BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

What can we learn about mortgage supply from online data?

by Agnese Carella, Federica Ciocchetta, Valentina Michelangeli
and Federico M. Signoretti

November 2020

Number 583

BANCA D'ITALIA

EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

What can we learn about mortgage supply from online data?

by Agnese Carella, Federica Ciocchetta, Valentina Michelangeli
and Federico M. Signoretti

Number 583 – November 2020

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of the Bank of Italy and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within the Bank of Italy, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6627 (print)
ISSN 1972-6643 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

WHAT CAN WE LEARN ABOUT MORTGAGE SUPPLY FROM ONLINE DATA?

by Agnese Carella, Federica Ciocchetta,
Valentina Michelangeli and Federico M. Signoretti*

Abstract

We exploit a novel dataset on mortgages offered by banks through Italy's main online mortgage broker, which works with banks representing over 80 per cent of mortgages granted, to gain an up-to-date assessment of loan supply conditions. Characteristics of mortgages are reported for about 85,000 borrower-contract profiles, constant over time, available at the beginning of each month starting from March 2018. We document that riskier applications, characterized by high loan-to-value ratios, low borrower's income and long maturity, are, on average, offered by a smaller number of banks that charge higher interest rates. Online banks tend to provide better price conditions than traditional intermediaries. We use the online rates offered to nowcast bank-level official (MIR) interest rate statistics, available only several weeks later. By relying on both regression analyses and machine learning algorithms, we show that the rates offered have significant predictive content for fixed-rate contracts, also after controlling for time-varying demand conditions, market reference rates, and unobserved time-invariant bank characteristics. Machine learning algorithms provide further improvements over regression models in out of sample predictions.

JEL Classification: G21, C81.
Keywords: mortgage, experimental data, risk-taking, nowcasting.
DOI: 10.32057/0.QEF.2020.583

* Bank of Italy, Economics, statistics and Research Directorate.

Contents

1. Introduction[1]

Can online data be used to inform policymakers on financial stability and monetary policy? To address this question we rely on a novel dataset on banks' mortgage offers, based on digital sources, which complements traditional data on lending rates such as the supervisory reports or the Credit Register (CR). In the past few years, a massive amount of online data has become available (Granello and Wheaton, 2004; Edelman, 2012), which has been used to nowcast (i.e. prediction of the very near future) and forecast economic indicators related to unemployment (Vicente *et al*., 2015; D'Amuri and Marcucci, 2017; Fondeur and Karamè, 2013), macroeconomic statistics (Ettredge *et al*., 2005), housing demand (Pangallo and Loberto, 2018) and local economic activity (Glaser *et al*., 2017). This paper builds on this literature and, to the best of our knowledge, presents the first application of online data to timely assess the evolution of mortgage supply by household risk and nowcast mortgage rates, crucial for both monetary policy and financial stability. In particular, our new dataset on mortgage offers does not suffer from bias coming from the demand side and constitutes an ideal setting to fully isolate mortgage supply changes, holding the demand constant. We show that the offered rates have significant predictive content for realized interest rates on fixed-rate contracts, also after controlling for time-varying demand conditions, market reference rates and unobserved time-invariant bank characteristics.

We exploit experimental data on mortgage offers from the major online mortgage broker in Italy, Mutuionline (MO).[2] Data are collected on a monthly basis for a large sample of Italian banks, accounting for over 80 per cent of residential real estate mortgages granted in 2018.[3] Upon submitting an application online, the prospective borrower has to specify several characteristics of the contract (mortgage type, interest type, amount, maturity), the borrower (age, income, job type), and the house (value, location). These elements concur to define a risk profile for the prospective borrower, according to which the broker selects and lists the banks willing to make an offer and the associated terms of the offered contract: the interest rate, both net and gross of fees, and the monthly instalment. The number of banks offering a contract and the contract terms may vary (and they typically do) for different profiles, as banks' willingness to grant a loan (through the broker) changes.

[1] We would like to thank Massimiliano Affinito, Emilia Bonaccorsi di Patti, Francesco Columba, Paolo Finaldi Russo, Marcin Kacperczyk, Silvia Magri, Ansgar Walther for their useful comments. We thank Enrico Sette for contributing to the initial design of the randomization of borrower characteristics. The opinions expressed are those of the authors and do not necessarily reflect those of the Bank of Italy or its staff. Please address correspondence to: agnese.carella@bancaditalia.it; federica.ciocchetta@bancaditalia.it; valentina.michelangeli@bancaditalia.it; federicomaria.signoretti@bancaditalia.it

[2] The underlying data have been provided free of charge by MutuiOnline.it for research purposes. They refer to fictitious customer profiles and contain no personal confidential information. Use of privately own data for research purposes does not imply the endorsement of the owner, its products or services.

[3] See Table A.1 in the Appendix for the list of banks associated with Mutuionline and Table A.2 for their main balance sheet variables.

Out of all the possible combinations of product characteristics, we have access to mortgage offers for about 85,000 fictitious profiles. Each profile is defined by a combination of: mortgage type (first home versus subrogation), interest-rate type (fixed versus adjustable), loan-to-value (LTV), maturity, applicant's age, income, job type, and house location. Our dataset is available at a monthly frequency from March 2018 onwards and within (approximately) the 10th day of each reference month, consistently with the timing of mortgage pricing decisions, which are typically taken by the banks at the beginning of the month. The 85,000 profiles are constant over time, while bank's willingness to supply a mortgage to a given risk profile and the associated contract terms do vary, reflecting changes in banks' loan supply policies. The contract terms offered by banks are binding, conditional on the accuracy of the information provided by the borrower.

From an econometrician's perspective, our dataset is an ideal setting for studying lending supply and, specifically, how this varies over time for different borrower-contract profiles. Indeed, as the borrower-contract profiles remain fixed over time, changes in the offers reflect *by construction* banks' supply choices. Differently from studies based on *realized* interest rates, we are immune from the associated typical endogeneity problems arising with borrowers' self-selection into specific banks or associated with discouraged borrowers that choose not to apply (see Michelangeli and Sette, 2016).

The dataset's features make it particularly attractive, both in terms of richness and frequency of the information provided, when compared with other data sources accessible for the Italian mortgage market. For Italy, Monetary Financial Institution (MIR) data provide monthly information at the bank level on interest rates by mortgage type, which is available about 30 days after the end of the reference period.[4] Data at the bank/borrower level are available monthly from the CR only for mortgage volumes and information on interest rates of new loans is available only quarterly from the Interest Rate Reporting (also called Taxia) for a sample of banks. Information on lending standards and loan terms and conditions is reported in the Bank lending Survey (BLS), carried out quarterly by the Bank of Italy. None of the above-mentioned traditional sources of data contains granular information on mortgage LTVs and maturity, nor on borrowers' income and other product characteristics.

We exploit the information from MO for three main purposes. First, we analyse the evolution of mortgage offered rates by risk profile over time. This is very important from a financial stability point of view, as it allows monitoring the banks' risk-taking behaviour in loan pricing. Second, we compare the characteristics of the contracts offered by online banks to those of traditional intermediaries. This

[4] MIR statistics include information on interest rates applied by monetary financial institutions to loans and deposits vis-á-vis households and non-financial corporations, both for new business and outstanding amounts, in the euro area. The financial institutions involved in the data collection are legally obliged to report monthly information to their National Central Banks, which in turn report to the ECB. The data are collected on a sample basis; in Italy the sample includes banks representing about 85 per cent of total outstanding loans and deposits to households and firms.

exercise provides some insights on the potential impact of digital technology (Fintech)[5] on the mortgage market and of its value for the final consumers. Third, we assess to what extent MO are useful to predict and nowcast MIR interest rates (which are available seven weeks after online data), using both regression analyses and machine learning algorithms. This exercise is particularly useful from a monetary policy perspective, as it may be used for a (very timely) assessment of the transmission of changes in policy interest rates to lending supply conditions.

Our main results are the following. For the risky profiles, characterized by higher LTV, longer maturity and lower borrower's income, the offered interest rates in our sample are systematically higher than for the other profiles; moreover, the number of banks offering a contract is substantially smaller. This evidence is suggestive of the fact that, on average, banks' correctly rank borrowers' riskiness in loan pricing decisions – at least in the period covered by the data. Moreover, in our sample we document that the reduction in mortgage interest rates was more pronounced for safer profiles (than for risky ones), which would not be consistent – *prima facie* – with banks engaging in excessive risk-taking.[6]

A comparison across bank categories indicates that online banks systematically and across risk profiles charge lower prices and offer contracts to a larger share of profiles than traditional intermediaries. For the banking groups that have both an online bank and a traditional brick-and-mortar one, we find that, on average, the online channel charges lower rates than the traditional one.

Finally, we study to what extent interest rates offered through the online platform can be used to predict official (MIR) rates and possibly nowcast them, obtaining an estimate of the realized rates about 50 days before these become available. The results show that MO offered rates do have predictive power on bank-level MIR data for fixed-rate contracts, after controlling for changes in the reference rate (the 10-year IRS), for a proxy for time-varying mortgage demand (lagged values of consumer confidence), and for the unobserved, constant over time, bank characteristics. While both standard regression techniques and machine learning algorithms have nowcasting power, the latter – in particular random forest algorithms – perform significantly better, based on standard measures of quality for out of sample predictions (root mean squared errors and share of correctly predicted cases). Machine learning techniques also allow determining which profiles are more relevant for predicting the final realized rate.

Our work relates to a recent strand of literature that exploits unconventional data sources to back traditional information and have a better understanding of the economic phenomena (Glaeser *et al.,*

[5] Fintech indicates the use of advanced technology in the provision of financial services.

[6] A fully-fledged assessment of excessive risk-taking would require more sophisticated analysis, which is beyond the scope of this paper.

2017; Glaeser *at al*., 2018). In particular, digital sources carry a huge potential in predicting economic outcomes of interest (e.g. Choi and Varian, 2012; Cavallo, 2012; Einav and Levin, 2014; Kang *et al.*, 2013, Wu and Brynjolfsson, 2015; Guzman and Stern, 2016; Pangallo and Loberto, 2018). Online platforms such as Yelp, Google, LinkedIn and immobilare.it provide researchers and policy-makers with crowdsourced data at the granular level, months before official statistics are available. Our paper builds also on the recent literature on machine learning, which has been used for default forecasting (Khandani *et al.,* 2010; Fuster *et al,* 2018; Albanesi and Vamossy, 2019, among others), for predicting distress in financial institutions (Chakraborty and Joseph, 2017), and for corporate default forecasting (Barboza *et al.,* 2017; Moscatelli *et al*., 2019).

The rest of the paper is organized as follows. In Section 2, we provide a detailed description of the data. In Section 3, we show some descriptive statistics and empirical evidence on the reliability of the data compared to MIR statistics and survey information from the BLS. In Section 4, we exploit MO data to assess credit supply conditions by borrowers' risk-profile. In Section 5, we describe how online banking differ from traditional banks in their supply of mortgages. In Section 6, we present a nowcasting exercise for MIR rates based on the MO rates (and some demand indicators). Section 7 concludes.

2. Mutuionline data

Mutuionline is an online mortgage broker that provides a quick outlook of the mortgage interest rates offered by affiliated banks, including all the major Italian banks, and puts the prospective borrower in touch with the bank making the preferred offer. In this section, we provide a description of the MO platform and of our experimental dataset.

2.1 Description of the Mutuionline platform

Figure 1 illustrates a screenshot of the main characteristics that the borrower has to specify in order to submit an application to the MO online platform: age, job type, income, mortgage type, rate type, house value, mortgage amount, and house location. The combination of all these characteristics defines a specific borrower profile. For each application, the platform shows the list of banks that are willing to grant the loan and the financial conditions they apply (Figure 2): net mortgage interest rate (hereinafter, named interest rate), annual percentage rate of charge (APR), monthly instalment, and commercial name of the contract. This is the pre- approval stage. Our data allow us to observe up to this stage of the process.

Next, the household selects the offer she prefers and proceeds to the stage in which she provides the broker with additional personal information (full name, date and municipality of birth, current address of residence, marital status, tax identification number, job position, etc.) and some other details on the property to buy (address, building type, building conditions, etc.). The broker forwards this information to the bank, which then reaches out to the borrower to finalize the contract. Finalization of the mortgage occurs at the bank's branch, or online if the bank does not have any physical branch. At this stage, for a given product, modification of the rate is not possible, unless information provided by the household is incorrect. In all other circumstances, banks active on the platform have a commitment to the broker not to modify the terms of the contract posted online during the mortgage settlement stages and, possibly due to reputational concerns, they have very low incentive not to respect such a commitment. In fact, according to MO reports, mortgage applications and mortgages actually concluded online exhibit strong similarities (Figure 3).[7] Moreover, products available on the platform are not specific to MO clients, but are the same products that are also available on the banks' website and therefore accessible to all.

2.2 Construction of our dataset

Our dataset includes the main characteristics of the applicant, contract, and house location, as well as the main contract terms offered by the banks upon accepting the application (Table 1). House value is fixed to 200,000 euros and mortgage amount varies with the LTV. Our data refer to the period from March 2018 (the starting point for our data) to August 2019.

By taking all the possible combinations of these characteristics, we obtain 85,000 fictitious profiles. We construct a matrix in which, for each profile and for each month, we record the banks that are willing to make an offer and the financial conditions offered (Figure 4).

In the rest of the analysis, we consider applications for first house purchase and renegotiations of the contract terms. We mostly focus on interest rates, distinguishing between fixed and adjustable rates. In addition, we report information on the share of banks that offer a product, for a given profile; the complement to 1 of this indicator – the "no offer rate" – is likely to be correlated with the probability that a bank denies credit to a given profile and thus is likely to be informative on actual loan rejection rates.

3. Descriptive analysis

In this section, we provide descriptive statistics on offered interest rates from MO and a comparison with official MIR rates and with information on changes in credit conditions based on the Bank Lending Survey (BLS).[8] We then show evidence on heterogeneity by borrower-contract

[7] In 2015 MutuiOnline intermediated about 2.5 billion euros of mortgages, which corresponds to about 6 per cent of the total amount of new loans for home purchase in Italy. Since then, this share is on the rise.

[8] The Bank Lending Survey is conducted quarterly by the NCBs in the Eurosystem and reports mainly on lending supply conditions. For more information and for data collected by Banca d'Italia for Italian banks, see https://www.bancaditalia.it/statistiche/tematiche/moneta-intermediari-finanza/intermediari-finanziari/indagine-credito-bancario/ index.html?com.dotmarketing.htmlpage.language=1

characteristics and by bank category. Figure 5 reports the distribution over time (average, median, 10th, 25th, 75th and 90th percentiles) of mortgage rates – both fixed and adjustable – offered by banks through MO. Fixed rates are generally higher than adjustable ones, with some heterogeneity across banks and profiles. From March 2018, rates have slightly decreased.

3.1. Interest rates: Mutuionline and MIR

A comparison between MO and official MIR rates is reported in Figure 6. MO banks account for about 85 per cent of the new mortgage originations, so that we can reasonably compare dynamics in our sample to those observed at the aggregate level. It should be noted, however, that MO and MIR rates might differ for a number of reasons and a direct comparison of the two series can be misleading. First, MIR rates are weighted by each bank's effective granted loan amounts, while in MO each profile is equally weighted (as data for granted volumes are not reported). Second, MO data are informative of the supply only, while MIR rates are the outcome of interacting supply and (time-varying) demand conditions.

The mean level of MIR rates is higher than the mean value in our sample (computed across all profiles), while the month-to-month change is very similar (Table 2). Moreover, for fixed-rate mortgages, the MIR-MO correlation is above 50 per cent; but negative and equal to -15 per cent for adjustable-rate mortgages. One potential explanation is that the distribution of granted adjustable-rate loans in our sample period – in which most of the loans are granted at fixed-rate – is skewed towards more risker borrowers; thus, the average MO rate is not representative of the actual adjustable rate.

3.2. Changes in banks' margins: Mutuionline and BLS

The BLS is a quarterly survey conducted on a sample of large Italian banks, which focuses on changes in the credit supply conditions. The BLS does not have questions on the banks' interest rates, which can be directly compared with MO, but it contains information on the margins. Specifically, we focus on the question in which banks are asked whether they increased, decreased, or kept stable margins on their mortgages over the past three months. The results are aggregated into an index, which is roughly a net percentage of banks reporting a tightening, weighted by the intensity of the reported change.[9] We compare the results from the BLS to those of the margins obtained using the MO dataset, which are calculated follows.[10] First, for all banks in the MO dataset we obtain a measure of the margin by subtracting the reference rate (10 year IRS for fixed rate mortgages, 3 month Euribor

[9] For details, please refer to the Note to Figure 7.

[10] MO does not contain data on mortgage margins; therefore, in our main analysis, we prefer to analyse directly mortgage rates. For the purpose of a comparison with BLS, we constructed margins for the MO dataset, making reasonable assumption on the reference rates.

for adjustable-rate mortgages) from the net mortgage rate. Second, we take the average margin across all banks in each month. Third, we compute the average over three months. Finally, we calculate the change in the quarterly margin. Figure 7 shows that the changes in margins in MO are very similar to the reported changes in the BLS. Even though the banks included in the Mutuionline sample differ from those in the BLS and even though Mutuionline data reflect real offers from the banks while BLS are more qualitative statistics, it is reassuring to find that the trends are very similar over time.

3.3 Mutuionline interest rates by borrower and contract characteristics

To identify the heterogeneity in pricing due to different levels of household or mortgage riskiness we breakdown mortgage rates by borrower and loan characteristics. This detailed information is a very valuable complement to traditional data, as these details are not available from other source. Panels A and B of Figure 8 show the evolution of fixed and adjustable interest rates, respectively, by main characteristics of either the mortgage (LTV and maturity, subfigures a and b) or the borrower (income, job type, subfigures c and d). Table A.3 in the Appendix reports additional summary statistics.

Riskier contracts, namely those with a high LTV (above 80 per cent), long maturity (30 years), or the fixed rate ones, are associated with higher average interest rates and higher dispersion (measured by standard deviation). Rates vary little by borrower characteristics (income level and job type).[11] Geographic characteristics, related to the province where the property is located, do not explain variation in mortgage prices, perhaps reflecting the fact that local differences in house prices, employment and economic growth are captured by other characteristics of the product (notably the LTV). The no-offer rate shows variability across both borrower and contract characteristics (see Table A.4 in the Appendix). In particular, riskier contracts have a higher probability of not being offered: for instance, the probability of not being offered exceeds 80 per cent for applications from borrowers without a permanent job and is above 90 per cent for loans with LTV greater than 80 per cent.

3.4 Mutuionline interest rates by bank category

Panels A and B of Figure 9 report the distribution of interest rates by banks' category: the largest five banking groups, other significant groups (subsidiaries of foreign banking groups included), less significant, online banks.[12] In order to account for the possibility that some intermediaries specialize in one mortgage type (fixed versus adjustable), which may imply different levels of risk, we analyse fixed- and adjustable-rate mortgages separately as in the previous section.

[11] Differences in interest rates by borrower characteristics are statistically significant.

[12] Most of online banks in the sample are online channels of traditional banks. See Table A.1 for the list of banking groups, online and traditional banks.

Overall, there is significant heterogeneity by banks' category (see also Table A.5 and Figure A.1 in the Appendix, for more detail and a comparison of the mean across groups). For fixed-rate mortgages, the top five largest groups offer the highest rates and show the highest dispersion; for adjustable-rate ones, the subsidiaries offer the highest rates with the largest dispersion.

4. Loan supply conditions by borrower-contract risk profile

From a financial stability perspective, mortgages characterized by a high LTV and a long maturity represent a significant source of risk[13] and should be carefully monitored. Differently from aggregate MIR data, MO data allow assessing mortgage pricing for different characteristics of the borrower and the contract and whether banks' willingness to engage in risky lending changes over time.

We construct two profiles of borrowers:[14]

1. Low risk profile: mortgage LTV equal to 60 per cent, mortgage maturity equal to 15 years and a permanent employment contract with net monthly income equal to 4000 euros;

2. High risk profile: mortgage LTV equal to 80 per cent, mortgage maturity equal to 20 years and a permanent employment contract with net monthly income equal to 2000 euros.

For both risk profiles, the age is fixed at 30 years and all mortgages are for first house purchase; furthermore, we distinguish between fixed and adjustable rates. Each profile is then analysed in terms of four indicators: average interest rate, month-to-month change of average interest rate, variation in terms of standard deviation and no-offer rate.

Figure 10 shows that the average interest rate associated with the high risk profile is significantly higher than the one associated with the low risk profile. The difference is equal, on average, to 20 basis points and is higher for fixed-rate mortgages (25 b.p., versus 14 b.p. for adjustable-rate ones). Over time, differences in the rates for the two profiles have become more pronounced, reaching 36 b.p. for fixed-rate mortgages and 18 b.p. for adjustable-rate ones in August 2019. These statistics also indicate that over the past year, the rates for both profiles have decreased, but the reduction for the low risk profile has been larger.

The month-to-month changes show that fixed rates have been on a decreasing trend in the last period, while adjustable-rate mortgages have been more stable over time. The only exception was

[13] See Gerlach-Kristen, 2018, Lydon and McCarthy, 2013, Magri, 2009, Whitley, Windram, and Cox, 2004 for a discussion of the main determinants on arrears on mortgages.

[14] We do not present results for the very risk profile (LTV above 80 per cent and fixed-term job) because very few banks are willing to accept these applications and, as a consequence, results are very sensitive to changes made by just one lender.

October 2018, when rates increased for fixed-rate mortgages and decreased for adjustable ones. The standard deviation is quite high for all profiles and mortgage types, and equal to, on average, about 30 basis points (or 20 per cent of the average mortgage rate).

Low risk profiles are widely offered by the banks, with an average no-offer rate of below 10 per cent. Moreover, after increasing slightly in January 2019, the rate decreased and flattened at around 0 per cent since April 2019, indicating that all the banks in the sample were willing to offer a mortgage to safe borrowers. For the high risk profile, instead, the number of offers is smaller: the no-offer rate is on average equal to 20 per cent, although also in this case it has been declining since January 2019.

5. Online banks versus traditional banks

Our novel dataset allows studying whether online banks adopt mortgage policies that differ from those chosen by traditional brick-and-mortar banks. This exercise may provide preliminary hints on the potential of digital technology and the benefits for the final consumer stemming from Fintech. Figure 11 shows that online banks and channels always offer, on average, lower mortgage rates and are characterized by lower no-offer rates, for both fixed- and adjustable-rate mortgages.

The difference between the interest rate offered by traditional and online banks is larger for fixed-rate mortgages than for adjustable-rate ones. Specifically, between March 2018 and August 2019 the average rate offered by online banks is about 17 p.p. lower for fixed-rate mortgages and about 9 p.p. lower for adjustable-rate ones than traditional banks. Online banks also exhibit a lower dispersion, on average equal to 0.36 p.p. for fixed-rate mortgage and 0.22 p.p. for adjustable-rate mortgages (vs 0.65 and 0.51 p.p. for traditional banks). This evidence could be consistent with the idea that the lower costs in the provision of financial services sustained by online banks – in connection with their use of advanced technologies – are at least to some extent passed-through to customers in terms of lower rates. It could also reflect superior skills by online banks to price risk.[15]

We also evaluate whether online banks differ from traditional banks in terms of loan rates and no-offer rates conditional on the riskiness of the borrowers. Previous results are confirmed, i.e. online banks offer the lowest possible price for given segments of clients based on their riskiness (in some cases the rate could be equal to that offered by traditional banks, Figure 12).

Finally, we evaluate whether online and traditional banks that belong to the same banking group offer different mortgage prices. In our sample, we have four groups with both online and traditional banks. On average, over the entire period, online banks charge lower rates than traditional banks that belong to the same banking group. However, as Table 3 shows, there is heterogeneity within risk

[15] A more thorough assessment of this issue would, however, require additional analysis.

profile, over time, and across banks. This implies that the banking group changes its mortgage policy across the different channels (online vs traditional) every single month, allowing also for changes across profiles. A closer look indicates that, for those profiles for which there is an offer from both traditional and online banks, the interest rate is quite similar within the banking group.[16]

6. Nowcasting mortgage rate for each bank

In this section we rely on both standard regression analyses and machine learning techniques to test whether MO data help predicting changes in (official) interest rates for the banks in our sample and their performance in terms of nowcasting. Among the machine learning techniques, we consider random forest (Breiman, 2001), a rather popular algorithm recently applied in economic context for forecasting (see Glaeser, 2017 and 2018). As shown in recent applications, random forest has a high predictive performance, generally better than other traditional approaches. Furthermore, by construction, the algorithm identifies the indicators with greater relevance in predicting the target variables and rank them in decreasing order of importance.

Table 4 shows results from the regressions of the month-to-month change in the MIR fixed interest rate (MIR delta fixed interest rate) on several covariates.[17] Columns 1 to 3 show results for standard OLS regressions; Column 4 presents those from the random forest algorithm. All model specifications include bank fixed effects. Following the literature on machine learning, we evaluate the quality of our estimation by looking at the out of sample Root Mean Square Error (RMSE), which measures the difference between the values predicted by the model and the observed ones. For regression analysis, we also consider the Adjusted R-squared, which measures how much of the total variance is explained by the regressors.

In Column 1 we regress month-to-month changes in the realized bank rates on monthly changes in the reference rate, i.e. the 10-year interest rate swap (IRS), and its lagged values. Only the coefficient for the contemporaneous change of the IRS is significant. The Adjusted R-squared is quite low and equal to 6 per cent. The out of sample RMSE is quite high, around 11 per cent. The number of correct predictions, measured as those within one standard deviation from the true value, are 66.66 per cent. Overall, the reference rate itself does not appear to be a satisfactory predictor of the change

[16] There are however several profiles for which we observe an offer only from the online banks; these profiles mostly include first home mortgages, from applicants with a fixed term contract, and characterized by low LTV.

[17] Results for variable mortgage rates are presented in Table A.6 in the Appendix. Overall, variable-rate mortgages display lower variability in the rates than fixed-rate mortgages. Moreover, the reference rate (3month Euribor) is about stable over the period and, thus, is not included in the regression. The coefficient for the MO average rate (computed across all profile) is not significant, but has the right sign. Machine learning algorithms generate about the same number of correct cases as the baseline OLS regression.

in bank rates.

In Column 2 we include also the change in the average rate (across all profiles) offered by each bank through MO and its lags. The coefficients of both contemporaneous and lagged values of this variable are statistically significant. The Adjusted R-squared increases to 31 per cent, the out of sample RMSE remains equal to 11 per cent, and the share of correct out of sample predictions increases to 72.22 per cent. Our variable based on the online dataset is very informative and improves significantly the ability to nowcast bank mortgage rates.

In Column 3 we account also for monthly (time-varying) controls for the demand, namely two lagged values of consumer confidence. These controls are not statistically significant and the Adjusted R-squared remains unchanged. The improvement associated with adding time-varying demand controls does not regard the in-sample estimation, but it regards the out of sample statistics. Indeed, the out of sample RMSE decreases and share of correct out of sample prediction increases to 83.33 per cent.

In Column 4 we use the random forest algorithm to predict the change in the realized interest rate. We account for the present and lagged changes in the interest rates associated with each of the 128 profiles obtained from MO, in addition to contemporaneous and lagged values of the reference rate and demand controls.[18] The algorithm assigns weights to all variables to get the best possible prediction. Results indicate that the use of machine learning techniques drives a further improvements in the out of sample RMSE, which decreases to 7.59 per cent (22 per cent lower than the out of sample RMSE resulting from the linear regression with the same set of regressors in Column 3). Moreover, the share of corrected predicted cases increases to about 89 per cent, almost 6 p.p. above the share estimated using linear regression techniques in Column 3. The random forest algorithm also provides valuable insights to identify which profiles are more important in driving the total variation. As shown in Figure 13, the variable that plays a larger role in explaining the change in realized interest rates is the contemporaneous change in the MO rates for the following profile: 80 per cent LTV, 10 year maturity, borrower’s age 40 years, monthly net income 4000 euros with a permanent job (Figure 12, panel A). The second and the third profiles in order of importance exhibit similar, low risk characteristics: 60 per cent LTV, with short maturity and employee with permanent contract. Overall, these statistics indicate that fixed-rate mortgages are typically chosen by safer borrowers.

Finally, we test whether profiles that emerge as more important for the interest rates variation are constant or change over time, which boils down to test whether banks change their pricing model over time. We limit the analysis to the four previous months. Figure 13, Panel B, indicates that the most

[18] The 128 profiles are possible combinations of LTV, maturity, age, income and job type and, therefore, they also capture different level of risk.

important profiles in driving the actual interest rates variation display low time variability, with the most common being those with high LTV, short maturity, granted to low risk borrowers.

7. Conclusion

This paper describes new data from the major online mortgage broker in Italy (Mutuionline) and shows how these data can be exploited for policy analysis. The dataset is extremely valuable as it provides very timely information and allows accounting for different sources of heterogeneity in the mortgage contract (house location, contract, and borrower characteristics). Our novel dataset allows to fully isolate the mortgage supply choices (conditional on demand characteristics) of the main banks in Italy and is a very useful complement to traditional data sources, such as supervisory reports and CR. These latter sources of data reflect an equilibrium between demand and supply, are available with some time lag and do not report granular and detailed information on the borrower and the mortgage contract.

Our analysis indicates that, on average, riskier contracts are characterized by higher mortgage rates and are generally offered by fewer intermediaries. Moreover, online banks tend to charge lower rates than traditional banks. Finally, we show that online data are informative on realized banks' rate, even controlling for demand and unobserved time-invariant bank characteristics, and are helpful for nowcasting them. In particular, random forest algorithms provide a better out-of-sample performance as compared to the linear regression model and can be used for identifying the profiles that are most important in driving the realized rates, which matter for assessing if there are changes in banks' lending strategies over time.

Our dataset can be exploited in many policy analyses, in addition to the one described in this paper. For instance, it can be used to evaluate how mortgage supply conditions vary in face of a big negative economic shock (such as Covid-19), changes in monetary policy or macroprudential policies directed towards some specific banks, merger and acquisitions. We aim at digging deeper on these issues in follow up analyses.

8. References

Albanesi, S., and D.F. Vamossy (2019), "Predicting Consumer Default: A Deep Learning Approach," *NBER Working Paper*, No. 26165.

Barboza, F., H. Kimura and E. Altman (2017), "Machine learning models and bankruptcy prediction", *Expert Systems with Applications: An International Journal*, Vol. 83(c): 405-417.

Breiman, L. (2001), "Random Forests," *Machine Learning*, Vol. 45: 5-32.

Cavallo, A. (2012), "Scraped Data and Sticky Prices," MIT Sloan Working Paper.

Chakraborty, C., and A. Joseph (2017), "Machine learning at central banks," *Bank of England Staff Working Paper*, No. 674.

Choi, H., and H. Varian (2012), "Predicting the Present with Google Trends," *Economic Record* Vol. 88:2-9.

D'Amuri, F., and J. Marcucci (2017), "The predictive power of Google searches in forecasting US unemployment," *International Journal of Forecasting*, Vol. 33(4): 801-816.

Einav, L., and J. Levin (2014), "The Data Revolution and Economic Analysis," *Innovation Policy and the Economy,* Vol. 14, https://doi.org/10.1086/674019.

Edelman, B. (2012), "Using internet data for economic research," *Journal of Economic Perspectives*, Vol. 26: 189-206.

Ettredge, M., J. Gerdes and G. Karuga (2005), "Using web-based search data to predict macroeconomic statistics," *Communications of the ACM*, Vol. 48: 87-92.

Fondeur, Y., and F. Karamé (2013), "Can Google data help predict French youth unemployment?" *Economic Modelling,* Vol. 30: 117-125.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2018), "Predictably unequal? the effects of machine learning on credit markets," *The Effects of Machine Learning on Credit Markets (November 6, 2018)*.

Gerlach-Kristen, P. and S. Lyons (2018), "Determinants of mortgage arrears in Europe: evidence from household microdata," *International Journal of Housing Policy*, Vol. 18(4): 545-567.

Glaeser, E.L., H. Kim and M. Luca (2017), “Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity,” *NBER Working Paper,* No. 24010.

Glaeser, E.L., S.D. Kominers, M. Luca, and N. Naik (2018), “Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life,” *Economic Inquiry*, Vol. 56(1).

Granello, D. H., and Joe E. Wheaton (2004), "Online data collection: Strategies for research," *Journal of Counseling & Development,* Vol. 82(4): 387-393.

Guzman, J. and S. Stern (2016), “Nowcasting and Placecasting Entrepreneurial Quality and Performance”. Working Paper.

Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). “Where not to eat? Improving public policy by predicting hygiene inspections using online reviews,” in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 1443-1448.

Khandani, A.E., A.J. Kim and A.W. Lo (2010), “Consumer credit-risk models via machine learning algorithms”, *Journal of Banking & Finance*, Vol. 34(1): 2767-2787.

Magri, S. (2009), Arrears on Mortgages: Differences Across Countries and Their Effect on the Pricing of the Loan. 22nd Australasian Finance and Banking Conference 2009. Available at SSRN: https://ssrn.com/abstract=1460700 or http://dx.doi.org/10.2139/ssrn.1460700

Lydon, R., and Y. McCarthy (2013), “What lies beneath? Understanding recent trends in Irish mortgage arrears,” *The Economic and Social Review*, Vol. 44(1): 117-150.

Michelangeli, V., and E. Sette (2016), “How does bank capital affect the supply of mortgages? Evidence from a randomized experiment,” *Bank of Italy Working Papers* (*Temi di Discussione*), No. 1051.

Moscatelli, M., S. Narizzano, F. Parlapiano and G. Viggiano (2019), “Corporate default forecasting with machine learning,” *Bank of Italy Working Papers* (*Temi di Discussione*), No. 1256.

Pangallo, M., and M. Loberto (2018), “Home is where the ad is: interest proxies housing demand,” EPJ Data science, Vol. 7(1): 47.

Vicente, M.R., A.J. López-Menéndez and R.Pérez (2015), "Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?" *Technological Forecasting and Social Change*, Elsevier, Vol. 92(c): 132-139.

Whitley, J., and R. Windram, and P. Cox (2004), "An Empirical Model of Household Arrears," *Bank of England Working Paper,* No. 214.

Wu, L., and E. Brynjolfsson (2015), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales", in *Economic Analysis of the Digital Economy*, eds. Avi Goldfarb, Shane M. Greenstein, and Catherine E. Tucker. Chicago: University of Chicago Press.

9. Tables and Figures

Table 1: Borrower and contract characteristics in Mutuionline

Characteristics of the borrower	
Age	30, 40 years old
Job type	self-employed, permanent job, fixed-term contract
Monthly net gross income	2000, 4000 euros
Characteristics of the loan	
Purpose of the contract	first home, subrogation
LTV	50, 60, 80, 85 per cent
Rate type	fixed, adjustable
Maturity	10, 15, 20, 30 years
Characteristics of the property	
Location	110 provinces
Characteristics of the bank offer	
Monthly instalment	
Interest rate and APR	
Contract name	

Source: Mutuionline

Table 2: Summary statistics for month-to-month change in mortgage rates

(*averages March 2018 - August 2019; percentage points*)

	Mean	Sd.	Min	Max
	Panel A. Fixed			
MIR – Aggregate	-0.01	-0.01	-0.12	0.09
MIR – Mutuionline sample	-0.01	0.04	-0.11	0.08
Mutuionline	-0.03	0.02	-0.14	0.06
	Panel B. Adjustable			
MIR – Aggregate	0.00	0.00	-0.07	0.07
MIR – Mutuionline sample	0.00	0.05	-0.08	0.07
Mutuionline	-0.01	-0.02	-0.14	0.03

Source: Mutuionline and MIR data.

Notes: "MIR-Aggregate" indicates MIR interest rates referring to all the banks in the MIR sample. "MIR – Mutuionline sample" refers to the bank-level MIR data, from supervisory reports, for the sample of banks that work with Mutuionline. In this case, the monthly rate is calculated as the average rate offered by those banks weighted by the amount of new originations in the month. "Mutuionline" refers to the simple average across all profiles of the rates offered by the banks that work with Mutuionline.

Table 3: Difference in interest rates of traditional versus online banks in the same banking group

(*averages March 2018 - August 2019; percentage points*)

Panel A. Low risk profile								
	Fixed				Adjustable			
	Group 1	*Group2*	*Group3*	*Group4*	*Group1*	*Group2*	*Group3*	*Group4*
Mar-18	0.00	0.00	-0.14	0.32	2.22	0.68	0.00	-0.20
Apr-18	0.00	0.00	0.04	0.35	2.27	0.63	0.00	-0.20
May-18	0.00	0.00	0.07	0.70	2.27	0.63	0.00	-0.20
Jun-18	0.00	0.00	0.02	0.66	2.27	0.63	0.00	-0.20
Jul-18	0.00	0.00	0.12	0.68	2.27	0.63	0.00	-0.20
Aug-18	0.00	0.00	0.48	0.74	2.27	0.63	0.00	0.05
Sep-18	0.00	0.00	0.45	0.80	2.27	0.63	0.00	0.05
Oct-18	0.00	0.00	0.48	-0.11	0.00	0.00	0.00	-0.10
Nov-18	0.00	0.00	0.36	-0.14	0.00	0.00	0.00	-0.10
Dec-18	0.00	0.00	0.38	-0.27	0.00	0.00	0.00	-0.10
Jan-19	0.00	0.00	1.57	0.10	0.00	0.00	0.00	1.00
Feb-19	0.00	0.00	0.10	-0.16	0.00	0.00	0.00	0.05
Mar-19	0.00	-0.01	0.25	-0.03	0.00	0.00	0.00	0.05
Apr-19	0.00	-0.01	-0.01	0.02	0.00	0.00	0.00	0.05
May-19	0.00	-0.01	0.17	-0.01	0.00	0.00	0.00	0.05
Jun-19	0.00	-0.01	0.12	-0.17	0.00	0.00	0.00	0.06
Jul-19	0.00	-0.01	0.99	-0.06	0.00	0.00	0.00	0.10
Aug-19	0.00	-0.01	0.91	-0.25	0.00	0.00	0.00	0.79
Panel B. High risk profile								
	Fixed				Adjustable			
	Group1	*Group2*	*Group3*	*Group4*	*Group1*	*Group2*	*Group3*	*Group4*
Mar-18	0.00	0.00	0.00	-0.14	2.07	0.83	-0.20	0.10
Apr-18	0.00	0.00	0.00	0.16	2.07	0.83	-0.20	0.10
May-18	0.00	0.00	0.00	0.18	2.07	0.83	-0.20	0.10
Jun-18	0.23	0.00	0.00	0.13	2.07	0.83	-0.20	0.23
Jul-18	0.23	0.00	0.00	0.22	2.07	0.83	-0.20	0.23
Aug-18	0.22	0.00	0.00	0.55	2.07	0.83	0.05	0.22
Sep-18	0.22	0.00	0.00	0.54	2.07	0.83	0.05	0.22
Oct-18	-0.05	0.00	1.90	0.57	0.00	0.00	0.00	-0.05
Nov-18	-0.05	0.00	1.95	0.46	0.00	0.00	0.00	-0.05
Dec-18	-0.05	0.00	1.90	0.48	0.00	0.00	0.00	-0.05
Jan-19	0.05	0.00	0.00	1.72	0.00	0.00	1.00	0.05
Feb-19	0.00	0.00	0.00	-0.01	0.00	0.00	-0.11	-0.10
Mar-19	0.00	0.00	-0.01	0.16	0.00	0.00	-0.11	-0.10
Apr-19	0.00	0.00	-0.01	-0.12	0.00	0.00	-0.11	-0.10
May-19	0.00	0.00	-0.01	0.08	0.00	0.00	-0.11	-0.10
Jun-19	0.00	0.00	-0.01	0.03	0.00	0.00	-0.10	-0.10
Jul-19	0.00	0.00	-0.01	0.89	0.00	0.00	-0.06	-0.10
Aug-19	0.00	0.00	-0.01	2.16	0.00	0.00	0.68	-0.10

Source: Mutuionline.

Notes: Low risk are profiles with mortgage LTV equal to 60 per cent, mortgage maturity equal to 15 years and a permanent employment contract with net monthly income equal to 4000 euros. High risk are profiles with mortgage LTV equal to 80 per cent, mortgage maturity equal to 20 years and a permanent employment contract with net monthly income equal to 2000 euros.

Table 4: Predicting interest rates using regression analysis and machine learning: Fixed-rate mortgages

	MIR delta fixed interest rate (b,t)			
	Only reference rate	Adding Mutuionline average	Adding controls for demand	Machine learning
	(1)	(2)	(3)	(4)
Delta IRS-10y (t)	0.203***	0.154	0.257***	
	(0.112)	(0.095)	(0.150)	
L.Delta IRS-10y (t)	0.174	0.113	0.107	
	(0.110)	(0.076)	(0.074)	
L2.Delta IRS-10y (t)	-0.089	0.031	0.068	
	(0.114)	(0.085)	(0.087)	
L3.Delta IRS-10y (t)	0.120	-0.141	0.023	
	(0.109)	(0.098)	(0.179)	
MO delta interest rate (b,t)		0.394***	0.388***	
		(0.073)	(0.074)	
L.MO delta interest rate (b,t)		0.107***	0.112***	
		(0.045)	(0.046)	
L2.MO delta interest rate (b,t)		0.197***	0.192***	
		(0.048)	(0.046)	
L3.MO delta interest rate (b,t)		0.253***	0.256***	
		(0.064)	(0.067)	
L.Consumer confidence			0.000	
			(0.002)	
L2.Consumer confidence			-0.009	
			(0.007)	
Bank FE	Y	Y	Y	Y
Observations	234	234	234	329
Adjusted R-squared	0.061	0.310	0.307	
Out of sample RMSE	0.1106	0.1109	0.0974	0.0759
Out of sample correctly predicted (%)	66.66	72.22	83.33	88.89

Notes: An observation is correctly predicted if its out of sample prediction is within one standard deviation of true value of the dependent variable. Standard errors clustered at bank level in parentheses in Columns 1-3.
b stands for banks and t for time (months). Ln indicates that the variable is lagged of n months.
*** p<0.1, ** p<0.05, * p<0.01

Figure 1: Mutuionline website: example of a mortgage application

Source: Mutuionline website.

Figure 2: Mutuionline website: example of acceptance of a mortgage application

Source: Mutuionline website.

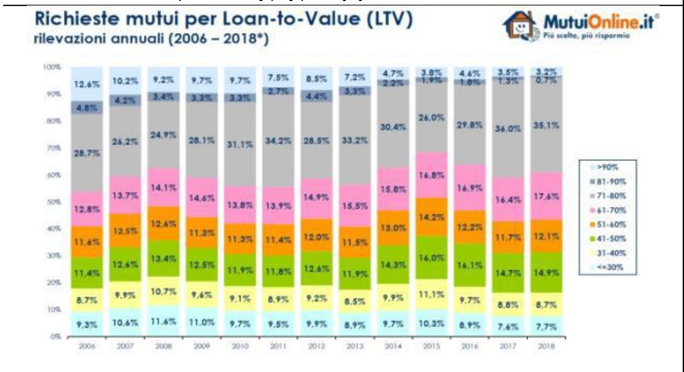
Figure 3: Comparison between mortgage applications and concluded contracts

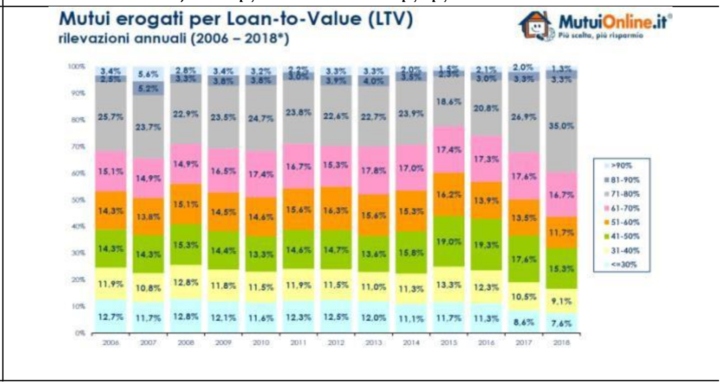
Panel A. Rate-type

a) Mortgage applications	b) Originated mortgages

Panel B. LTV

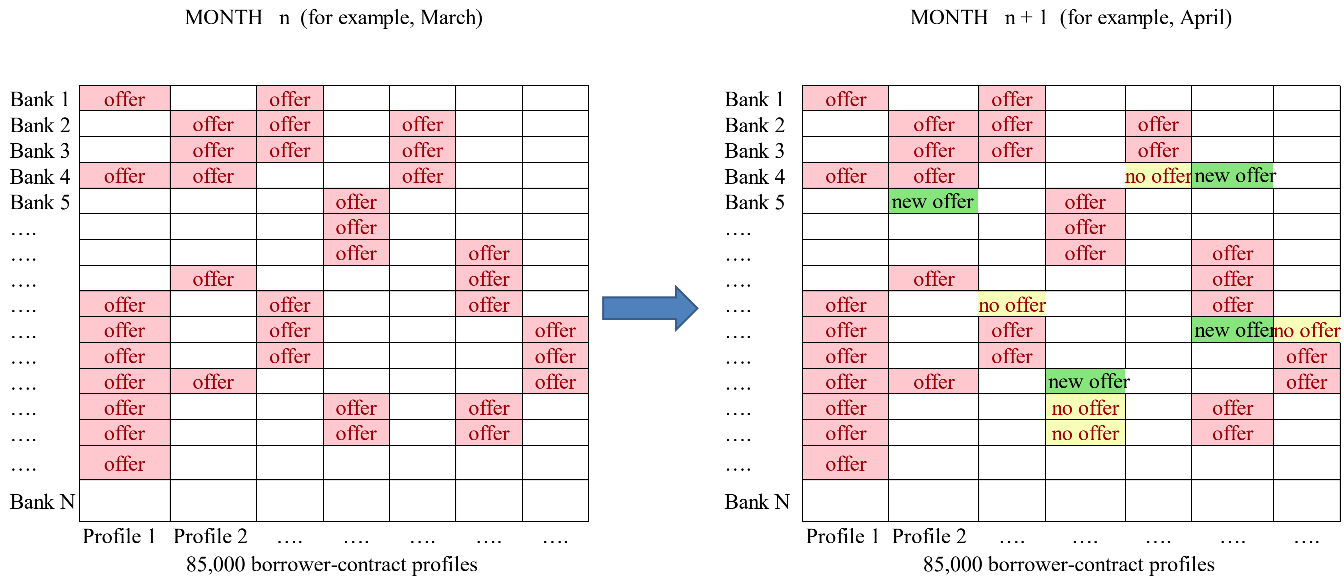
a) Mortgage applications	b) Originated mortgages





Source: Mutuionline data.

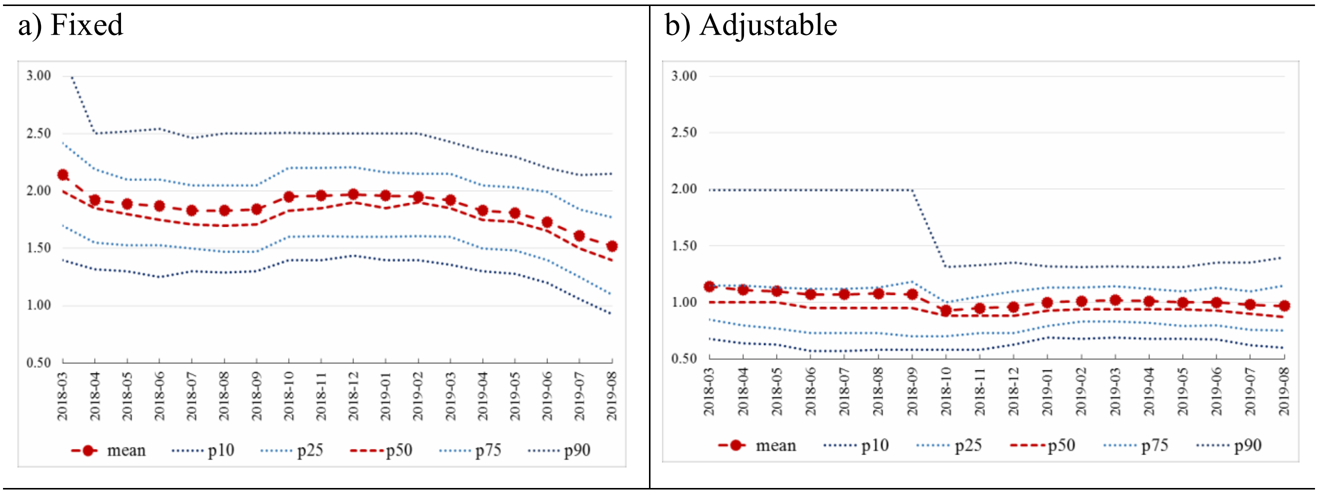
Figure 4: Dataset construction



Notes: Each profile is defined by a combination of: mortgage type (first home versus subrogation), interest-rate type (fixed versus adjustable), LTV, maturity, applicant's age, income, job type, and house location.

Figure 5: Mutuionline interest rates

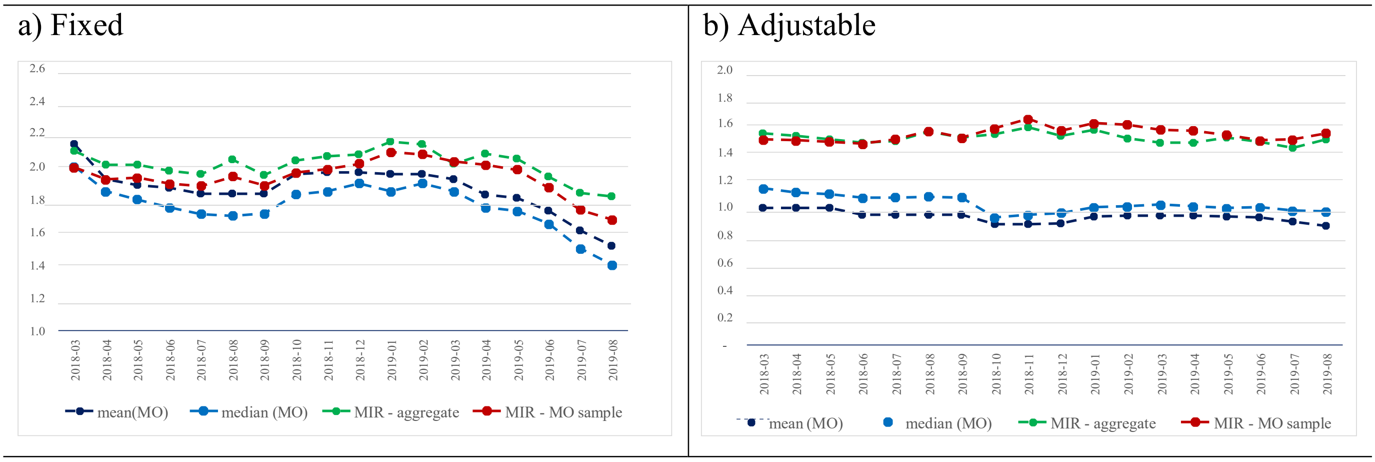
(*percentage points*)



Source: Mutuionline data.

Figure 6: Mutuionline interest rates and official MIR rates

(*percentage points*)

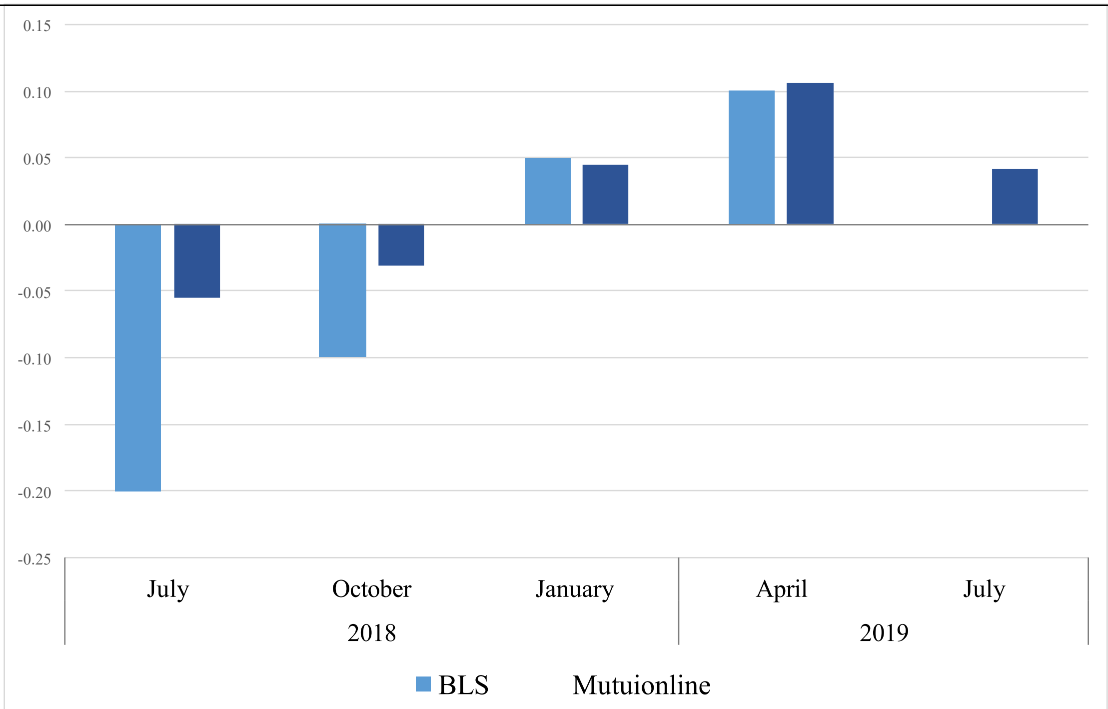


Source: Mutuionline and MIR data.

Notes: "MIR aggregate" indicates MIR interest rates referring to all the banks in the MIR sample. "MIR – MO sample" refers to the bank-level MIR data, from supervisory reports, for the sample of banks that work with Mutuionline. In this case, the monthly rate is calculated as the average rate offered by those banks weighted by the amount of new originations in the month. "MO" refers to the simple average across all profiles of the rates offered by the banks that work with Mutuionline.

Figure 7: Changes in banks' margins from BLS and Mutuionline data

(*diffusion index and percentage points*)



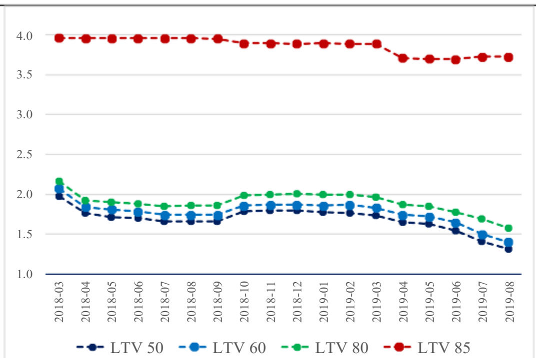
Source: Mutuionline and BLS data.

Notes: for BLS, diffusion index. Positive values indicate an increase in margins compared with the previous quarter. The diffusion index is constructed on the basis of the following weighting scheme: 1=increased considerably, 0.5= increased somewhat, 0=basically unchanged, -0.5= decreased somewhat, -1= decreased considerably. The range of variation of the index is from -1 to 1. For Mutuionline: percentage points; difference between MO net mortgage rate and the reference rate (10 year IRS for fixed rate mortgages, 3 month Euribor for adjustable-rate mortgages); 3-month average of monthly averages across banks.

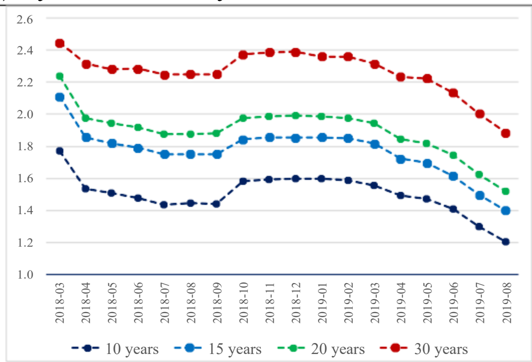
Figure 8: Mutuionline interest rates by loan and borrower's characteristics
(percentage points)

Panel A. Fixed-rate mortgages

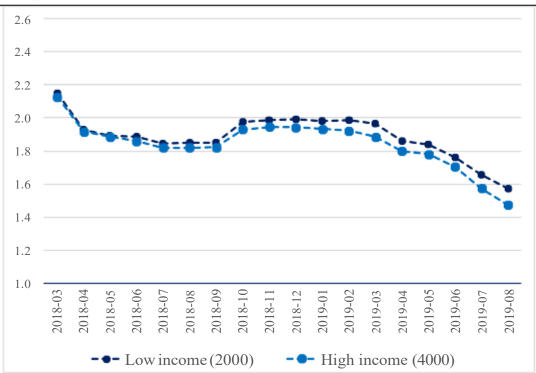
a) by LTV



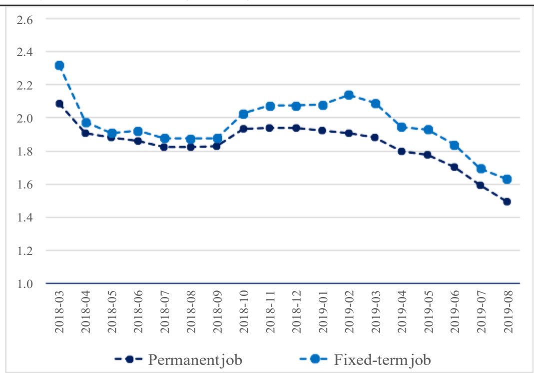
b) by loan maturity



c) by income

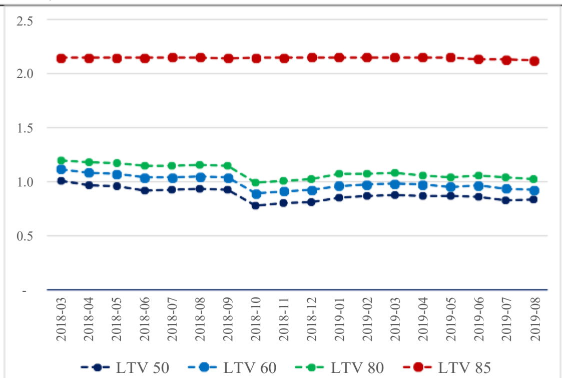


d) by borrower job type

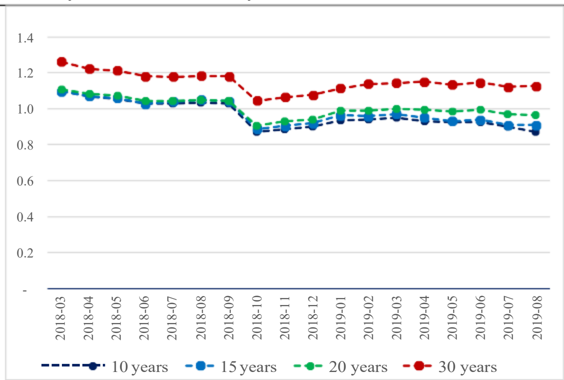


Panel B. Adjustable-ratemortgages

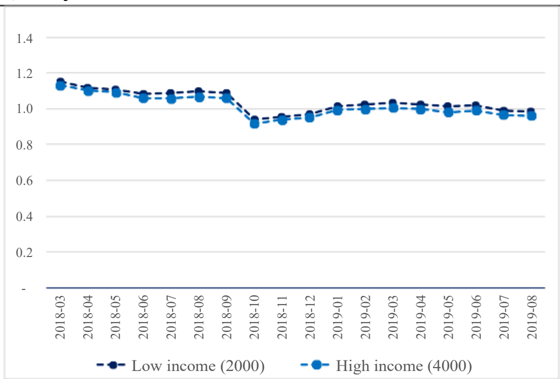
a) by LTV



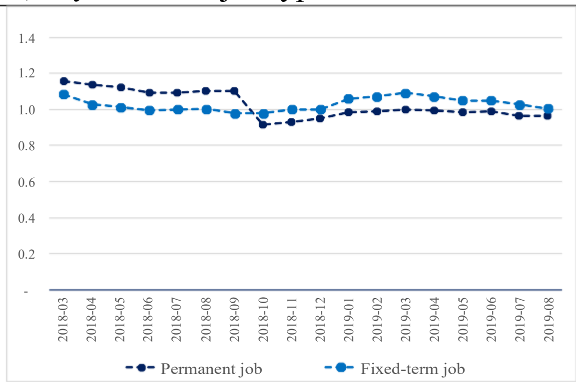
b) by loan maturity



c) by income



d) by borrower job type

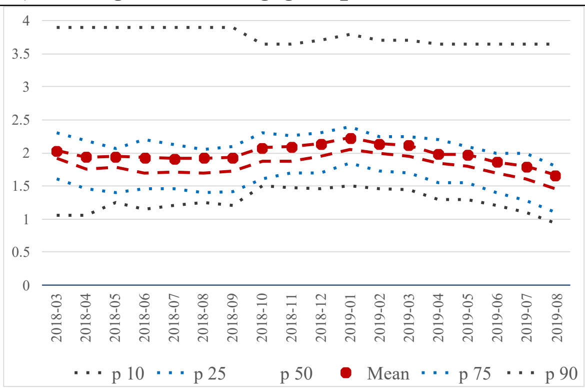


Source: Mutuionline data.

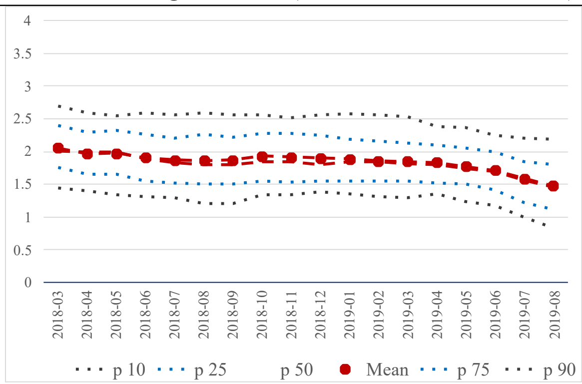
Figure 9: Mutuionline interest rates by bank category (*percentage points*)

Panel A. Fixed–rate mortgages

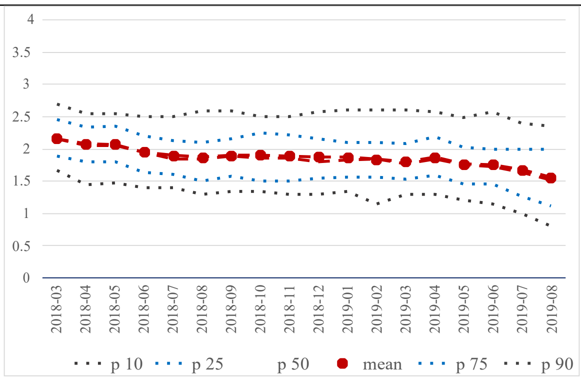
a) 5 Largest banking groups



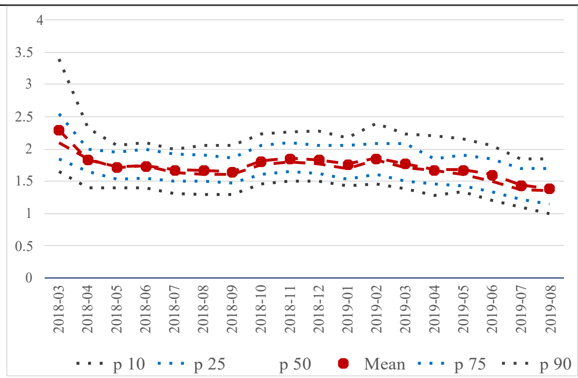
b) Other Significant (subsidiaries included)



c) LSI

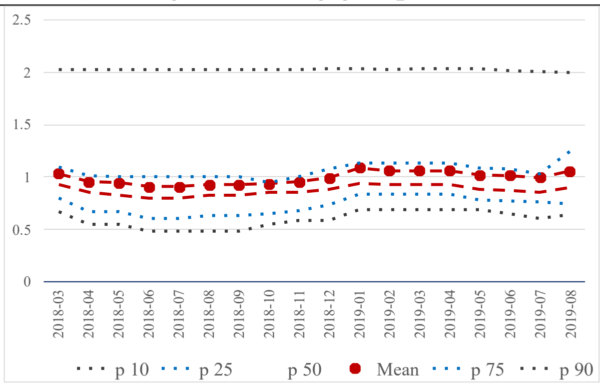


d) Online banks

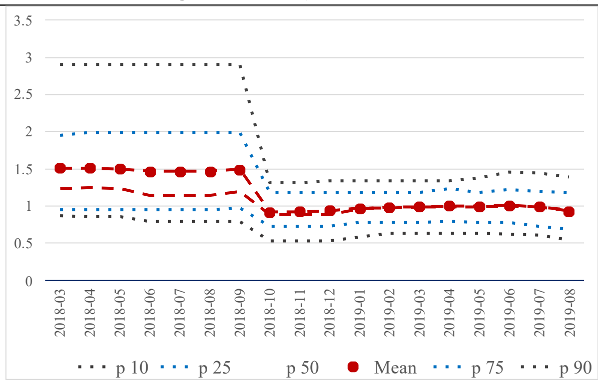


Panel B. Adjustable-rate mortgages

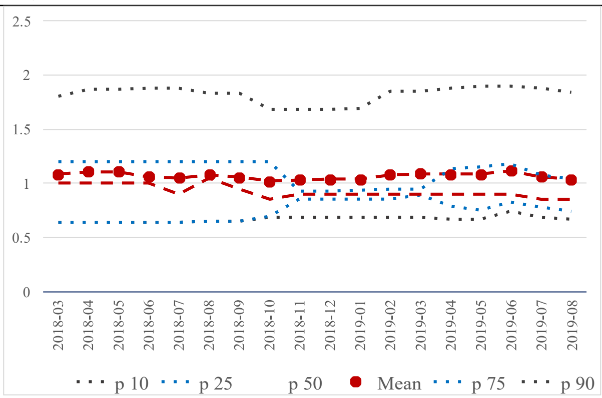
a) 5 Largest banking groups



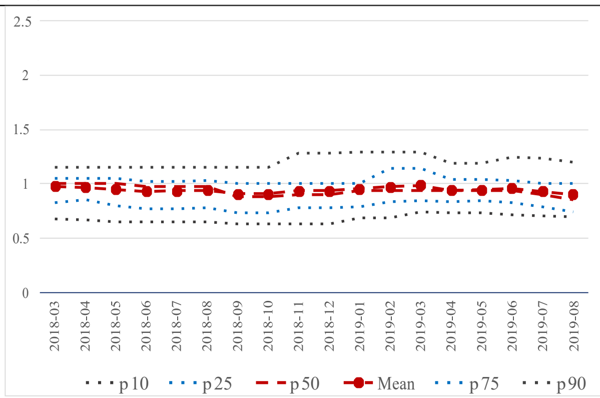
b) Other Significant (subsidiaries included)



c) LSI



d) Online banks



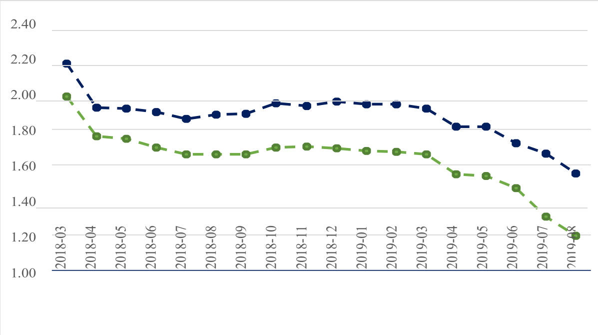
Source: Mutuionline data.

Figure 10: Mortgage rates by risk profiles

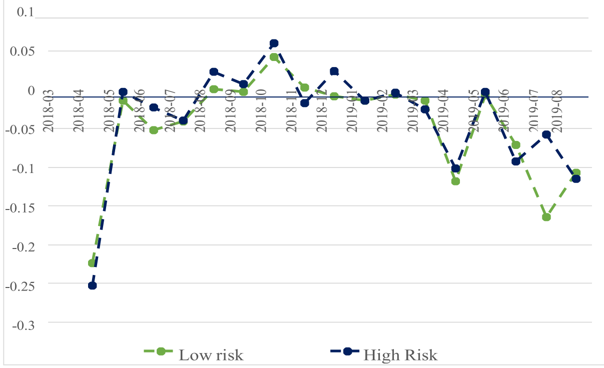
(*percentage points*)

Panel A. Fixed-rate mortgages

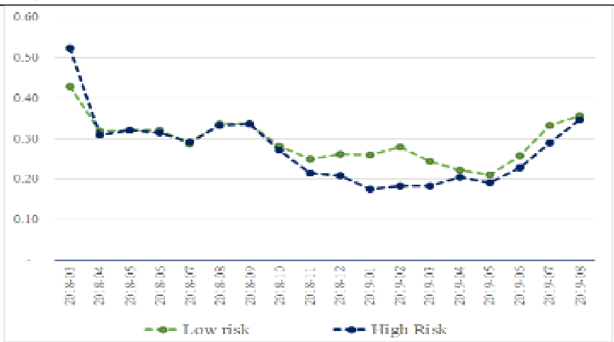
a) Levels



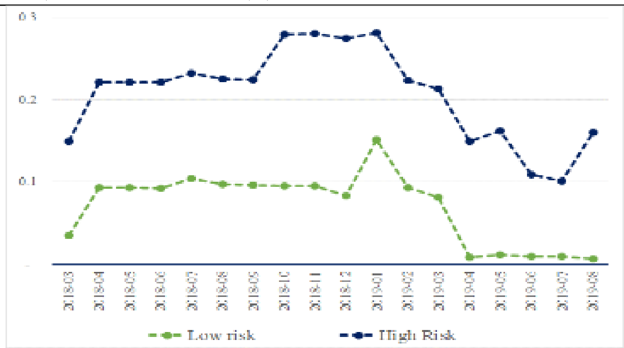
b) M-to-m changes



c) Standard deviation

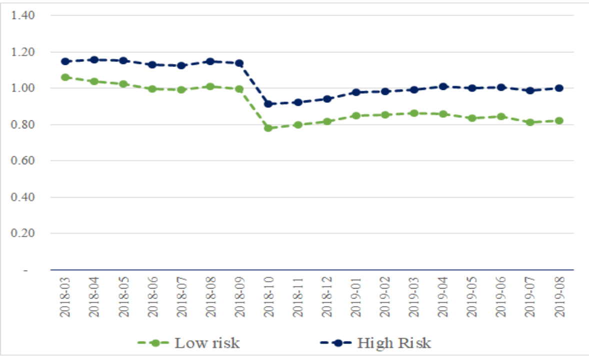


d) No-offer rate (1)

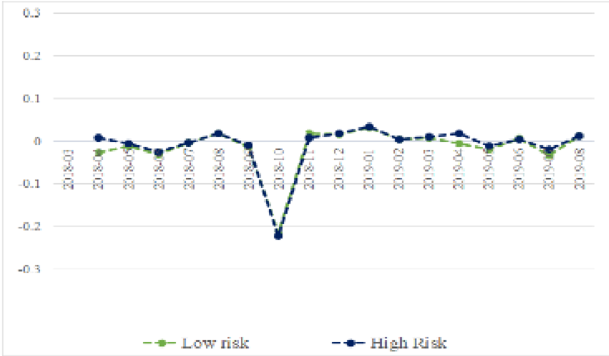


Panel B. Adjustable-rate mortgages

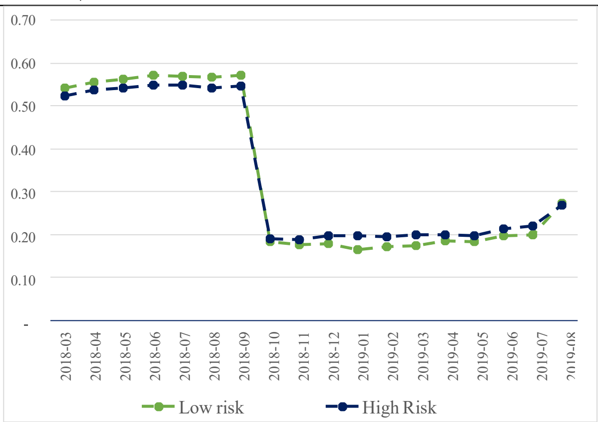
a) Levels



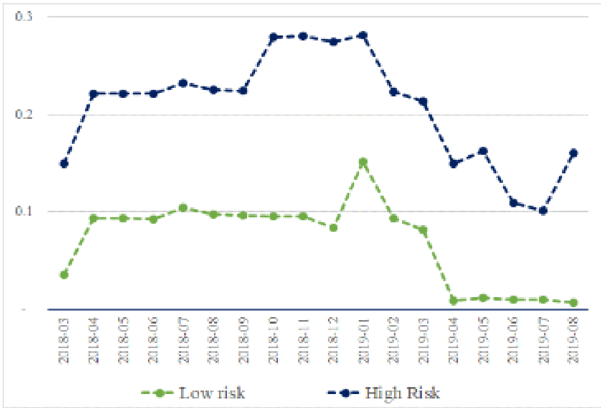
b) M-to-m changes



c) Standard deviation



d) No-offer rate (1)



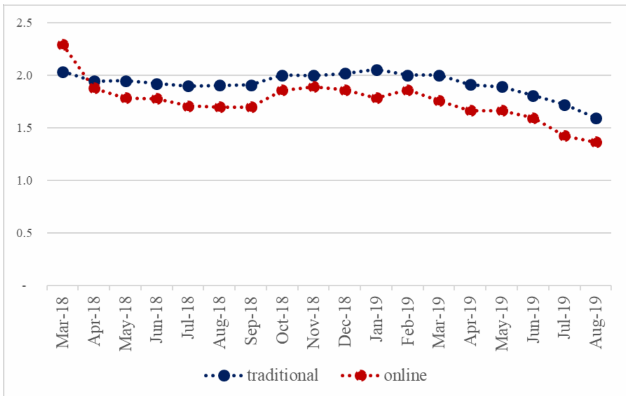
Source: Mutuionline data.

Notes: (1) No-offer rate is defined as the average rejection rate for each given profile.

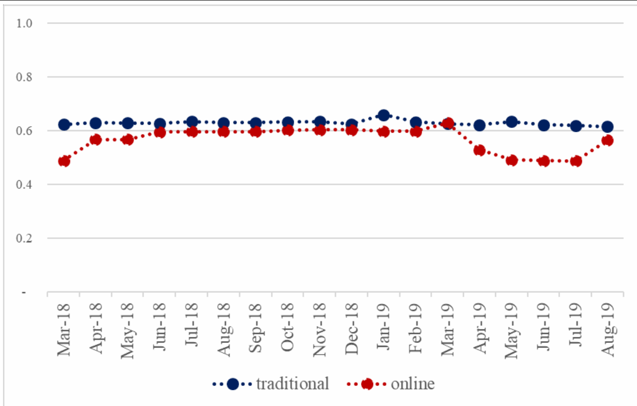
Figure 11: Traditional versus online banks
(*percentage points*)

Panel A. Fixed-rate mortgages

a) Interest rates

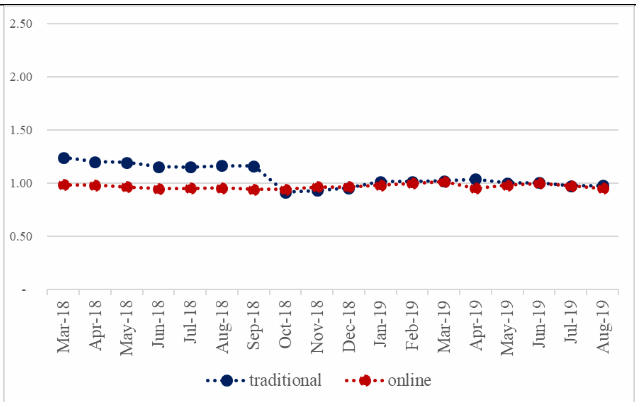


b) No-offer rate (1)

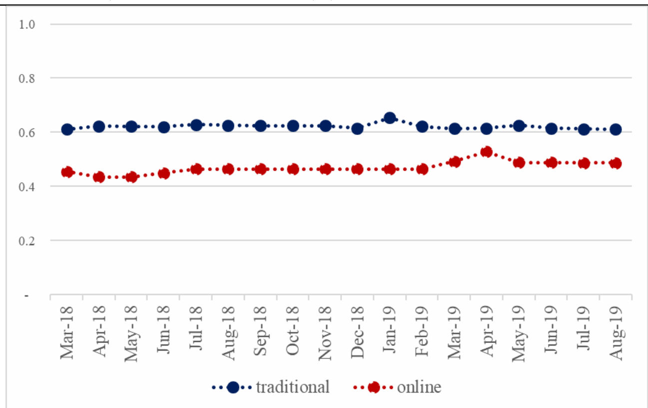


Panel B. Adjustable-rate mortgages

a) Interest rates



b) No-offer rate (1)

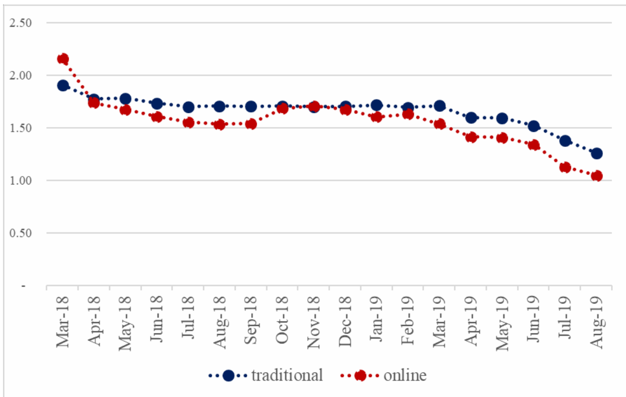


Source: Mutuionline data.
Notes: (1) No-offer rate is defined as the average rejection rate for each given profile.

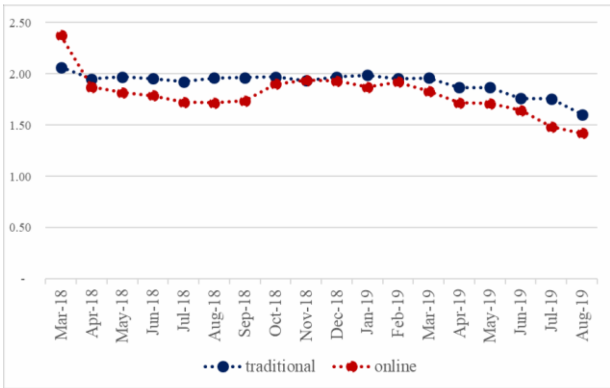
Figure 12: Traditional versus online banks by risk profile
(*percentage points*)

Panel A. Fixed-rate mortgages

c) Low risk

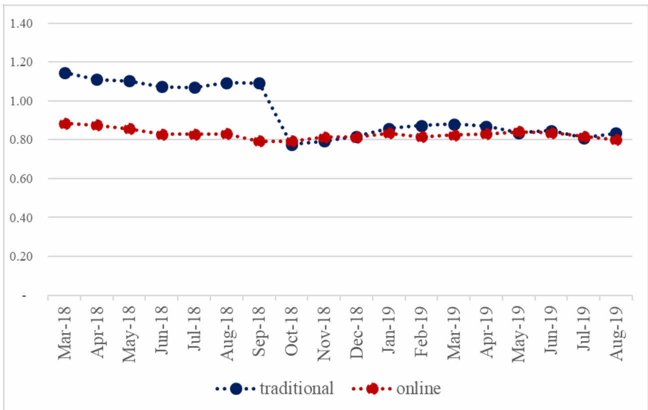


d) High risk

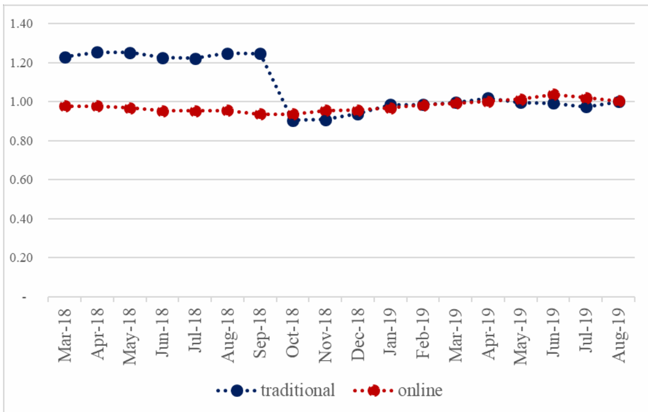


Panel B. Adjustable-rate mortgages

c) Low risk



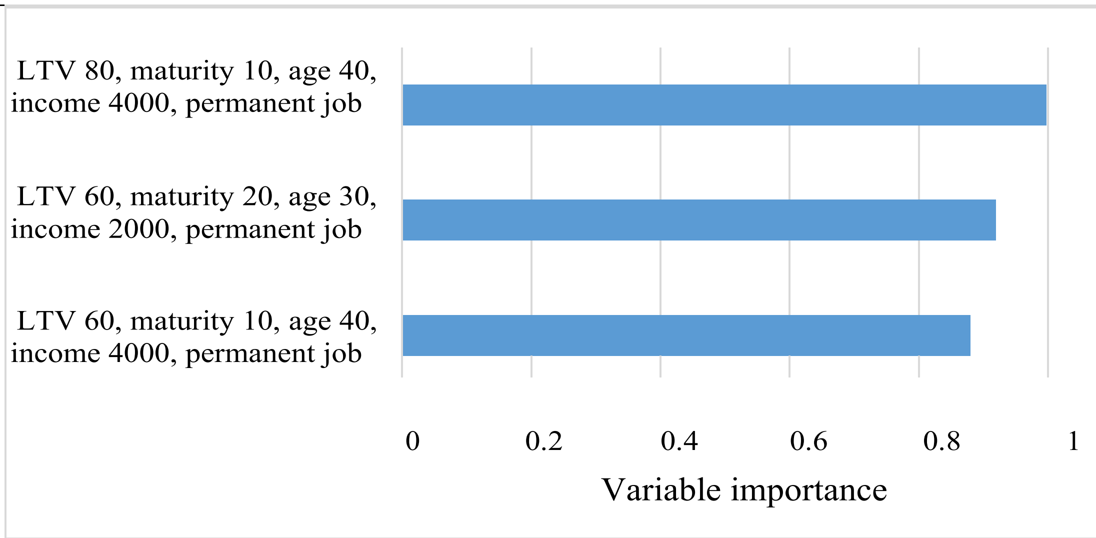
d) High risk



Source: Mutuionline data.

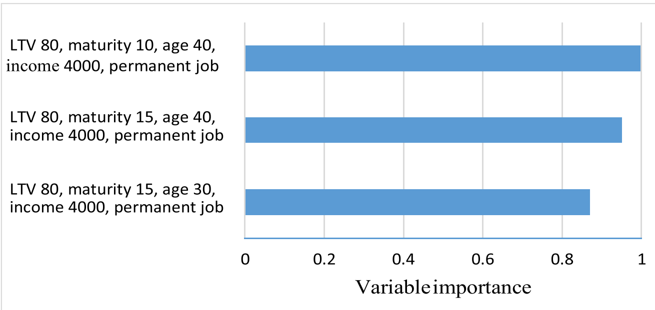
Figure 13: Random Forest, fixed rate

Panel A: Profile importance June 2019

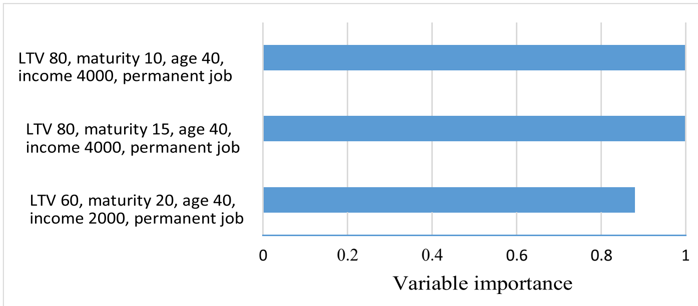


Panel B: Profile importance over time

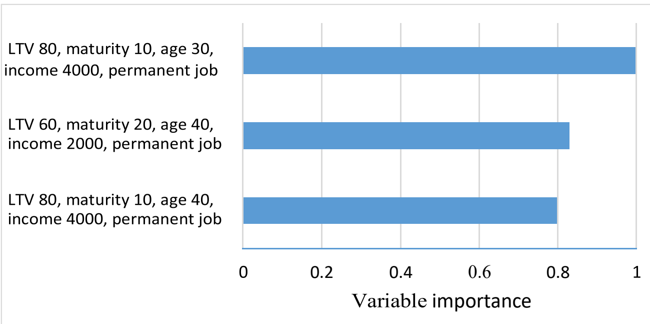
a) May 2019



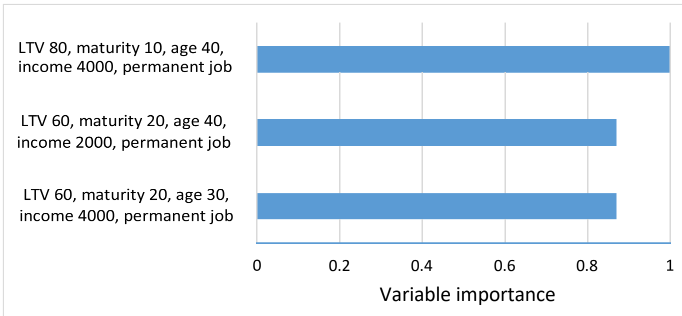
b) April 2019



c) March 2019



d) February 2019



Source: Mutuionline data.
Notes: The variable importance provides a score that indicates how useful each indicator was in predicting the target variable in the random forest. The importance varies between 0 and 1 and 1 is assigned to the most informative variable.

10. Appendix

Table A.1: List of banks in the Mutuionline dataset

Bank name	Class	Online	All periods
GRUPPO UNICREDIT: Unicredit	SI (top 5)		YES
GRUPPO INTESA SAN PAOLO	SI (top 5)		
Intesa Sanpaolo			YES
Carisbo			
Banca CR Firenze			
Cassa di Risparmio del Friuli Venezia Giulia			
Cassa di Risparmio del Veneto			
Cassa dei Risparmi di Forlì e Romagna			
Banco di Napoli			
Cassa di Risparmio di Pistoia e della Lucchesia			
GRUPPO MPS	SI (top 5)		
Banca Monte dei Paschi di Siena			YES
Widiba		X	YES
UBI	SI (top 5)		
UBI Banca			YES
IW Bank			YES
GRUPPO BPM	SI (top 5)		
Banco BPM			YES
Webank		X	YES
GRUPPO CREDEM: CREDEM	SI		YES
GRUPPO BPER	SI		
BPER Banca			YES
Banco di Sardegna			YES
Cassa di Risparmio di Bra			YES
GRUPPO CARIGE	SI		
Banca CARIGE			YES
Banca Monte Lucca			YES
GRUPPO MEDIOBANCA	SI		
CheBanca!		X	YES
BNL - Gruppo BNP Paribas	SI (EU)		
BNL			YES
Hello bank!		X	YES
GRUPPO Deutsche Bank	SI (EU)		YES
GRUPPO CREDIT AGRICOLE	SI (EU)		YES
FriulAdria			
Carispezia			
Cariparma - Crédit Agricole			
Bancadinamica (CARISMI)		X	
ING - Direct	SI (EU)	X	
GRUPPO BANCA SELLA	LSI		
Banca Sella			YES
GRUPPO BANCO DESIO			
Banco di Desio e della Brianza	LSI		YES
Banca d'Alba Credito Cooperativo	LSI		YES
Banca Macerata	LSI		YES
Extrabanca	LSI		YES
Banca Popolare di Spoleto	LSI		
Banca Popolare di Puglia e Basilicata	LSI		YES
BCC Banca di Ancona	LSI		
Banca Popolare Pugliese	LSI		YES
Emil Banca	LSI		YES
CrediFriuli	LSI		YES
BCC Bene Vagienna s.c.	LSI		YES
Banca di Verona Credito Cooperativo Cadidavid	LSI		YES
BCC ravennate, forlivese e imolese Soc. coop	LSI		
BCC Casalgrasso e Sant'Albano Stura	LSI		YES

Notes: The list of banks refers to the period March 2018-August 2019. SI stands for significant institution according to SSM definition; SI(EU) means other EU significant institution; LSI means less significant institution.

Table A.2: Main balance sheet characteristics

(December 2018; percentage values)

	Loan-to-asset	Share of loans for house purchase	Cost-to-income	Diversification index	ROA	Liquidity ratio	Net interest-to-asset	T1 ratio	RWA-to-asset ratio	T1-to-asset ratio	NPL ratio
					MUTUIONLINE SAMPLE: TOTAL						
Mean	55.5	27.3	73.9	0.5	0.2	13.4	1.3	18.3	40.5	6.7	11.4
Quart. 1	48.6	15.4	63.4	0.4	0.1	5.5	1.1	12	33.5	5	5.9
Median	60.5	24.8	71.4	0.5	0.2	9.9	1.3	15.7	41.9	6.4	11.2
Quart. 2	67	33	79.2	0.6	0.4	24.5	1.6	18	47	7.6	15
W. Avg.	52.2	24.2	63	58.2	0.4	7.3	1	17.4	41.4	7.2	11.8
					MUTUIONLINE SAMPLE: SIGNIFICANT						
Mean	59.5	29.2	72.7	0.5	0.1	5.2	1.1	16	37.7	5.9	14.1
Quart. 1	49.8	20.8	64.5	0.5	0	1.1	1	11.2	32.4	4.6	11.3
Median	62.8	29	72.2	0.5	0.2	6	1.1	14.7	38.7	5.6	15.1
Quart. 2	66.7	32	79.1	0.6	0.4	7.7	1.2	18	43.8	6.9	16.9
W. Avg.	50.2	22.4	61.4	60.1	0.4	7.2	1	17.8	41.8	7.5	12.3
					MUTUIONLINE SAMPLE: SIGNIFICANT EU						
Mean	66.8	34.6	73.2	0.5	0.2	7.5	1.5	12.7	44.8	5.4	7.2
Quart. 1	60.9	32.5	63.7	0.4	0.3	5.5	1.5	10.6	36.5	4.3	4.7
Median	68.8	33.6	64.2	0.5	0.4	10.1	1.6	10.6	40.5	5.4	5.9
Quart. 2	68.9	38	71.4	0.6	0.6	10.1	1.6	12	45	6	7.6
W. Avg.	68.7	33.1	71.2	48.9	0.2	6.7	1.4	13.6	44.1	6	10.2
					MUTUIONLINE SAMPLE: LESS SIGNIFICANT						
Mean	51.6	20	72.4	0.5	0.3	25.6	1.4	22.4	43.1	8.1	10.1
Quart. 1	44.5	13.2	62.2	0.4	0.2	19.6	1.3	14.6	41.1	6.3	6.3
Median	58.4	20.3	71.4	0.4	0.3	27.5	1.4	16.2	45.7	7.4	9.5
Quart. 2	60.3	23.8	75.5	0.5	0.4	31.1	1.6	19.7	49	9	14.1
W. Avg.	52.9	19.5	67.8	55.9	0.4	19.4	1.3	19.1	44.5	8.5	10.4
					TOTAL BANKING SYSTEM						
Mean	49.6	15	73.7	0.4	0	22.6	1.5	17	38.6	7.3	12.3
Quart. 1	42.6	3.6	63.6	0.3	0.1	10.2	1.2	12.7	32.5	5.4	7.1
Median	53.4	14.7	71.5	0.4	0.3	25	1.6	15.9	42.3	7.1	11.7
Quart. 2	61.2	22.4	80.2	0.5	0.5	32	1.9	20.7	48.8	9.4	16.7
W. Avg.	47.2	19.3	60.8	55.2	0.4	10.1	1	16.7	37.4	6.2	11.1

Source: Supervisory reports.

Table A.3: Distribution of mortgage rates by borrower and contract characteristics
(*averages March 2018 - August 2019; percentage points)*

	Mean	Median	Sd.	P25	P75
LTV					
50	1.27	1.13	0.58	0.80	1.70
60	1.37	1.25	0.56	0.90	1.75
80	1.48	1.40	0.58	0.99	1.85
85	3.01	2.55	0.87	2.13	3.90
Maturity					
10	1.23	1.12	0.59	0.85	1.45
15	1.36	1.29	0.63	0.88	1.70
20	1.43	1.35	0.65	0.90	1.84
30	1.68	1.68	0.71	1.05	2.20
Rate type					
Fixed	1.87	1.77	0.59	1.50	2.10
Adjustable	1.03	0.94	0.45	0.78	1.13
Income					
2000	1.44	1.30	0.68	0.93	1.85
4000	1.41	1.28	0.66	0.90	1.80
Job type					
Permanent job	1.42	1.30	0.65	0.90	1.84
Other	1.45	1.27	0.74	0.90	1.84
Age					
30	1.43	1.30	0.67	0.90	1.84
40	1.43	1.29	0.67	0.90	1.84
Mortgage type					
First home	1.38	1.20	0.67	0.88	1.80
Subrogation	1.47	1.35	0.67	0.97	1.87
Geographical area					
North-East	1.42	1.30	0.66	0.91	1.81
North-West	1.43	1.29	0.70	0.88	1.85
Centre	1.43	1.30	0.66	0.90	1.83
South	1.42	1.29	0.66	0.92	1.83
Islands	1.44	1.30	0.69	0.90	1.85

Source: Mutuionline data.

Table A.4: Distribution of no-offer rates for main borrower and contract characteristics (1)
(*averages March 2018 - August 2019)*

	Mean	Median	Sd.	P25	P75
LTV					
50	0.46	0.00	0.50	0.00	1.00
60	0.46	0.00	0.50	0.00	1.00
80	0.51	1.00	0.50	0.00	1.00
85	0.94	1.00	0.24	1.00	1.00
Maturity					
10	0.62	1.00	0.49	0.00	1.00
15	0.60	1.00	0.49	0.00	1.00
20	0.57	1.00	0.50	0.00	1.00
30	0.59	1.00	0.49	0.00	1.00
Rate type					
Fixed	0.61	1.00	0.49	0.00	1.00
Adjustable	0.57	1.00	0.49	0.00	1.00
Income					
2000	0.61	1.00	0.49	0.00	1.00
4000	0.57	1.00	0.49	0.00	1.00
Job type					
Permanent job	0.35	0.00	0.48	0.00	1.00
Other	0.83	1.00	0.38	1.00	1.00
Age					
30	0.59	1.00	0.49	0.00	1.00
40	0.59	1.00	0.49	0.00	1.00
Mortgage type					
First home	0.58	1.00	0.49	0.00	1.00
Subrogation	0.60	1.00	0.49	0.00	1.00
Geographical area					
North-East	0.59	1.00	0.49	0.00	1.00
North-West	0.59	1.00	0.49	0.00	1.00
Centre	0.59	1.00	0.49	0.00	1.00
South	0.58	1.00	0.49	0.00	1.00
Islands	0.61	1.00	0.49	0.00	1.00

Source: Mutuionline data.
Notes: (1) No-offer rate is defined as 1 – (share of banks that offer a product for each given profile).

Table A.5: Distribution of interest rates and APR for banks' categories
(averages March 2018 - August 2019; percentage points)

a) Fixed-rate mortgages

	Mean	P10	P25	P50	P75	P90	Min	Max	Sd.
				5 largest banking groups (Significant groups)					
Interest rates	1.98	1.25	1.5	1.8	2.2	3.65	0.5	4.05	0.77
APR	2.17	1.44	1.69	1.98	2.36	3.83	0.69	4.43	0.79
				Other significant groups					
Interest rates	1.87	1.25	1.55	1.85	2.2	2.58	0.4	3.73	0.5
APR	2.05	1.45	1.74	2.05	2.34	2.69	0.78	3.29	0.47
				Less significant groups and banks					
Interest rates	1.89	1.3	1.51	1.8	2.2	2.56	0.85	4.4	0.54
APR	2.1	1.5	1.71	2.02	2.37	2.88	1.03	4.72	0.55
				Subsidiaries					
Interest rates	1.81	1.2	1.5	1.71	2.15	2.44	0.61	3.08	0.48
APR	1.94	1.35	1.61	1.85	2.29	2.57	0.82	3.13	0.45
				Online banks and channels					
Interest rates	1.75	1.3	1.5	1.7	1.99	2.2	0.6	3.84	0.4
APR	1.82	1.37	1.57	1.76	2.03	2.28	0.6	3.95	0.39

b) Adjustable-rate mortgages

	Mean	P10	P25	P50	P75	P90	Min	Max	Sd.
				5 largest banking groups (Significant groups)					
Interest rates	0.99	0.58	0.73	0.88	1.08	2.03	0.33	2.28	0.45
APR	1.15	0.74	0.9	1.02	1.23	2.12	0.43	2.52	0.47
				Other significant groups					
Interest rates	1.02	0.58	0.8	1	1.2	1.4	0.24	1.95	0.31
APR	1.18	0.78	1	1.2	1.34	1.55	0.51	2.03	0.28
				Less significant groups and banks					
Interest rates	1.07	0.65	0.7	0.9	1.15	1.83	0.42	3.43	0.52
APR	1.25	0.77	0.88	1.1	1.36	2.05	0.52	3.8	0.55
				Subsidiaries					
Interest rates	1.38	0.66	0.86	1.07	1.46	2.9	0.4	2.9	0.77
APR	1.51	0.8	0.97	1.23	1.47	2.98	0.55	3.27	0.79
				Online banks and channels					
Interest rates	0.94	0.68	0.79	0.95	1.04	1.25	0.53	1.59	0.22
APR	0.99	0.77	0.88	0.99	1.07	1.24	0.57	1.61	0.2

Source: Mutuionline data.

Table A.6: Predicting interest rates using regression analysis and machine learning: Adjustable-rate mortgages

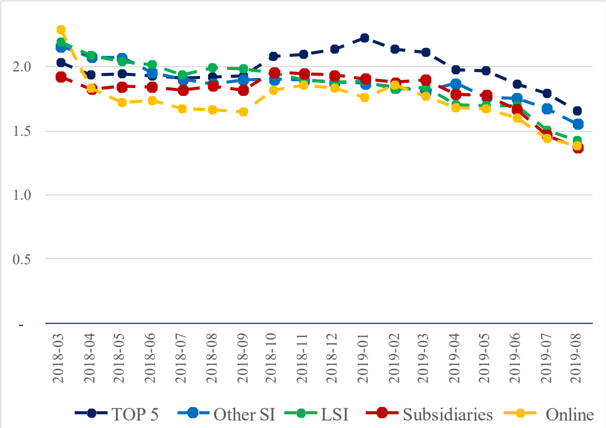
	MIR delta adjustable interest rate (b,t)		
	Only Mutuionline average	Adding controls for demand	Machine learning
	(1)	(2)	(3)
MO delta interest rate (b,t)	0.006	0.017	
	(0.029)	(0.033)	
L.MO delta interest rate (b,t)	0.086	0.097	
	(0.060)	(0.071)	
L.Consumer confidence		-0.005	
		(0.006)	
L2.Consumer confidence		-0.003	
		(0.013)	
Bank FE	Y	Y	Y
Observations	288	288	320
Adjusted R-squared	-0.069	-0.073	
Out of sample RMSE	0.1606	0.9038	0.1901
Out of sample correctly predicted (%)	82.35	0	82.35

Notes: An observation is correctly predicted if its out of sample prediction is within one standard deviation of true value of the dependent variable. Standard errors clustered at bank level in parentheses in Columns 1-2. b stands for banks and t for time (months). Ln indicates that the variable is lagged of n months.
*** p<0.1, ** p<0.05, * p<0.01

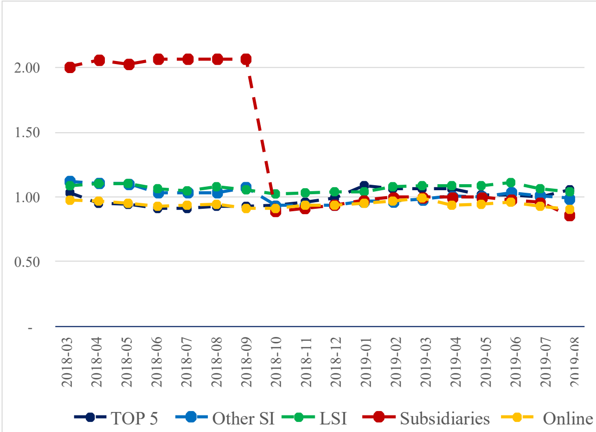
Figure A.1: Mutuionline interest rates by bank type: mean value

(percentage points)

a) Fixed



b) Adjustable



Source: Mutuionline data.