ESTIMATION OF DEMAND FOR MORTGAGE LOANS USING LOAN-LEVEL DATA

Evgeniy M. Ozhegov

Higher School of Economics, Laboratory of Empirical Analysis of Industrial Organization, Perm, Russia

Abstract This report contains structural and econometric model for estimating the demand on mortgage loans. The demand for loan can be represented as two functions: probability of borrowing and the loan amount depending on borrower-specific characteristics, contract terms and set of macrovariables. The decision-making process on borrowing can be described as the sequence of decisions on: 1) choosing the credit program; 2) approving of a borrower; 3) choosing contract terms from feasible set; 4) and loan performance. Following Attanasio, Goldberg and Kyriazidou (2008) and Philips and Yezer (1996) the author proposes the econometric approach that deals with endogeneity and self-selection of borrowers when estimating the demand-for-loan equations and specifies the structure of data that is required for implementation of that approach.

Keywords: Demand for loans, sample selection, consistent estimates

JEL classification: C31, D12, D14, G21

Introduction

Demand for loan as well as for mortgage loan are the function of probability of credit contract agreement and functions of credit contract terms on characteristics of borrower, aim of crediting, expected loan performance and some macroeconomic variables.

Econometric estimation of parameters of that functions facing with inconsistency driven by endogeneity and sample selection. Endogeneity is generated by simultaneity in borrower and credit organization decisions on explanatory variables in demand equations. Sample selection arises when decision-making process of borrowing is made sequentially and some explanatory variables are observed partially in different stages of crediting.

However, this challenges in estimation process was avoided in recent papers that studied crediting process. 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 LTV decision), if and when to refinance or default (the termination decision), and the choice of mortgage instrument itself (the contract decision). Rachlis and Yezer (1993) then suggested a system of four simultaneous equations for mortgage lending analysis: (1) borrower's application, (2) borrower's selection of mortgage terms, (3) lender's endorsement, and (4) borrower's default.

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 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.

Previous models that tackled sample selection bias in lending analysis are not appropriate to estimate the loan amount or LTV ratio. The probit model of Ross (2000) and bivariate probit model used by Yezer, Phillips, and Trost (1994) and Phillips and Yezer (1996) are suitable for estimating a binary outcome. The following papers studied the dependence of the decision on loan amount as well as different endogenous variables on the exogenous ones.

Zhang (2010) investigated the sample selection bias and interaction between pricing and underwriting decisions using standard Heckman model. Bocian, Ernest, and Li (2008) used 3SLS for the simultaneous decisions on pricing and credit rating and found the empirical evidence that non-white borrowers are more likely to receive higher-priced subprime credits than similar white borrowers. Ambrose et al. (2004) constructed a simultaneous equation system of LTV and house value, which is used as a proxy for loan amount to account for endogeneity.

Other literature on mortgage choice has focused on the optimal mortgage contract given uncertainty about future house prices, household income, risk preferences, and, in some papers, mobility risk. Campbell and Cocco (2003) examine household choice between FRM and ARM in an environment with uncertain inflation, borrowing constraints, and income and mobility risk. They demonstrate that an ARM is generally attractive, but less so for a risk-averse household with a large mortgage, risky income, high default cost, or low probability of moving. Coulibaly and Li (2009) using survey data also found the evidence that more risk-averse, with risky income and low probability of future move borrowers prefers fixed rate mortgage contracts.

Forthowski, LaCour-Little, Rosenblatt and Yao (2011) studied the demand for mortgage loans from the point of choosing of adjusted rate mortgage versus fixed rate mortgage as a function on expected mobility. They find that, with all else equal, who self-select into ARM estimates their probability of moving in the future as relatively high.

Leece (2001) investigated the choice of ARM-FRM in the UK market dependent on the expected level of rates. Thus with sustainable low interest rates households intends to lock into fixed rate mortgage. In order to construct consistent and unbiased estimates he used linear additive model with time-dependent explanation variables.

Firestone et al. (2007) analyzed the prepayment behavior of low- and moderate-income borrowers. Main findings that non-white borrowers prepay more slowly than white ones. Results are stable during the time. The data contains the performance of 1.3 million loans originated from 1993 to 1997.

Courchane (2007) studied differences in pricing for different ethnicities after controlling of other pricing and underwriting parameters.

LaCour-Little (2007) was also focused on the question of choosing the credit program among low- and moderate-income borrowers. Using the loan-level from only one financial organization he founds that LMI borrowers are more likely to choose Federal Housing Administration insured mortgage programs and Special programs that assumed less down payments and higher score of expected risks due to high levels of current debt or weaker credit history. He also finds that nonprime loans preferred for those borrowers who are time limited to provide full documentation.

Some recent papers discussed the theoretical framework of optimal mortgage contraction. Thus, Nichols et al. (2005) showed that rejection rates vary directly with interest rates in the mortgage market and inversely in the personal loan market. The theoretical model in this paper demonstrates that the discrete levels of mortgage credit supply and the positive relationship between interest and rejection rates arise from a separating equilibrium in the mortgage market. This separation does rely on simple observation that processing an application through the underwriting process is costly, and is only partially covered by the application fee. When a subprime lender tries to locate too close (in credit risk space) to prime lenders, the application costs overwhelm credit losses to the point where it is less costly to lower credit standards and accept a higher proportion of applicants. Equilibrium requires that the subprime lender move a substantial distance from prime lenders, thus leading to a discrete and segmented mortgage market of those borrowers who may apply for prime mortgage and for those ones who are subprime.

Ghent (2011) discussed the dynamic demand for mortgage loans and steady state equilibrium for borrowers with hyperbolic compared to exponential discounting and the preference of such borrowers on the set of traditional fully amortizing mortgages and no-down-payment mortgages. The main findings of this paper that young households and retirees are more likely to choose NDP mortgage that arises when those households behave hyperbolically.

Piskorski and Tchistyi (2010, 2011) follow DeMarzo and Sannikov (2006) and pose the theoretical model of choosing the optimal mortgage contract that maximizes both lender's and borrower's combined surplus. As a result of this papers it provide a prediction of higher default rates for adjusted rate mortgages when the interest rate increases but shows that, nevertheless, ARM is optimal mechanism for mortgage contraction.

Karlan and Zinman (2009) found different method for solve the endogeneity problem when modeling the loan amount equation. They generated the truly random sample of credit proposals by sending letters with it to former borrowers. Using simple Heckman model they estimated the elasticities of demand for consumer credits to maturity and interest rate for different risk types of borrowers.

Attanazio, Goldberg and Kyriazidou (2008) introduced more progressive approach of managing the sample selection problem when modeling the empirical demand for loan equation. They studied the existence of credit constraints in different income segments. Using loan-level data of car loans they found that low-income households has positive elasticity of demand for car loans on the maturity and zero reaction of demand to interest rate change that means that those households are credit constraint. For doing that they used three-stage estimation methodology. At the first stage they estimated the participance equation. At the second stage the endogenous variables equations are being estimated by semi-parametric regression with correction for self-selection. Then endogenous variables in the demand equation was replaced by fitted values and the parameters was estimated also by semi-parametric regression. The only one motivation of using semiparametric regression is that the error terms of the loan amount, endogenous variables error terms and error term from the participation equation are correlated in non-linear way.

The main contribution of this paper for estimation of parameters of mortgage borrowing process is construction of structural and econometric model that gains consistent estimates of demand-for-loan function using loan-level individual data.

Structural and econometric model

Recent researches accumulated knowledge on the form of demand-for-mortgage function. It can be represented in linear way by following equation:

$$ln\, L = \beta_L D + \gamma_L C + \delta_L F + \psi_L P + \mu_L M + e_L \qquad (1)$$

where L is usually loan amount (or Loan-to-Value ratio), D are socio-demographic characteristics of borrower, C are the contract terms, F – parameters of the desired and bought property, P are contract performance characteristics, and M is macroeconomic variables or variables of financial market. All of them can be divided into endogenous (which are being choosing by borrower and credit organization respectively) and exogenous ones.

Table 1. Explanatory variables in demand equation

Variables	Endogenous for borrower	Endogenous for credit organization	Exogenous
Contract terms	Down payment; Maturity; Annual payment; Date of contract agreement; Program choice (ARM/FRM, Prime/nonprime, Conventional/Special/FHA programs); Self-selection for participation in mortgage.	Loan limit; Program parameters (minimum down payment, maximum maturity).	Program parameters (interest rate, insurance, Government Subsidied Enterprises); Cost of application.
Socio-Demographic characteristics	Number of co-borrowers; Aggregated income of co-borrowers; Aggregated expenses of co-borrowers; Income of borrower; Providing of full documentation.	Probability of creditworthness (FICO score of riskiness); Flag of endorsement.	Expenses of borrower; Age; Number of children; Marriage status; Level of education; Parameters of job; Nationality/Race; Expected mobility; Recent credit history.
Desired property	Value.		Specification of flat.
Loan performance	Month of first delinquency; Date of first delinquency; Flag of delinquency; Default, Refinancing, Prepayment.		Loss given default.
Macrovariables			Yield on Treasury notes; Refinancing rate; Volatility of interest rate; Unemployment rate; Volume of new construction.

Borrowing process can be represented by following sequence of decisions:

1. Application of borrower.

Future borrower realizes the neccesity of borrowing, chooses the credit organization and credit program that reflects her preferences, fills an application form with demographic characteristics.

2. Approval of borrower.

Considering application form and recent credit history, credit organization endorses the application or not, inquires the form data and set the limit loan amount when endorsed.

3. Choice of credit terms.

The approved borrower makes a choice on contract agreement and, when agreed, on property to buy and credit terms from feasible set: loan amount not more than limit, downpayment, annual payment and maturity determined by credit program.

4. Loan performance.

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

Econometric model repeats the steps of structural one:

1. Using instrumental variables for endogenous demographic characteristics:

$$D^{en} = Z_D \beta_D + e_D \quad (2)$$

where D^{en} is a vector of endogenous socio-demographic characteristics, Z_D are instrumental variables for demographics.

2. Modeling the probability of application:

$$y_1 = \begin{cases} 1, if\ D\beta_D^1 + M\beta_M^1 + e_1 \geq \alpha_1 \\ 0, if\ D\beta_D^1 + M\beta_M^1 + e_1 < \alpha_1 \end{cases} \quad (3)$$

where $y_1 = 1$ is an application decision, $D = (D^{ex}, \widehat{D}^{en})$ is a vector of exogenous demographics and fitted endogenous demographics, M – macrovariables.

3. Modeling the probability of endorsement for all applied:

$$y_2 = \begin{cases} 1, if\ D\beta_D^2 + M\beta_M^2 + E[e_2|y_1 = 1] \geq \alpha_2 \\ 0, if\ D\beta_D^2 + M\beta_M^2 + E[e_2|y_1 = 1] < \alpha_2 \end{cases} \quad (4)$$

where $y_2 = 1$ is an endorsement decision.

4. Choice of loan amount limit for all endorsed:

$$\bar{L} = D\beta_D^{\bar{L}} + M\beta_M^{\bar{L}} + E[e_{\bar{L}}|y_2 = 1] \quad (5)$$

where $\bar{L}$ is a decision on loan limit.

5. Modeling the probability of contract agreement:

$$y_3 = \begin{cases} 1, \text{если } D\beta_D^3 + M\beta_M^3 + \hat{L}\beta_{\hat{L}}^3 + E[e_3|y_2 = 1] \geq \alpha_3 \\ 0, \text{если } D\beta_D^3 + M\beta_M^3 + \hat{L}\beta_{\hat{L}}^3 + E[e_3|y_2 = 1] < \alpha_3 \end{cases} \quad (6)$$

where $y_3 = 1$ is an agreement decision; $\hat{L}$ is a fitted value of loan amount limit.

6. Choice of credit terms and property:

$$\begin{cases} \mathrm{C}_i = D\beta_D^{C_i} + M\beta_M^{C_i} + \mathrm{C}_{-i}\beta_{C_{-i}}^{C_i} + V\beta_V^{C_i} + F\beta_F^{C_i} + E\left[e_{C_i}\middle|y_3 = 1, C \in \bar{C}\right] \\ V = D\beta_D^V + M\beta_M^V + C\beta_C^V + F\beta_F^V + E[e_V|y_3 = 1, C \in \bar{C}] \end{cases} \quad (7)$$

where $C = (\mathrm{C}_i, \mathrm{C}_{-i})$ is a vector of contract terms (Loan amount, Maturity, Downpayment, Interest rate), $\bar{C}$ is a feasible set of contract terms determined by credit program, V is a property value, F is a property characteristics.

7. Modeling the probability of contract events and loss given credit event:

$$\begin{cases} y_4 = k, if \;\; D\beta_D^4 + M\beta_M^4 + \hat{C}\beta_C^4 + \hat{V}\beta_V^4 + U_k\beta_{U_k}^4 + E[e_4|y_3 = 1] \in \mathcal{L}_k \\ U_k = M\beta_M^4 + \hat{C}\beta_C^4 + \hat{V}\beta_V^4 + E\left[e_{U_k}\middle|y_4 = k\right] \end{cases} \quad (8)$$

where $y4$=k is a fact of k-th credit event, $\hat{C}$ and $\hat{V}$ are fitted values of credit terms and property value, U_k is a loss given k-th event.

Conclusion and discussion

Proposed model can perfectly manage endogeneity caused by simultaneity by instrumenting and fitting endogenous explanatory variables by multistage estimation procedure.

Inconsistency of estimates generated by sample selection will be released by introducing and estimation of bias term like $E[e|y = 1]$ in outcome equations. Effectiveness of this correction depends on accuracy of assumptions on distribution of error terms in selection equations. Thus, it is appropriate to use IMR in outcome equations when selection equation terms are normally distributed. More complicated assumptions on error term distributions are forcing to use semi-parametric methods for correction of sample selection bias. But this estimates will be less effective in terms of standard errors.

More debating point is about rationality of borrower and credit organization when decision-making. Sequential estimation procedures like multivariate probit or multistage Heckman procedure have no assumptions on rationality of agents. Using partially observed data in selection equations consider lack of borrower's ability to predict her and credit organization future decisions. Full rationality of agents assumes that borrower in every stage of decision-making process can predict outcomes on next stages and this prediction affects her present choice. Model of full rational borrowing process should contains fitted predictions on future outcomes as explanatory variables in all equations (2-8) and his equations should be estimated as system of simultaneous equations. And this case looks like very complicated for estimation because of discrete and continuous variables equations that are biased by sample selection.

References

Ambrose, B., LaCour-Little M., Sanders, A. 2004. The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Loan Level Analysis. Real Estate Economics, 32(4), pp. 541–569.

Attanasio, O.P., Goldberg, P.K., Kyriazidou, E. 2008. Credit Constraints in the Market for Consumer Durables: Evidence from Micro Data on Car Loans. International Economic Review, 49(2), pp. 401-436.

Bocian, D., Ernst, K., Li, W. 2008. Race, Ethnicity and Subprime Home Loan Pricing. Journal of Economics and Business, 60, pp. 110–124.

Campbell, J. Y., Cocco, J. 2003. Household Risk Management and Optimal Mortgage Choice. The Quarterly Journal of Economics, 118(4), pp. 1449–1494.

Campbell, J. Y., Cocco, J. 2011. A Model of Mortgage Default. NBER Working Paper Series.

Clapp, M. 2006. Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks. Real Estate Economics, 34(2), pp. 243-273.

Coulibaly, B., Li, G. 2009. Choice of Mortgage Contracts: Evidence from the Survey of Consumer Finance. Real Estate Economics, 37(4), pp. 659-673.

Courchane, M. 2007. The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain? Journal of Real Estate Research, 29, pp. 399-439.

DeMarzo, P. M., Sannikov, Y. 2006. A Continuous-Time Agency Model of Optimal Contracting and Capital Structure. Journal of Finance, 61, pp.2681–2724.

Firestone, S., Van Order, R., Zorn, P. 2007. The Performance of Low-Income and Minority Mortgages. Real Estate Economics, 35(4), pp. 479-504.

Follain, J. R. 1990. Mortgage choice. AREUEA Journal, 18(2), pp. 125–144.

Forthowski, E., LaCour-Little, M., Rosenblatt, E., Yao, V. 2011. Housing Tenure and Mortgage Choice. Journal of Real Estate Finance and Economics, 42, pp. 162-180.

Ghent, A., 2011. Subprime Mortgages, Mortgage Choice, and Hyperbolic Discounting, Working Paper, Zicklin School of Business.

Karlan, D., Zinman J. 2009. Observing Unobservables: Identifying Information Assymetries with a Consumer Credit Field Experiment. Econometrica, 77(6), pp.1993-2008.

LaCour-Little, M. 2007. The Home Purchase Mortgage Preferences of Low- and Moderate-Income Households, Real Estate Economics, 35(3), pp. 265–290.

Leece, D. 2001. Regressive Interest Rate Expectations and Mortgage Instrument Choice in the United Kingdom Housing Market, Real Estate Economics, 29(4), pp. 589-613.

Leece, D. 2006. Testing a Theoretical Model of Mortgage demand on UK Data. Applied Economics, 38, pp. 2037-2051.

Maddala, G., Trost R. 1982. On Measuring Discrimination in Loan Markets. Housing Finance Review, 1, pp. 245–268.

Nichols, J., Pennington-Cross, A., Yezer, A. 2005. Borrower Self-selection, Underwriting Costs, and a Subprime Mortgage Credit Supply, Journal of Real Estate Finance and Economics, 30(2), pp. 197-202.

Piskorski, T., Tchistyi, A. 2010. Optimal Mortgage Design. Review of Financial Studies, 23, pp. 3098—3140.

Piskorski, T., Tchistyi, A. 2011. Stochastic House Appreciation and Optimal Subprime Lending. Review of Financial Studies, 24, pp. 1407—1446.

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, pp. 87–102.

Philips, R., Trost R., Yezer, A. 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, pp. 197-215.

Rachlis, M., Yezer A. 1993. Serious Flaws in Statistical Tests for Discrimination in Mortgage Markets. Journal of Housing Research, 4, pp. 315–336.

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

Zhang, Y. 2010. Fair Lending Analysis of Mortgage Pricing: Does Underwriting Matter?, Office of the Comptroller of the Currency, OCC Economics Working Paper 2010)