	1.6
Regional property value	37.7	37.6	6.5	37.0	34.5	5.9
Refinancing rate	9.5	8.5	1.8	8.9	8.3	1.3
Variables	Defaulted borrowers (166)			Nondefaulted borrowers (2633)		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
Age of client	36	35	7.7	34	33	7.7
Borrower's income	33.5	21.8	43.8	30.2	25.0	23.2
Income of coborrowers	15.8	11.9	11.5	17.5	14.6	11.3
Credit limit	0.9	0.7	1.0	1.1	1.0	0.6
Loan amount	0.9	0.7	1.0	1.1	0.98	0.5
Contract rate	14.1	14	1.5	9.8	11	4.2
Maturity	13.9	15	4.5	15.9	15	5.2
Downpayment	0.8	0.5	1.1	0.9	0.7	0.7
Flat value	1.7	1.3	1.8	1.9	1.7	0.98
Monthly payment	12.9	9.1	14.2	12.6	11.4	6.7
LTV	0.56	0.59	0.22	0.57	0.59	0.16
DTI	0.45	0.42	0.17	0.45	0.41	0.18
«Credit duration»	44.8	45.9	4.0	27.9	31.0	13.8
Unemployment rate	9.3	8.9	0.9	8.6	8.6	1.5
Regional property value	44.3	49.9	7.4	37.4	37.4	6.3
Refinancing rate	10.7	11	2.2	9.7	8.8	1.9

Note: ***, **, * — significance level 1, 5 and 10% correspondingly. Variables with high discrimination power in italics. To test homogeneity of means, medians and variances t-test and ANOVA-test, Wilcoxon-Mann-Whitney test, Bartlett' test were applied. Borrower's income is calculated for 1349 approved and 31 rejected applicants and 147 defaulted and 1001 nondefaulted borrowers with declared borrowers income. Income of coborrowers is calculated for 558 approved and 16 rejected applicants and 64 defaulted and 436 nondefaulted borrowers with declared income of coborrowers.

[8] DTI – ratio between monthly payment and monthly income.

[9] Affordability coefficient reflects the ratio between an annual income of mean household and a value of mean flat.

Table A3. Results of discrimination analysis (cont.)

Variables	*p*-value	
	Pearson's chi-squared test	Fisher's exact test
Approved/Rejected applicants		
Gender	0.677	—
Category of employment	—	0.000***
Education level	0.000***	—
Marital status	—	0.508
Borrower's income (categorical variable)	—	0.000***
Income of coborrowers (categorical variable)	—	0.000***
Time of mortgage application	0.412	—
Type of creditor	0.017**	—
Region of mortgage loan	—	0.000***
Type of credit program	0.000***	—
Defaulted/Nondefaulted borrowers		
Gender	0.031**	—
Category of employment	—	0.000***
Education level	—	0.001***
Marital status	—	0.004***
Borrower's income (categorical variable)	0.000***	—
Income of coborrowers (categorical variable)	0.000***	—
Maturity (categorical variable)	—	0.000***
LTV (categorical variable)	0.000***	—
DTI (categorical variable)	0.000***	—
Type of contract rate	—	0.000***
Time of mortgage application	—	0.000***
Type of creditor	0.000***	
Region of mortgage loan	—	0.065*
Type of credit program	0.000***	—

Note: ***, **, * — significance level 1, 5 and 10% correspondingly. Variables with high discrimination power in italics.

Table A4. Results of instruments' relevance test

	(I)			(II)			(III)		
Equation	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
LTV	2.27	2.40	2.03	3.08	2.07	2.07	2.32	2.11	2.17
Log. of rate	157.4	92.0	51.4	184.9	93.0	52.0	238.0	99.7	51.31
Log. of maturity	4.70	2.53	2.08	5.88	3.22	2.12	6.16	2.98	2.18
Prob. of insurance	3.46	2.33	2.36	2.03	3.16	2.31	2.19	2.62	2.32
Log. of loan limit	16.84	5.57	3.40	14.36	5.94	3.31	16.83	5.16	3.21
10% critical values	1.42	1.36	1.35	1.42	1.36	1.35	1.42	1.36	1.35
5% critical values	1.58	1.49	1.46	1.58	1.49	1.46	1.58	1.49	1.46
1% critical values	1.89	1.75	1.71	1.89	1.75	1.71	1.89	1.75	1.71

Note: In the table cells there are conditional F-statistics of excluded instruments.
Critical values are provided.
For each equation, models (I) are calculated with market-level instruments fixed in the month of application, models (II) with market-level instruments fixed one month before the month of application, and models (III) for two months before the month of application.
For each equation, model (1) was estimated for $M = 1$, model (2) for $M = 2$, model (3) for $M = 3$.

Table A5. Results of discrimination analysis after *PD* estimation

Variables	True defaulted borrowers (39)	False defaulted borrowers (127)	*p*-value (*t*-test)
	Mean	Mean	
Age of client	37.9	35.9	0.1839
Borrower's income	48.5	29.3	0.1890
Income of coborrowers	13.1	16.3	0.3252
Credit limit	940.2	891	0.8516
Loan amount	940.2	889.4	0.8468
Contract rate	15.8	13.6	0.000***
Maturity	2	2.9	0.000***
Downpayment	1.4	0.6	0.0120**
Flat value	2.3	1.5	0.1004
Monthly payment	15.6	12.0	0.3477
LTV	0.41	0.61	0.000***
DTI	0.42	0.46	0.3353
«Credit duration»	45.3	43.1	0.0001***
Unemployment rate	8.9	9.4	0.000***
Regional property value	48.1	43.2	0.000***
Refinancing rate	11.2	10.6	0.1133

Note: ***, **, * — significance level 1, 5 and 10% correspondingly. Variables with high discrimination power in italics. Borrower's income are calculated for 32 true defaulted and 115 false defaulted with declared borrowers income. Income of coborrowers are calculated for 11 true defaulted and 53 false defaulted with declared income of coborrowers

Table A6. Descriptive statistics for collateral (defaulted contracts)

Variables	Mean	Std. Dev.	Min	Max
Assessed property value, mln Rus. rub.	1.71	1.80	0.33	15.3
Total square, sq. meters	55.15	43.85	22.4	390.5
Number of storeys	6.95	3.42	2	18
Regional market property value , thou. rus. rub. /1 sq. meter	44.28	7.40	28.80	51.30
Regional assessed property value ,thou. rus. rub. /1 sq. meter	30.11	13.58	7.11	70.84

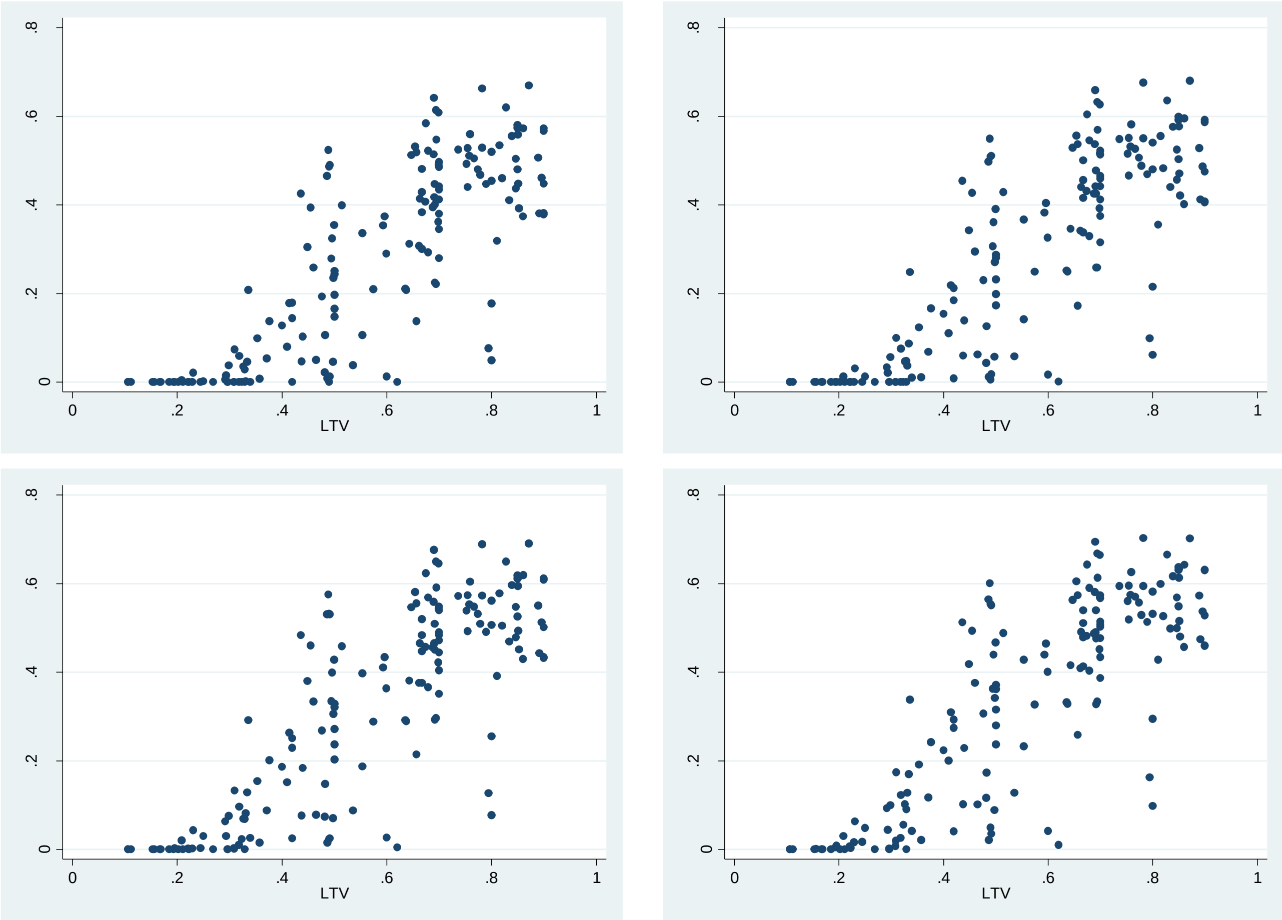

Figure A2. Dependence between *LGD* and LTV ratio at the origination

Table A7. Empirical estimation of credit risk parameters for defaulted loans of different categories

Variable	Description	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
		Verified income (146 defaulted loans)				LTV≤0.7 (121 defaulted loans)				Regional operator AHML (138 defaulted loans)			
ELGD 1	*C*=0	0.28	0.22	0	0.67	0.21	0.20	0	0.64	0.27	0.22	0	0.67
ELGD 2	*C*=5% of the current collateral value	0.31	0.22	0	0.68	0.23	0.21	0	0.66	0.29	0.22	0	0.68
ELGD 3	*C*=10% of the current collateral value	0.33	0.22	0	0.69	0.25	0.21	0	0.68	0.31	0.23	0	0.69
ELGD 4	*C*=15% of the current collateral value	0.35	0.23	0	0.70	0.28	0.22	0	0.69	0.34	0.23	0	0.70
EAD	Mean *EAD* (mln Rus. rub)	1.98	2.55	0.19	22.3	1.79	2.72	0.19	22.3	1.87	2.22	0.19	22.3
EL_1	*C*=0 (mln Rus. rub)	0.76	1.42	0	11.9	0.58	1.51	0	11.9	0.67	1.17	0	11.9
EL_2	*C*=5% of the current collateral value (mln Rus. rub)	0.80	1.47	0	12.4	0.62	1.57	0	12.4	0.72	1.22	0	12.4
EL_3	*C*=10% of the current collateral value (mln Rus. rub)	0.85	1.53	0	13.0	0.67	1.63	0	13.0	0.76	1.28	0	13.0
EL_4	*C*=15% of the current collateral value (mln Rus. rub)	0.90	1.59	0	13.5	0.72	1.69	0	13.5	0.81	1.33	0	13.5
		Unverified income (19 defaulted loans)				LTV>0.7 (44 defaulted loans)				Other creditors (27 defaulted loans)			
ELGD 1	*C*=0	0.22	0.21	0	0.56	0.47	0.13	0.05	0.67	0.31	0.20	0	0.64
ELGD 2	*C*=5% of the current collateral value	0.24	0.22	0	0.58	0.49	0.13	0.06	0.68	0.34	0.21	0	0.66
ELGD 3	*C*=10% of the current collateral value	0.27	0.22	0	0.60	0.51	0.12	0.08	0.69	0.37	0.21	0	0.68
ELGD 4	*C*=15% of the current collateral value	0.29	0.23	0	0.61	0.54	0.12	0.09	0.70	0.40	0.20	0.01	0.69
EAD	Mean *EAD* (mln Rus. rub)	1.24	0.86	0.31	3.49	2.18	1.29	0.48	5.73	2.06	3.33	0.21	17.7
EL_1	*C*=0 (mln Rus. rub)	0.36	0.52	0	1.94	1.07	0.66	0.02	2.75	0.90	2.06	0	10.8
EL_2	*C*=5% of the current collateral value (mln Rus. rub)	0.38	0.54	0	2.01	1.11	0.69	0.03	2.88	0.94	2.12	0	11.1
EL_3	*C*=10% of the current collateral value (mln Rus. rub)	0.41	0.56	0	2.07	1.16	0.71	0.04	3.02	0.99	2.18	0	11.4
EL_4	*C*=15% of the current collateral value (mln Rus. rub)	0.44	0.57	0	2.14	1.21	0.74	0.05	3.15	1.04	2.24	0.02	11.8

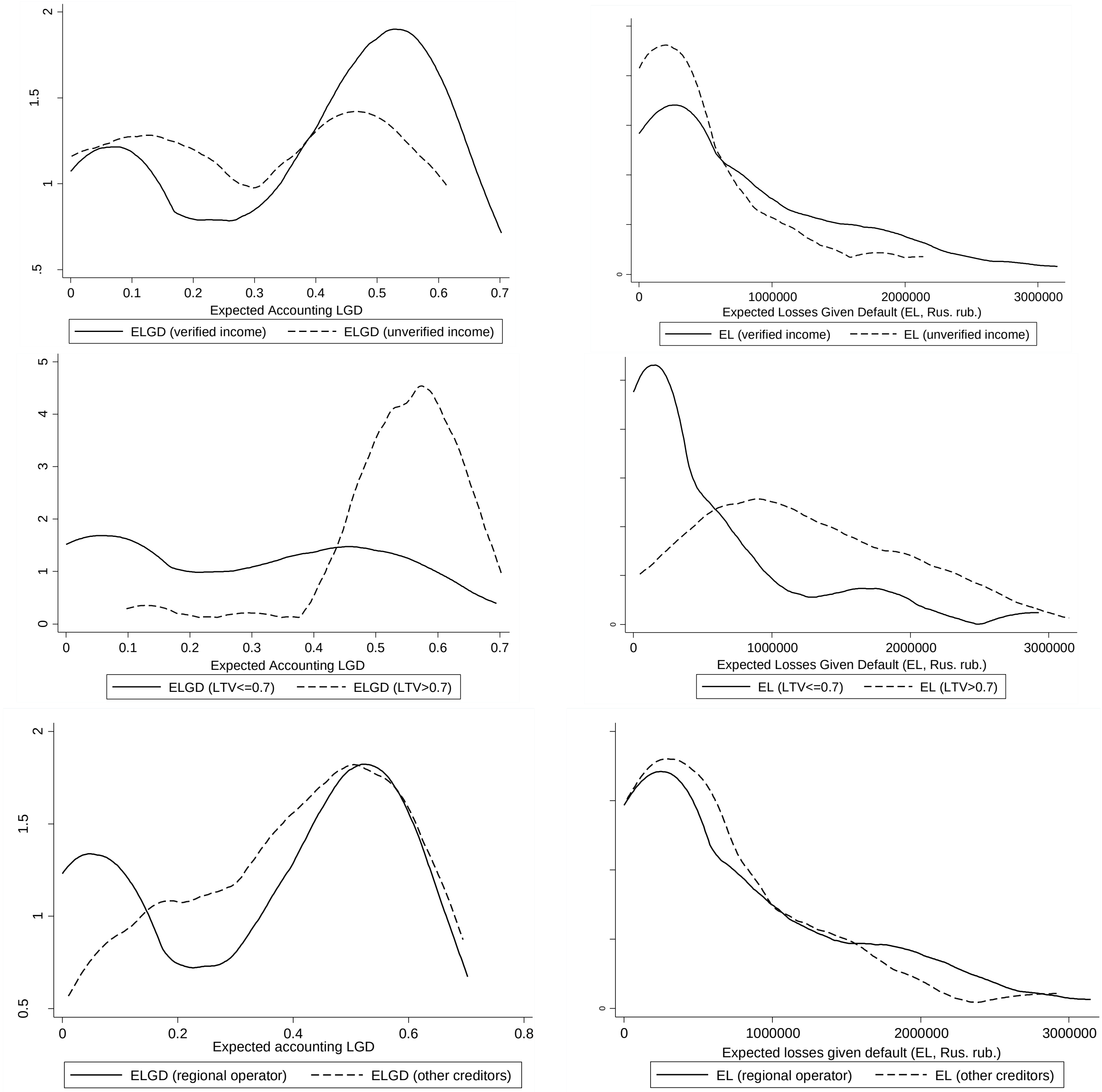

Figure A3. Empirical distributions of *LGD* and *EL* for defaulted loans of different categories

Note: Graphs show empirical density functions for accounting *LGD* and *EL* (less 5 mln Rus. rub.) for defaulted loans, when total costs equal C=15% of the current collateral value. Nonparametric smoothing is used – kernel smoothing with (kdensity) the Epanechnikov kernel. Results are robust for C=0, 5% and 10%.

Table A8. Expected interest income for defaulted loans of different categories

Variables	All signed contracts	Verified income	Unverified 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	LTV≤0.7	LTV>0.7	Regional operator AHML	Other creditors
Originated mortgage loans (mln Rus. rub.)	2910	1090	1820	1.2	9.1	31.4	307	547	347	1301	105	255	6.1	2544	366	1190	1720
Expected interest income at the portfolio level (mln Rus. rub.)	3480	1210	2270	1.6	3.8	28.5	288	579	439	1665.4	138	330	6.7	3005	475	1390	2090
Expected interest income per mortgage contract (mln Rus. rub.)	1.2	1.1	1.4	1.6	0.2	0.6	0.7	1.0	1.4	1.5	1.6	1.6	1.1	1.2	1.6	0.9	1.7
Expected interest income per 1 mln Rus. rub. of loan amount (mln Rus. rub.)	1.2	1.1	1.2	1.3	0.4	0.9	0.9	1.1	1.3	1.3	1.3	1.3	1.1	1.2	1.3	1.2	1.2

Note: Expected interest income at the portfolio level is calculated for all nondefaulted mortgage borrowers under assumption that they will be not defaulted during maturity. It is the difference in the sum of annuity payments and loan amount. Expected interest income per mortgage contract and per 1 mln Rus. rub. of loan amount is calculated as the ratio expected interest income at the portfolio level to number and volume of originated loans, correspondingly.

References

AHML: Agency of Home Mortgage Lending (2009). Godovoy Otchet [Annual Report] // AHML.

Ambrose B.W., LaCour-Little M., Huszar, Z.R. (2005). A note on hybrid mortgages, *Real Estate Economics* 33(4), 765-782.

Angrist J., Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton.

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

Archer W.R., Ling D.C., McGill G.A. (1996). The effect of income and collateral constraints on residential mortgage terminations, *Regional Science and Urban Economics* 26(3), 235-261.

Bajari P., Chu C.S., Park M. (2008). An empirical model of subprime mortgage default from 2000 to 2007, working paper, National Bureau of Economic Research 14625.

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., Dokko J., Shan H. (2010). The depth of negative equity and mortgage default decisions, Division of Research & Statistics and Monetary Affairs, Federal Reserve Board.

BIS: Basel Committee on Banking Supervision (2006). Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework (2006), Bank for International Settelments document.

Bellotti T., Crook J. (2012). Loss given default models incorporating macroeconomic variables for credit cards, *International Journal of Forecasting* 28(1), 171-182.

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

Clapp J.M., Goldberg G.M., Harding J.P., LaCour-Little M. (2001). Movers and shuckers: interdependent prepayment decisions, *Real Estate Economics* 29(3), 411-450.

Das, M., Newey, W. K., Vella, F. 2003. Nonparametric estimation of sample selection models, *The Review of Economic Studies* 70(1), 33-58.

Deng Y., Quigley J.M., Order R. (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options, *Econometrica* 68(2), 275-307.

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

Felsovalyi A., Hurt L. (1998). Measuring loss on Latin American defaulted bank loans: a 27-year study of 27 countries, *Journal of Lending and Credit Risk Management* 80, 41-46.

Foote C.L., Gerardi K., Willen, P.S. (2008). Negative equity and foreclosure: Theory and evidence, *Journal of Urban Economics* 64(2), 234-245.

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

Gupton G.M., Stein R.M. (2005). LossCalc V2: Dynamic Prediction of LGD Modeling Methodology, working paper, Moody's KMV.

Huang X., Oosterlee C.W. (2012). Generalized beta regression models for random loss-given-default, *Journal of Credit Risk* 4(7), 45-70.

Jackson J.R., Kaserman D.L. (1980). Default Risk on Home Mortgage Loans: A Test of Competing Hypotheses, *Journal of Risk Insurance* 47(4), 678-690.

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.

Lekkas V., Quigley J.M., Order R. (1993). Loan Loss Severity and Optimal Mortgage Default, *American Real Estate and Urban Economics Association Journal* 21(4), 353-371.

Leow M., Mues C. (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.

Moody's MILAN Methodology for Rating Russian RMBS (2008), Moody's.

Ozhegov, E.M. (2015). Identification in a class of nonparametric simultaneous equation models with sample selection, *Quantile* 13, 15-23.

Pavlov A.D. (2001). Competing risks of mortgage termination: who refinances, who moves, and who defaults? *The Journal of Real Estate Finance and Economics* 23(2), 185-211.

Pennington-Cross A. (2003). Subprime and prime mortgages: Loss distributions, working paper.

Pennington-Cross A., Ho G. (2010). The Termination of Subprime Hybrid and Fixed Rate Mortgages, *Real Estate Economics* 38(3), 399-426.

Phillips, R., Yezer, A. (1996). Self-Selection and Tests for Bias and Risk in Mortgage Lending: Can You Price the Mortgage If You Don't Know the Process? *Journal of Real Estate Research* 11, 87-102.

Qi M., Yang X. (2009). Loss given default of high loan-to-value residential mortgages, *Journal of Banking and Finance* 33(5), 788-799.

Qi M., Zhao X. (2011). Comparison of modeling methods for Loss Given Default, *Journal of Banking and Finance* 35(11), 2842-2855.

Quercia R.G., Stegman M.A., Davis W.R. (2007). The Impact of Predatory Loan Terms on Subprime Foreclosures: The Special Case of Prepayment Penalties and Balloon Payments, *Housing Policy Debate* 18(2), 311-346.

Rachlis, M., Yezer A. (1993). Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets, *Journal of Housing Research* 4, 315-336.

Radaev, V.V., Kuzina O.E. (2011). Monitoring of the financial behavior of the population. http://www.hse.ru/org/projects/47265530.

Ross, S.L. (2000). Mortgage Lending, Sample Selection and Default, *Real Estate Economics* 28(4), 581-621.

Sanderson, E., Windmeijer, F. (2014). A Weak Instruments F-test in linear IV models with multiple endogenous variables, Discussion paper 14/644, University of Bristol, Department of Economics.

Schuermann T. (2004). What do we know about Loss Given Default? working paper, No 04-01.

Sigrist F., Stahel W.A. (2011). Using the censored gamma distribution for modeling fractional response variables with an application to loss given default, *ASTIN Bulletin* 2(41), 673-710

Stock, J.H., Yogo M. (2005). Testing for weak instruments in linear IV regression. In: D.W.K. Andrews and J.H. Stock (Eds.), Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg, 80-108. New York: Cambridge University Press.

Vandell K.D. (1978). Default risk under alternative mortgage instruments, *The Journal of Finance* 33(5), 1279-1296.

Vandell K.D. (1995). How ruthless is mortgage default? A review and synthesis of the evidence, *Journal of Housing Research* 6(2), 245-264.

Yang B.H., Tkachenko M. (2012). Modeling exposure at default and loss given default: empirical approaches and technical implementation, *Journal of Credit Risk* 2(8), 81-102.

Yashkir O., Yashkir Y. (2013). Loss Given Default Modeling: Comparative Analysis, *Journal of Risk Model Validation* 7(1), 25-59.

Yezer, A., Philips, R., Trost R. 1994. Bias in Estimates of Discrimination and Default in Mortgage Lending: the Effects of Simultaneity and Self-Selection, *Journal of Real Estate Finance and Economics* 9, 197-215.

Zhang Y. (2013). Fair Lending Analysis of Mortgage Pricing: Does Underwriting Matter? *The Journal of Real Estate Finance and Economics* 46(1), 131-151.