Journal of Real Estate Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjer20

Can Social Capital Variables Help to Determine Loan to Value Approved by Banks?

Josep Maria Raya

JOURNAL OF REAL ESTATE RESEARCH
https://doi.org/10.1080/08965803.2021.2024678

Can Social Capital Variables Help to Determine Loan to Value Approved by Banks?

Josep Maria Raya

Escola Superior de Ciencies Socials I de la Empresa, Tecnocampus, Universitat Pompeu Fabra, Mataró, Barcelona

ABSTRACT
This paper sheds some light on the effects of social capital variables (social network data, physical appearance, etc.) on loan-to-value (LTV), a crucial variable to evaluate systemic risk. Using a unique database created by merging several sources of data, we show that the introduction of social capital variables are shown to be statistically significantly related to LTV. In particular, Facebook likes in a month and creditworthiness are a negative determinant of LTV while beauty and certain personality traits play a role in borrowers obtaining a higher LTV. We distinguish these effects depending on the LTV variable used: loan-to-appraisal (entirely under the control of lender) and loan-to-transaction (in which the transaction price can also be influenced). As policy implications we found that social capital variables capture information that would otherwise be unobservable using only the traditional variables in the sense that they are related to information lenders may have at lending that the researchers do not observe.

KEYWORDS
social networks; mortgages; beauty premium; bank performance; loan-to-value

Introduction

The internet is changing banking (King, 2014). The use of big data to improve efficiency in the banking sector and competition from producers of mass information entering parts of the value chain in financial products are two examples.

One of the most important tasks of a financial institution is adequately assessing the risk of borrowers. To do this, lenders effectively calculate probabilities of default for each borrower. Traditionally, banks have used internal indicators drawing on previous relationships with customers, past transactions, available information on income and wealth, and so on. In many cases, these models of borrower financial strength (Avery et al., 2004) are specific to each entity.

Agarwal and Hauswald (2010) found that supplementing these models with the human touch of loan officers decreases the default rate. Recently, to improve the performance of the model, related research verified the utility of traditional variables, such as income, age, FICO score[1], and debt information. They have also focused on the social capital variables of borrowers, which involve online reputation, social status and contacts.

CONTACT Josep Maria Raya josep.raya@upf.edu Escola Superior de Ciencies Socials I de la Empresa, Tecnocampus, Universitat Pompeu Fabra, C/Ernest Lluch 32, Mataró 08302, Barcelona.

In this respect, social capital variables have received much attention. These models use variables such as the number of followers on Facebook, number of likes, groups, events, time spent in social networks, whether the profile is updated, connections and their characteristics, neighbourhood in which the individual lives, etc.[2]. Companies such as Neo Finance, Creditech, and LenddoEF use social credit scores in microfinance. Some such models use peer pressure to increase the likelihood of paying outstanding claims (De Cnudde et al., 2019; Montalvo, 2014)[3].

In general, this paper contributes to the literature on the importance of social capital variables in the mortgage market (Agarwal et al., 2011; Bandyopadhyay, 2021; Ergungor & Moulton, 2014). Woolcock (1998) defines social capital as the norms, values, and trust in a social network, which enables cooperative and shared actions. Social capital variables is different from soft information (Agarwal et al., 2011) which can be defined as the information difficult to verify, quantify or store that a lender could obtain about the borrower's ability and willingness to repay (the likelihood of receiving future income or wealth shock, or whether the borrower has sufficient self-control and financial discipline). The main aim of this paper is to shed light on the effect that information on social capital variables introduces in the loan to value granted. Social capital variables capture information that would otherwise be unobservable using only the traditional variables. On the one hand, previous papers that use social capital variables are based on small consumer credit and self-disclosed social media information (Ge et al., 2017; Lin et al., 2013). Additionally, we use non-self-disclosed information. We avoid, therefore, self-disclosed information. Online self-disclosure is explained by the general tendency to choose short-term rewards by neglecting long-term risks as well as by tendencies toward a problematic social-networks-use (Ostendorf et al., 2020). This is the first such paper to focus on the mortgage market. On the other hand, previous papers that have attempt to capture the importance of nontraditional variables to determine loan performance in the mortgage market have not used social capital variables (Saengchote, 2013; Wanzare & Scheule, 2010). Li et al. (2020) showed empirical evidence to address the impact of social capital on mortgage delinquency. The paper used a linear regression model to explain the likelihood of mortgage delinquency and country-level dataset between 1999 and 2011. The also used a local level index of social capital which included information of the density of civic, religious and sports organizations. They concluded that the assessment of default risk should take social capital into consideration, besides the factors already documented in the literature. Apart from this paper, this is the first paper that explicitly uses social capital variables in the mortgage market. Unlike Li et al. (2020), we use individual data and individual measures of social capital variables rather than a local-level index. We also obtain that social capital variables explain mortgage market outcomes. However, we can disaggregate the effects of social capital variables in the effect of social network information, the beauty premium and a trustworthy appearance on loan to value. Our paper also differs from Guiso et al. (2013) in which they focused on the willingness of strategic default and its determinants. The paper used survey data and estimates probit models on strategic default. They found that people that people who considered immoral to default was less willing to default in a hypothetic situation. Also people who trusted banks less and was angrier about economic situation was more likely to strategic default. In our paper we also obtain that social capital variables determines a mortgage market output, however, unlike Guiso et al. (2013), we use a compilation of several different sources of microdata, but not survey data. That is, we focus on the effect social network information, the beauty premium and a trustworthy appearance on loan to value of a real mortgage. Also, we do not focus on strategic default, which is a rare practice in Spain, rather we focus on the loan to value approved by the bank.

In particular, we analyse how social capital variables are important determinants of the loan to value approved by the bank. We focus on loan to value as a dependent variable. Discussion among policy-makers, economists, and commentators has emphasized the pivotal role of loan-to-value (LTV) ratios in measuring expected credit losses (Claessens et al., 2013; Crowe et al., 2013; Lim et al., 2011; Mayordomo et al., 2020). In this respect, maximum LTV ratios on mortgages have been adopted by many countries as a macro-prudential instrument to address systemic risk (Montalvo & Raya, 2018; Wong et al., 2011). We use a unique dataset that has information on the mortgagor, characteristics of the loan and social capital variables of the mortgagor (viz., number of photos of leisure activities and restaurants and number of LinkedIn contacts). The results show that a higher number of Facebook likes is also a determinant of LTV.

Finally, an additional motivation of the paper is that most studies are conducted US data. In this paper, we use data from the Spanish mortgage market. The economy of Spain during the boom years offers an excellent setting in which to analyse mortgage market questions. In addition to higher price growth rates, Spain exhibited overly soft lending standards and excessive risk-taking, a behaviour made possible using over-appraising (defined as the difference between housing collateral values computed by the appraisers and actual transaction prices). Overappraising was an extensive bad practice of Spanish banks during the boom years to circumvent macroeconomic prudential policies (Montalvo & Raya, 2018). The paper is also novel in analysing the importance of social capital variables to overappraising. In addition, the Spanish mortgage market is very different from the US mortgage market. Therefore, analysing the effect of social capital variables in the loan to value in Spain is also a novelty in itself. The results show that a higher number of LinkedIn contacts diminishes overappraising. The importance of overappraising highlights larger differences among loan-to-appraisals and loan-to-transaction prices. In this sense, we distinguished these effects depending on the LTV variable used: loan to appraisal (entirely under the control of lender) and loan to transaction (in which the transaction price cannot be influenced).

Although the most important source of social capital variables used in the paper is social network information, through the use of social network information, we have collected additional information allowing us to test whether physical appearance affects lenders (Duarte et al., 2012; Gonzalez & Loureiro, 2014; Ravina, 2008) . In the "beauty premium" theory, attractiveness is associated with health, intelligence and competence (Eagly et al., 1991). Wilson and Eckel (2006) ran a version of a trust game in which trustors viewed the trustees' faces and found that attractive trustees were perceived as more trustworthy. In this sense, it is useful to control for attractiveness to correctly estimate trustworthiness. To measure the effect of physical appearance on the mortgage market, we chose a typical profile photo from Facebook. Then, by using the MTurk service, subjects were rated on their physical attractiveness and the first impressions they made regarding their trustworthiness and creditworthiness, which might matter in a person-to-person interaction over and above their demographic characteristics. For the first time in regard to mortgage loans, we test the relationship between the beauty premium and perceived trustworthiness. Not only a beauty premium but also a "happiness" premium is observed in the results. In addition, the results related to creditworthiness confirm the existence of bad practices during these boom years in Spain.

The paper is structured as follows. First, we contextualize the Spanish mortgage market. Then, we present the different sources of data used. Next, we conduct the analysis, and finally, we end with some concluding remarks.

Spanish Mortgage Market: Boom and Bust

During the first decade of this century, Spain experienced one of the most remarkable housing booms among developed economies. In the period of 1998—2007, the housing prices in Spain tripled in nominal terms. The rate of increase in house prices peaked in 2004 above 15% in annual terms. One interesting characteristic of this housing boom in Spain is the coincidence of the remarkable rise in prices with a huge increase in supply. An average of 600,000 dwellings were built yearly between 2000 and 2007 (Department of Public Works). This number exceeded, for some years, combined new construction in the other 4 large EU economies, i.e., Germany, France, UK and Italy.

This housing boom was reflected in a credit boom, with rates of growth that peaked above 25% in 2006, of which 15 points was related to housing, construction and property development. The average number of conceded mortgages was more than 1 million per year. These amounts are quite remarkable if we consider that in Spain, the annual average number of households in that period was 15.5 million (Spanish Statistic Bureau). Likewise, these numbers were also made possible by very strong competition in the mortgage origination business. Spanish financial institutions offered the lowest mortgage rates of the Euro area. In fact, during the 2003–06 period, the average mortgage rate in the Euro area was 4.51, while the average in Spain was 3.71 (Bank of Spain). Higher competitive pressure implied that managers of financial institutions could increase profits only by originating many new mortgages.

Mortgages in Spain had some different characteristics than those granted in the United States. First, the predominance of adjustable-rate mortgages (ARMs)[4] accounted for approximately 98% of the stock of mortgages. Second, in Spain, 99% of loans were granted by deposit institutions, while in the US, such institutions only granted 50% of the total. Third, of the total mortgages granted, 70% were securitized in the US, while in Spain, that figure was only 13.5% (Bank of Spain). Finally, in Spain, as in the UK, interest rates are product specific and do not reflect borrowers' creditworthiness, as in the United States. Products depends on the banks[5]. During boom years, differences in spread were reduced (due to intense competition among banks) but still existed because not every bank assumed the same level of credit risk. Akin et al. (2014) show that banks with worse corporate governance problems soften standards to an even larger degree. Additionally, not every borrower has the same probability of obtaining every product. However, in those years, excessively soft credit standards offered access to credit to almost every potential borrower (Akin et al., 2014). As a result, interest rates were independent of creditworthiness, and mortgage amounts due to soft credit standards were discussed.

The specific agency mechanism that inflated the bubble in the US was quite different from the forces at work in the Spanish case. In both cases, lax standards and excessive credit were the ultimate causes of housing price inflation (Akin et al., 2014). In the case of the US, those lax standards for mortgage granting were the result of perverse incentives in the housing finance sector related to the securitization process and the possibility of taking out of the banks' balance sheets securitized mortgages. An alternative pathway can be documented for Spain (Montalvo & Raya, 2018): real estate appraisal firms were encouraged to introduce an upwards bias in appraisal prices to satisfy their owners or their most important clients (i.e., the banks).

Spanish banking regulations did not allow for deconsolidating securitized mortgages; therefore, banks could not improve their capital ratios by securitizing mortgages, as in the US, where securitization was a mechanism to transfer risk. However, Spanish banks needed to finance the increasing number of mortgages that they were originating. For this purpose, traditionally, they issue large amounts of covered bonds backed by their portfolio of mortgages. However, only mortgages with a loan-to-value ratio of 80% or below can count for the backing portfolio. This is the point at which Spanish appraisal companies played a role similar to rating agencies: their valuations determined what mortgages could be included in the pool of mortgages that could back the issuance of covered bonds.

In this case, the perverse incentive was the fact that most of the appraisals were performed by companies owned by financial institutions. Appraisal companies concentrated 76% of their turnover in just one or two clients, which were banks. An indirect effect of the influence of appraisals companies, owned by or including banks as participants, on the rest of the sector is the fact that, given the technical conditions to calculate an appraisal, once a large part of the market tends to produce overappraisals, then other companies that use those dwellings as comparables for new appraisals (with a valuation requiring six comparables) will translate that overappraisal to other real estate properties. In a period of rising home values, granting larger mortgages may not seem risky. By this logic, during the housing boom, financial institutions were prone to opening the market to weak borrowers with financial constraints. To this end, appraisers were encouraged to introduce an upwards bias in appraisal prices. As the appraisal was the value used by financial institutions to determine the loan-to-value ratio, even for new mortgages, this artificial increase in appraisal prices allowed them to draw larger mortgages.

The excessive dependence on the real estate industry, together with a softening of credit standards, caused the economic and financial crisis to hit Spain more severely than other developed economies. Household debt in Spain (loans to households for mortgages and consumer credit) was 91% of GDP in 2010, compared to just below 106% in the UK and 95% in the USA, but substantially higher than in France and Germany, with 69% and 64%, respectively.[6] In this context of an economic recession in Spain, one of the most controversial issues is the Spanish Mortgage Law, which was approved in 1909. Spanish mortgages are loans with full recourse, in contrast to the limited recourse mortgage loans in most states of the US. While in these US states the guarantee is the dwelling, this is not the case in Spain. In the event of a mortgage foreclosure, the lender seizes the mortgagor's dwelling, which is sold at auction at a price generally quite below its market price. After that, borrowers still hold a debt consisting of the outstanding mortgage debt minus the auction price of the dwelling plus the interest for late payment, which is generally quite high.

Higher household debt coupled with a rapid increase in the unemployment rate led to a growing number of families being unable to service their mortgage debt. Default rates grew from 0.5% in 2006 to 6.3% in 2014 (Bank of Spain). The Spanish government implemented some measures against this situation, such as a 'code of good banking practices' and 'urgent measures to strengthen protection for mortgage holders' (Royal Decree-Law 27/2012). Both measures established the immediate moderation of late-payment interest applied to that group and encouraged the renegotiation and refinancing of the loans of these families. Refinancing is a typical practice in Spain. The main goal of refinancing is to improve the interest rate[7]. The current mortgage will be 'replaced' by the new mortgage; thus, the current mortgage will be cancelled, and a new mortgage deed with new conditions will be drafted and signed. Refinancing is cheaper than setting up a whole new mortgage, as the costs for the mortgage tax do not have to be paid (only the costs of the valuation and registration). However, with more than 1 million families in which all members were unemployed (Gutiérrez & Delclòs, 2016), refinancing was not an option for lenders, and the number of foreclosures grew exponentially[8].

Data

In this section, we analyse our dataset, which is a compilation of several different sources of data: a baseline dataset and data for social capital variables, that is, data on social networks and data from an experiment using a profile photo from social networks. Next, we explain every source of data separately.

Baseline Data

This dataset combines information from many different sources. First, we obtained market information from a residential real estate intermediary[9] with branches in most of the Spanish provinces. The company made approximately 4% of the total home sales on the open market in Spain in recent years[10]. We matched those residential units with information on the financial intermediary that provided the mortgage. These data, corresponding to residential mortgages originating between 2005 and 2007, include information on loan and borrower characteristics, including real estate appraisal value and lender identity.

For a sample of 3,307 observations, we could match this information with information on the financial intermediary. Therefore, for these houses, we had the demographic characteristics of the family and financial information in addition to the characteristics of the dwelling, including the actual market price. To improve our ability to merge data from different sources, we matched each residential unit with their unique identification number ('número catastral') using the database of the 'Catastro', the Spanish household registry, which is the institution in charge of the georeferenced cartography of all real estate assets in the Spanish territory. Using this identification number, we could improve the matching of the information described previously with the information of the Official

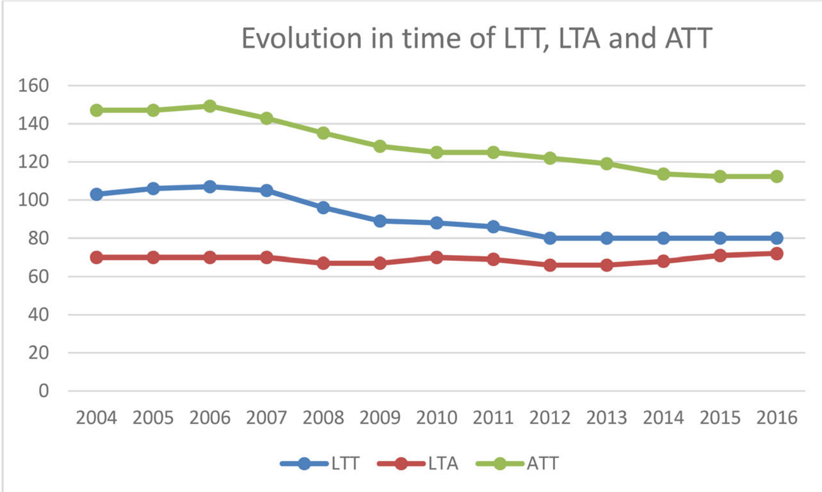

Figure 1. Evolution in time of LTT, LTA and ATT. *Source:* Bank of Spain. Bover et al. (2019).

Property Registry[11], which also includes the size of the loan, its characteristics and the price[12] that was declared to the Registry[13]. The dataset is unique in that, in addition to transaction price, financial information (mortgage amount conceded, appraisal price, spread, type of financial institution, etc.) and characteristics of the dwelling, it also includes information about the buyer (which includes the name of the mortgagor). Every observation pertains to a single transaction.

The average transaction price (156,005 €) is considerably lower than the appraisal value (195,214 €). As a result, the appraisal-to-transaction-price ratio (ATT) is, for the whole period, 1.29. This fact confirms an overappraising behaviour by financial institutions (Akin et al., 2014). The average loan-to-appraisal (LTA) value for the boom period was 87.52%. As a result of overappraisals, loan-to-transaction (LTT) was even higher. To put these figures into context, we report Figure 1, in which using information from Bover et al. (2019), we show the evolution of LTV, LTA and ATT from 2004 to 2016 in Spain. During our time period, overappraisal was even higher than in our dataset (over 40%). From 2007, ATT decreased until it reached 12% in 2016. As a result, the difference between LTT and LTA was reduced from almost 40 percentage points of difference (2007) to less than 10 (2016). Additionally, LTT was reduced from 107% (2006) to 80% (2016). That is, our time period is very unusual: rapid price appreciation, overappraisal, and mortgage competition lead to an increase in LTVs (LTT). Thus, although higher prices, all else-mortgage loan amounts equal, should reduce LTT, we observed higher LTT because the increase in the loan amount was higher than the price increase. That is, lenders granted riskier loans in this housing-boom period. In this sense, a higher ratio of loan value relative to house value indicates a higher level of softness and worse screening and quality control by the bank, which will result in a higher probability of default in the future. In fact, maximum loan-to-value (LTV) ratios on mortgages have been adopted by many countries as a macroprudential instrument to address systemic risk. However, Spain regulators did not observe this practice, as LTA was constant (or suffered less variation than LTT) throughout the period.

In terms of the price of the loan, we have two sources of information: the type of benchmark rate and the spread. Lenders use as benchmark interest rates either the reference interest for mortgage loans (RIML) or the Euribor. In our sample, 84% of the loans were priced using the Euribor. The spread is defined as the difference between the gross loan rate and the reference rate. The average spread is 0.85.

We include several borrower characteristics that enable us to infer the risk profile of the borrower. The first variable is household monthly income, with an average just above 1,550 €in both boom-and-bust periods. Second, our dataset provides information on the borrower's labour status and the type of contract, if occupied. From labour status information, we know whether the borrower is working in the private sector, is working in the public sector, is self-employed or is not working, and for those who are employed, the labour contract type information enables us to identify borrowers working with a permanent contract or with a temporary contract. Our sample mainly consists of active workers, and the share of borrowers working with a temporary contract is 35%. The loans in the sample period originated from 86 lenders. In terms of the characteristics of the house, apart from market price and appraisal value, we also have information on its location. We distinguish between coastal and interior provinces. Almost half of the houses purchased are in coastal provinces. The remaining variables in the equations are mainly controls and include other borrower characteristics (marital status, education, age and number of holders) and bank, year and location dummies. Including bank fixed effects permits us to control for the level of competition since the level of competition each bank faces affects both LTT and LTA.

In our dataset, we cannot account for borrowers' credit history (as is typical in Spain). Studies on mortgage price differentials in the US have found that credit history is statistically significant in models estimating mortgage lending. Borrowers must have an acceptable credit history to be granted a mortgage. However, we think that in our case, this omission is not that problematic. We should expect the potential bias caused by this omission, if there is any, to be fairly small. The main reason is that we control for most of the relevant variables that determine the risk profile of a borrower, such as the type of labour contract (indefinite or fixed-term), income, education, age, marital status and type of job. We think that these variables are excellent proxies for a borrower's credit history. One should expect credit history to be determined by past realizations of these variables. Indeed, we consider these variables to be better predictors of the probability of default than credit history[14].

Data on Social Capital Variables

To collect information on social capital variables, we first searched for information on the mortgagor in social networks. To this end, we have used not only the name but also the other available personal information in our dataset (age, nationality, city, marital status, etc.) to ensure that we have identified the correct individual.

We searched for information on Facebook, Twitter and LinkedIn[15]. We identified a subsample of 512 individuals from whom we would have information from at least one social network (usually Facebook[16]). We identified the entire subsample on Facebook but only 41 Twitter accounts and 102 LinkedIn accounts. We collected all information available from the whole period observed. The period ranges from the moment the user created the account (usually, approximately 2008) until 2019[17]. That is, in many cases the information about social capital variables is generated after origination, but even in this case is a valuable information. On the one hand, if the period of observation of the social capital variable is close to the one of origin, we can assume that the social capital variable is similar in both periods. This is the case of the information from Facebook, which is a social network very widespread in 2007. On the other hand, even though the period of observation of the social capital variable is not close to the one of origin, social capital variables can proxy for unobservables at the time of origination. This is the case of the information from LinkedIn, which is a social network popularised after 2010. For instance, a higher number of contacts in LinkedIn in 2013 can proxy unobservables in 2009 (such as the potential of the network of the borrower social status and reputation). As social status and reputation continually grows throughout the individual's professional career, make sense that social capital variables collected many years later (such as linkedin contacts) are reasonable proxies for these variables. Similar, we can say about social stigma cost which can be proxy from social media engagement (photos posted, number of likes …).

Then, we created a file for every mortgagor in which we compiled all the public information that we found: photos (classified by profile photos, photos from trips and other leisure activities and photos from restaurants), number of friends, contacts, number of posts, interactions, and update frequency. From this information, we computed several numeric variables for every individual. The variables finally included in the model are[18] as follows: the number of photos from trips and other leisure activities (with a mean of 9.01 photos, ranging from 0 to 117); the number of photos from restaurants (with a mean of 1.55 photos, ranging from 0 to 54); the number of LinkedIn contacts (with a sample mean of 15.39, ranging from 0 – no LinkedIn account – to 502 contacts) and the monthly average of Facebook likes (with a mean of 50.40, ranging from 0 to 11,875). Therefore, we computed not only stock variables (number of photos or LinkedIn contacts) but also flow variables (monthly average of Facebook likes).

To measure the effect of physical appearance on the mortgage market, we chose a typical profile photo from Facebook, that is, a photo in which the focus of the image is the face. If possible, an image with a neutral or muted background and natural light was chosen. The photo chosen was as close to the period of mortgage data (2005-2007) as possible. Since the borrowers created their Facebook accounts after 2006, the photo was usually from 2007-2010. The subjects were also rated on their physical attractiveness and the first impressions they make regarding their trustworthiness and creditworthiness, which might matter in a person-to-person interaction over and above the demographic characteristics. This is a similar situation to that faced by lenders. As we have shown, in Spain, 99% of loans were granted by deposit institutions. These institutions use face-to-face interviews with potential borrowers and do not deal with the social network information of the borrower. There is a large literature on beauty in the social sciences. Most people agree on the definition of attractiveness, and according to research in social psychology, they also agree on who is attractive and who is not (Feingold, 1992; Langlois et al., 2000). In addition, we define as trustworthy a recipient who sends a fair

Table 1. Characteristics of the mortgages*.

	Sample mean (SD)
Amount of the Loan (€)[1]	171,211
	(70,513)
Appraisal Value (€)[1]	195,214
	(69,813)
Loan to Value (%)	87.56
	(18.64)
Spread (%)	0.85
	(0.43)
Appraisal to Market Price (%)	128.93
	(22.25)
Market Price (€)[1]	156,005
	(60,073)
Spanish Housing price (€ per square meter)	1838.16
	(143,38)
Reference Interest Rate** (% **of total)	
RIML (ref)	15.16
Euribor	84.84
Financial Institution** (% **of total)	
Commercial bank	39.61
Savings Bank	51.62
Individually rescued[2]	8.56
FROB owned	30.65
Rest	12.41
Nonbank financial institutions	8.77
Number of observations	3,307

[1]Variables are in real terms.

share of money back to the sender, even if he or she has no obligation to do so (Ravina, 2008).

Similarly, we asked the raters to evaluate a closely related variable that is relevant in credit markets, which is the ability to repay (creditworthiness). Finally, we also asked the raters to evaluate professionalism. The rating procedure is used in the literature on beauty (Duarte et al., 2012) and works as follows. Each picture was evaluated by three female and three male raters, and the average rating was used for the analysis. Raters only have the profile photo as information. The rating is on a 7-point scale, ranging from "Extremely Attractive/Professional/Happy/Creditworthy/Trustworthy" to "Not Attractive/ Professional/Happy/Creditworthy/Trustworthy at All". Our raters were workers on Amazon's Mechanical Turk (MTurk) service and were unaware of the research objective[19]. We also checked differences in evaluation between raters from the MTurk service and real lenders for a small subsample of profile photos (see Appendix 1).

The attribute with the highest value was trustworthiness (4.01), while the attribute with the lowest value was attractiveness (3.47).

Table 1 displays the descriptive statistics.

Data Robustness Check

One the one hand, Ravina (2008) uses lenders as raters. Even though Duarte et al (2012) use the MTurk service for raters as we did in this paper, one criticism of our data is that Amazon's workers may not have the same preferences in evaluating attractiveness,

Table 1 (cont.): Characteristics of the borrowers.

	Sample mean (SD)
Labour Status	
Public sector employee	10.32
Private sector employee or self-employed	86.63
Nonoccupied	3.06
Type of Contract	
Permanent	62.17
Temporary	34.77
Marital Status	
Married	31.04
Single, widowed, separated	68.96
Education	
Compulsory	54.52
Secondary (noncompulsory)	33.4
University degree	12.07
Number of Holders	
One	31.71
Two	57
Three	11.01
Location	
Interior	57.7
Coastal	42.93
Income in real terms (thousand €)	1,563
	(0.65)
Age	33.77
	(9.18)
Number of observations	3,307

Percentages for categorical variables.

happiness, creditworthiness, trustworthiness and professionalism as real lenders. To test this hypothesis, we convinced 10 lenders to rate a small random sample (40) of our individuals (every lender rated 5 individuals, and we computed the mean of these evaluations). In Table 2, we present a mean comparison test. Although lenders rate attractiveness and professionalism higher and happiness, creditworthiness and trustworthiness lower, these differences are very low and are statistically insignificant[20]. Additionally, a researcher can be worried about the correlation between these variables and social network variables (i.e., attractiveness can be correlated with the number of likes). That is, multicollinearity is a concern. To check this point, we computed the variance inflation factor (VIF) in all regressions. The mean VIF value is approximately 2, and the maximum value is 3.61 (Table 3 creditworthiness). Values under 10 do not cause multicollinearity.

On the other hand, as we have seen, the sample shrinks when we include social media information. In this situation, a potential selection bias[21] in the reduced sample must be analysed. The absence of social network information in the rest of the sample might be informative in itself. That is, individuals for which we do not have information about their social networks (either because they do not use it or because they do not have any public information) might differ in important ways than those that do. To test this hypothesis, we estimated the models presented in the next section (Table 3), only with baseline information for the sample with (512 observations) and without social network information (3,307 observations). The results (Table A1) show that if there is any selection bias, it is very small. Although some differences emerged in the estimated coefficients, these differences were due to the lower efficiency of the estimates in the small sample

Table 1 (cont.) Social capital variables.

	Sample mean (Standard deviation)
Social Networks	
Number of leisure photos	9.01 (17.62)
Number of photos from restaurants	1.55 (3.92)
LinkedIn contacts	15.39 (71.37)
Facebook likes	50.40 (606.96)
Experimental Data	
Attractiveness	3.47 (0.94)
Trustworthiness	4.01 (0.91)
Creditworthiness	3.97 (0.92)
Professionalism	3.89 (1.03)
Number of observations	512

Facebook likes and the number of photos of leisure and restaurants are observed from 2006 to 2012, while linkedIn contacts is observed from 2010 to 2019.

(as seen in the higher standard deviations of all the coefficients, including the constant). However, the results in terms of significance are almost identical. Another explanation for the difference in coefficients is that individuals are, to some extent, different. For example, knowing that an individual doesn't have Facebook is informative and permits to better identify unobservables.

Results

Baseline Estimation

Our empirical strategy consists of estimating reduced-form equations on the determinants of LTT, LTA and ATT. Both loan principal and appraisal value may be chosen by the bank (in this sense, it can be endogenous), whereas the transaction value is not[22]. All the equations are estimated for the period 2005 to 2007, that is, the final years of the boom period. In all cases, we include the Spanish Housing Price per square meter (Ministry of Public Works)[23], location (province) and monthly time dummies. Additionally, we include a dummy that identifies the financial institution that granted the mortgage. We add spread, borrower characteristics and social capital variables. Social capital variables can capture unobservables as well as effects that are consequence of some degree of subjectivity in the lending process (Duarte et al., 2012). All in all, we estimate this specification:

$$\begin{aligned} LTT/LTA/ATT_i = {} & \alpha + \beta_1 Borrower\ Characteristics_i + \Sigma\beta_{2k} Financial\ institution_{ik} \\ & + \beta_3 Benchmark\ reference\ interest\ rate_i + \beta_4 Spread_i \\ & + \beta_5 Social\ Capital\ Variables_i + \Sigma\beta_{6t} Time_t + \Sigma\beta_{7j} Location_{ij} \\ & + \beta_3 Spanish\ Housing\ price_t \end{aligned} \tag{1}$$

Table 2. Mean comparison test among lenders' and MTurk workers' evaluations.

	Difference	p-values
Attractiveness	0.37	0.23
Trustworthiness	−0.04	0.45
Creditworthiness	−0.11	0.37
Professionalism	0.10	0.28

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3. Determinants of the LTT and LTA.

Variables	LTT	p-value	LTA	p-value	ATT	p-value
Coastal	10.92**	0.05	6.12	0.16	4.75	0.40
Spread	19.85***	0.00	7.97	0.14	11.30***	0.00
Spanish Housing price	0.06	0.11	0.06	0.17	0.01	0.86
Education						
Secondary (noncompulsory)	3.17	0.37	−0.13	0.96	2.87	0.36
University degree	3.17	0.54	−3.45	0.55	4.87	0.37
Age	−0.04	0.98	−2.64	0.28	1.95	0.20
Age^2	−0.004	0.84	0.04	0.29	−0.03	0.14
Number of Holders						
Two	6.95**	0.03	2.57	0.34	4.15	0.20
Three	7.44*	0.08	3.89	0.33	3.33	0.50
Spanish	−5.91*	0.09	−6.53*	0.08	0.71	0.82
Married	−1.25	0.71	0.91	0.75	−3.39	0.30
Public sector employee	−3.92*	0.06	−4.11	0.32	1.87	0.70
Unemployed	−10.09	0.38	−5.26	0.42	−10.40	0.15
Permanent	−2.83*	0.09	0.15	0.95	−4.40*	0.08
Income in real terms (thousand €)	−0.23	0.93	−1.44	0.54	1.66	0.41
Euribor	−21.31**	0.01	−1.54	0.81	−24.31***	0.00
Social Network Variables						
Number of leisure photos	−0.74	0.21	−0.18	0.72	−0.46	0.32
Number of photos from restaurants	−0.41	0.32	−0.67	0.38	−0.17	0.77
LinkedIn contacts	0.007	0.66	0.05	0.13	−0.04*	0.09
Facebook likes	−0.001*	0.08	−0.001***	0.00	0.0004	0.74
Attractiveness	1.74**	0.03	1.53**	0.05	0.08	0.94
Happiness	2.68***	0.00	2.31***	0.00	−0.69	0.47
Trustworthiness	−0.65	0.54	−0.17	0.87	−0.01	0.99
Creditworthiness	−2.25**	0.04	−2.09*	0.08	0.23	0.84
Professionalism	−1.81	0.12	−1.34	0.18	−0.73	0.43
Constant	109.10***	0.00	115.1***	0.00	123.1***	0.00
Bank, Year and Location controls	Yes		Yes		Yes	
Observations	512		512		512	
Adjusted R-squared	0.24		0.25		0.34	

Thus, in Table 3[24], we present the regression results of the loan-to-transaction-value equation and its components: the loan-to-(appraisal)-value equation and the overappraisal equation. First, we focus on social capital variables. The introduction of these variables increases the goodness of fit of the models[25]. In particular, the social capital variables slightly increase the adjusted R-squared value (from 0.19 to 0.23) in the case of LTT and LTA as well as in the case of the overappraisal model (from 0.28 to 0.32)[26]. In this sense, these variables capture previously unobserved information. As in Berg et al. (2018), we find that social network information complements rather than substitutes for traditional variables. The correlation between traditional variables and social capital variables is only approximately 5%. This suggests that a lender that uses information from both sources can make superior lending decisions.

Otherwise, once we control for the rest of the variables affecting LTT, LTA and ATT, the previous evidence is confirmed. Increasing LinkedIn contacts by 100 is associated with a reduction in overappraisal of 4 percentage points. As in "social credit score" algorithms, the number of LinkedIn contacts captures social status, so people with higher social status may have enough savings or wealth that they do not need an overappraisal to obtain a mortgage (Montalvo et al., 2020). In this respect, the introduction of LinkedIn contacts improves the information that the bank has about mortgagors. It seems that a mortgage given to an individual with a higher number of LinkedIn contacts is, all else constant, a better mortgage (in the sense that it is less overappraised). However, the number of photos of applicants engaging in leisure activities or in restaurants does not affect LTT, LTA or ATT. Finally, 1000 more Facebook likes in a month are associated with a decrease in LTT by 0.9 points (and LTA by 1.1 points). The more engagement a borrower has with his or her social media site, the less likely he or she is to ask for a higher LTT. A possible explanation is a higher social stigma cost in the case of these borrowers.

In the estimation models, the beauty premium is also observable. Attractive people are rewarded by a boost of almost 2 points in the loan-to-value ratio. That is, people with the same sociodemographic and financial situation show a 5-point increase in their loan-to-value ratio because of being perceived as happy (3 points) and attractive. On the one hand, this implies more credit; on the other hand, this loan can be riskier. Finally, a one-point-higher perception of creditworthiness is associated with a decrease in LTT and LTA by 2.5 points. This result can be explained by the fact that creditworthy people may be responsible and do not ask for a higher LTV.

In addition, in the bubble economy, workers with temporary contracts obtained the same LTVs (both statistically and economically speaking) during the boom as workers with permanent contracts. Additionally, borrowers who were not employed obtain identical loan-to-value ratios to employed borrowers in the boom (Akin et al., 2014). Despite the fact that the small size of the sample invites us to be cautious, these results seems to suggest overly soft lending standards and excessive risk-taking in the boom[27]. Variables that correlate with borrowers' ability to repay and bargain are insignificant. Only spread, the number of holders, the use of Euribor versus RIML and the characteristic of being Spanish are significant. Besley et al. (2013) argue that a regression on the loan principal (LTT) can explain whether lending standards are soft.

In this context, it is not surprising that traditional creditworthy variables do not positively affect the loan-to-value ratio; this fact is just another signal of a bad practice. Our findings also align with previous evidence on the Spanish boom (Montalvo & Raya, 2018). As we can observe, if the regulator wants to monitor the specific determinants of the ratio of loan to market value, such monitoring is biased if it is based on the loan-to-appraisal equation. Because of overappraisal, the regulator cannot monitor important determinants of the ratio of loan to market price that are not significant in the loan-to-appraisal equation. This is the case for the variables for coastal location, labour status (and the type of contract, which is significant at 10%), having the Euribor as the reference interest rate, the spread and the number of mortgage holders[28]. In all these cases, the variables are significant in the LTT equation but not in the LTA equation. For instance, working in the private or public sector reduces the ratio of loan to market price relative to being unemployed. Furthermore, we find that a higher spread does not impact LTA. In summary, the effect of the determinants of the loan-to-value ratio is generally underestimated when we use the appraisal value rather than the market price.

Concluding Remarks

In this paper we have obtained evidence about an increase in the goodness of fit of the models with the introduction of the social capital variables, improving, therefore, the credit-risk screening process. Thus, the number of contacts in LinkedIn helps to identify overappraising. In this respect, this variable seems to capture social status and reputation in a causal model, as in social credit score models. However, scoring methods that use mainly internet data face the possibility of manipulation. In this sense, this result is useful, but only once we have controlled for the rest of the traditional scoring variables (Einav et al. 2012). Additionally, the more engagement a borrower has with his or her social media site, the less the LTV (Ge et al., 2017, Iyer et al., 2015). One possible reason is that these borrowers care about social stigma costs. Thus, they are less likely to ask for a loan with a higher LTV. Favourable evidence of the existence of beauty premium is also obtained in regard to mortgage loans. Also, a negative effect of creditworthiness on the LTV evaluated with regard to the appraisal value can be interpreted as additional evidence of bad practices in the Spanish mortgage market during the boom years.

The results of the analyses reported in this paper suggest that the potential bias that might arise from the omission of variables, if there is any, should be modest (Oster, 2019). However, all the interpretations made on the paper should be tempered. On the one hand, external validity is needed, as data were collected during boom years. On the other hand, the results would benefit from actual loan default information. Additionally, a larger sample size, different definitions of social capital variables or simply extrapolating the analysis to other housing markets will help to confirm these results. In any case, at least this paper shows that some unobserved effect that is correlated with social capital variables might lead to higher or lower LTV, a basic variable for macroprudential policies, although further research is needed.

Notes

1. Fair Isaac Corp. The exact formula used is unknown but the following are, approximately, the weights of each component: history of previous payments (35%), use of credit (30%), length of credit history (15%), types of credit used (10%), amount of credit obtained recently (10%).
2. Zest Finance has found that loan applicants who only use uppercase or lowercase letters are less likely to repay loans.
3. The impact of other nontraditional variables used to improve models of credit default can be shown in (Jagtiani & Lemieux, 2018; Khandani et al., 2010; Kruppa et al., 2013; Netzer et al., 2019). These cases include variables from machine learning and big data.
4. This includes the uses of a French repayment method, which is a well-known method characterized by payment of constant instalments in each period wherein Interest decreases as loan periods pass and amortized capital increases in every new period.
5. Two examples of typical mortgage product during boom years. First is from the Banco de Santander during 2007. Reference index: euribor. Spread: 0.25 (4.97%). Term: 30 years Annual interest rate review. 0.5% of early amortization commission. Financing up to 80% of the housing value. Fixed payment amortization system. Offer conditional on payroll direct debit and hiring home insurance. The second is from Barclays during 2005. Reference index: euribor. Spread: 0.45 (2.90%). Term: 30 years Annual interest rate review. 0.5% of early amortization commission. Offer conditional on payroll direct debit and hiring home insurance. Fixed payment amortization system.
6. See Cecchetti et al. (2011), in particular Table A2.1.

7. This can be done in a number of ways. First, when interest rates are as low as they are now, even homeowners with relatively recent mortgages may benefit from refinancing. Second, to take advantage of better personal financial circumstances (if you have more savings or a higher income, you may qualify for better credit conditions). Finally you can also change mortgage conditions, refinancing into fixed rate mortgages or negotiating a longer mortgage term to reduce monthly repayments (to alleviate worse financial circumstances).
8. Raya (2018) examines the determinants of foreclosures in Spain.
9. For confidentiality reasons, we cannot identify the company.
10. This number excludes social housing and residential units that had some type of public subsidy.
11. 'Registry' refers to the Spanish Official Property Registry ('Registro de la Propiedad'), which archives the property titles of all real estate assets.
12. Notice that prices declared to the real estate registration office do not have to coincide with market prices since there is an extended practice of using money that is undeclared to the tax authority as part of the payment in real estate transactions.
13. To be sure that the matching was properly performed, we compared the common variables available in our constructed dataset and the information from the Official Property Registry (size of the loan, appraisal price).
14. A decrease in the accuracy of credit scores based on borrowers' credit history for predicting loan delinquency has been proved in a Fitch study in the US. Indeed, some banks have abandoned credit scores for other risk analyses based on, among other factors, borrowers' employment. These more reliable scores use as inputs variables such as the ones we use here regarding borrowers' labour status.
15. We also checked other social networks, such as Instagram, but fewer than 2% of the individuals had an active Instagram account.
16. In 2017, Facebook had 1.97 billion monthly active users. The seventh-ranked photosharing app, Instagram, had over 600 million monthly active accounts. LinkedIn had 106 million monthly active accounts.
17. In particular, Facebook information is observed from 2006 to 2012 while LinkedIn information is observed from 2010 to 2019.
18. Other variables computed, such as frequency of profile updates or the number of friends, depend crucially on the amount of information the individual sets to be publicly accessible. We also compiled information on Twitter and Instagram (followers, number of posts, etc.), but, as we have pointed out previously, the sample was very small.
19. The Cohen's kappas and t tests suggest a fairly high level of agreement across raters, although for every item, raters 1 and 5 were relatively generous, while rater 6 was relatively ungenerous.
20. A variance comparison test also could not reject the null hypothesis of equal variance.
21. In fact, the major source of drop outs were people with a private social network profile. Thus, the selection bias, if any, is not due to differences in activity.
22. Note that loan-to-transaction equations are models that also approximate the probability of future default, because the correlation between the loan-to-value ratio and the future probability of default is well known (Wong et al., 2011). In this sense, sociodemographic and social capital variables at time "t" are implicitly explaining default at time "$t+n$"; i.e., we are using social capital data to predict loan performance. We have no information about the effective future default of these loans. However, we know for a very small subsample whether the dwelling ends in foreclosure. We present some results in this respect in the Appendix 2 (Tables A.3 and A.4)
23. The variable is not statistically significant. In fact, the housing price trend is captured by the time dummies. We have excluded this variable in the models presented in the Appendix.
24. In the Appendix 1 (Table A.2) we present, for a subsample, the same estimation adding text analysis variables as predictors of LTT, LTA and ATT.
25. Models without social capital variables are presented in Table A.1 of the Appendix.

26. We have also computed the root of mean squared error (RMSE). In all cases, the difference between the models with and without social capital variables is similar to that found in terms of R-squared. That is, the LTT is reduced from 19.05 to 16.86 when we introduce social capital variables. For LTA, the reduction is from 19.14 to 17.3, and for ATT, RMSE is reduced from 19.06 to 16.7.
27. In this respect, in Table A.1, we present results of the estimation where we do not yet include social networks indicators. Again, variables like education, age, income, permanent occupation are all statistically insignificant.
28. In fact, the only explanatory variable that is significant in both equations is the characteristic of being from Spain, which reduces LTA and LTT. Diaz-Serrano and Raya (2014) explore discrimination in the Spanish mortgage market.
29. Similarly, Dorfleitner et al. (2016) examine information derived from the description texts to the probability of successful funding and to the default probability in peer-to-peer lending for two leading European platforms.
30. Obviously, with 121 observations, we have a problem of representativeness. However, as will be seen below, we obtain additional evidence from a recent literature result.
31. For 1 observation, we do not know the selling price.
32. For the whole sample of 3,307 observations, the number of homes is 318 (65 of which were foreclosed upon).

Acknowledgements

The author thanks Claudia Vargas, Elena Fernández and Helena Bonnin for excellent research assistance, especially with respect to data compilation

Data

The data that support the findings of this study are available from a Real Estate Company (REC). Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors with the permission of the REC.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

The author acknowledges the MCINN (ECO2016-78816-R) for financial support. Ministerio de Ciencia, Innovación y Universidades.

References

Agarwal, S., Ambrose, B. W., Chomsisengphet, S., & Liu, C. (2011). The role of soft information in a dynamic contract setting: Evidence from the home equity credit market. *Journal of Money, Credit and Banking*, *43*, 633–655. https://doi.org/10.1111/j.1538-4616.2011.00390.x

Agarwal, S., & Hauswald, R. (2010). Distance and private information in lending. *Review of Financial Studies*, *23*, 2757–2788. https://doi.org/10.1093/rfs/hhq001

Akin, O., Montalvo, J. G., García Villar, J., Peydró, J.-L., & Raya, J. M. (2014). The real estate and credit bubble: Evidence from Spain. *SERIEs*, *5*, 223–243. https://doi.org/10.1007/s13209-014-0115-9

Avery, R. B., Calem, P. S., & Canner, G. B. (2004). Consumer credit scoring: Do situational circumstances matter? *Journal of Banking & Finance*, *28*, 835–856. https://doi.org/10.1016/S0378-4266(03)00202-4

Bandyopadhyay, A. (2021). *Quantifying soft information, mortgage market efficiency & asset pricing implications*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3755698.

Berg, T., Burg, V., Gombović, A., & Puri, M. (2018). *On the rise of FinTechs–Credit scoring using digital footprints*. National Bureau of Economic Research.

Besley, T., Meads, N., & Surico, P. (2013). Risk heterogeneity and credit supply: Evidence from the mortgage market. *NBER Macroeconomics Annual*, *27*, 375–419. https://doi.org/10.1086/669185

Bover, O., Torrado, M., & Villanueva, E. (2019). The loan to value ratio for housing in Spain over the period 2004-2016. *Banco Espana Artic*, *3*, 19.

Cecchetti, S. G., Mohanty, M. S., & Zampolli, F. (2011). *The real effects of debt*. BIS Working Papers 352

Claessens, S., Ghosh, S. R., & Mihet, R. (2013). Macro-prudential policies to mitigate financial system vulnerabilities. *Journal of International Money and Finance*, *39*, 153–185. https://doi.org/10.1016/j.jimonfin.2013.06.023

Crowe, C., Dell'Ariccia, G., Igan, D., & Rabanal, P. (2013). How to deal with real estate booms: Lessons from country experiences. *Journal of Financial Stability*, *9*, 300–319. https://doi.org/10.1016/j.jfs.2013.05.003

De Cnudde, S., Moeyersoms, J., Stankova, M., Tobback, E., Javaly, V., & Martens, D. (2019). What does your Facebook profile reveal about your creditworthiness? Using alternative data for microfinance. *Journal of the Operational Research Society*, *70*, 353–363. https://doi.org/10.1080/01605682.2018.1434402

Diaz-Serrano, L., & Raya, J. M. (2014). Mortgages, immigrants and discrimination: An analysis of the interest rates in Spain. *Regional Science and Urban Economics*, *45*, 22–32. https://doi.org/10.1016/j.regsciurbeco.2013.12.004

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., & Kammler, J. (2016). Description-text related soft information in peer-to-peer lending–Evidence from two leading European platforms. *Journal of Banking & Finance*, *64*, 169–187. https://doi.org/10.1016/j.jbankfin.2015.11.009

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies*, *25*, 2455–2484. https://doi.org/10.1093/rfs/hhs071

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., Longo, L. C. (1991). What is beautiful is good, but ... : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109. https://doi.org/10.1037/0033-2909.110.1.109

Einav, L., Jenkins, M., & Levin, J. (2012). Contract pricing in consumer credit markets. *Econometrica*, *80*, 1387–1432.

Ergungor, O. E., & Moulton, S. (2014). Beyond the transaction: Banks and mortgage default of low-income homebuyers. *Journal of Money, Credit and Banking*, *46*, 1721–1752. https://doi.org/10.1111/jmcb.12164

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, *94*, 334. https://doi.org/10.1037/0022-3514.94.2.334

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, *111*, 304. https://doi.org/10.1037/0033-2909.111.2.304

Gao, Q., Lin, M., & Sias, R. W. (2018). Words matter: The role of texts in online credit markets. *Journal of Financial and Quantitative Analysis*. Advance online publication. http://dx.doi.org/10.2139/ssrn.2446114.

Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems*, *34*, 401–424. https://doi.org/10.1080/07421222.2017.1334472

Gonzalez, L., & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, *2*, 44–58. https://doi.org/10.1016/j.jbef.2014.04.002

Guiso, L., Sapienza, P., & Zingales, L. (2013). The determinants of attitudes toward strategic default on mortgages. *The Journal of Finance*, *68*, 1473–1515. https://doi.org/10.1111/jofi.12044

Gutiérrez, A., & Delclòs, X. (2016). The uneven distribution of evictions as new evidence of urban inequality: A spatial analysis approach in two Catalan cities. *Cities*, *56*, 101–108. https://doi.org/10.1016/j.cities.2016.04.007

Iyer, R., Khwaja, A. I., Luttmer, E. F. P., & Shue, K. (2015). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, *62*, 1554–1577. https://doi.org/10.1287/mnsc.2015.2181

Jagtiani, J., & Lemieux, C. (2018). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, *48*(4), 1009–1029.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*, 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

King, B. (2014). *Breaking banks: The innovators, rogues, and strategists rebooting banking*. John Wiley & Sons.

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications.*, *40*, 5125–5131. https://doi.org/10.1016/j.eswa.2013.03.019

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390. https://doi.org/10.1037/0033-2909.126.3.390

Li, L., Ucar, E., & Yavas, A. (2020). Social capital and mortgage delinquency. *Journal of Real Estate Finance and Economics*. Advance online publication. https://doi.org/10.1007/s11146-020-09775-4.

Lim, C. H., Costa, A., Columba, F., Kongsamut, P., Otani, A., Saiyid, M., Wezel, T., & Wu, X. (2011). *Macroprudential policy: What instruments and how to use them? Lessons from country experiences*. SSRN. https://ssrn.com/abstract=1956385 or http://dx.doi.org/10.2139/ssrn.1956385.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, *59*, 17–35. https://doi.org/10.1287/mnsc.1120.1560

Mayordomo, S., Rachedi, O., & Moreno, M. R. (2020). *Bank regulatory capital arbitrage: evidence from housing overappraisals*. https://www.norges-bank.no/contentassets/f7f999e4d23c4dcc8c95f266ffdb81fe/overappraisal_20201031.pdf

Montalvo, J. G. (2014). El impacto del "Big Data" en los servicios financieros. *Papeles de Economía Española*, 43–58.

Montalvo, J. G., Piolatto, A., & Raya, J. (2020). Transaction-tax evasion in the housing market. *Regional Science and Urban Economics*, *81*, 103526. https://doi.org/10.1016/j.regsciurbeco.2020.103526

Montalvo, J. G., & Raya, J. M. (2018). Constraints on LTV as a macroprudential tool: A precautionary tale. *Oxford Economic Papers*, *70*, 821–845. https://doi.org/10.1093/oep/gpy007

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, *56*(6), 960–980.

Ostendorf, S., Müller, S. M., & Brand, M. (2020). Neglecting long-term risks: Self-disclosure on social media and its relation to individual decision-making tendencies and problematic social-networks-use. *Frontiers in Psychology*, *11*, 543388. https://doi.org/10.3389/fpsyg.2020.543388

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*, 187–204. https://doi.org/10.1080/07350015.2016.1227711

Ravina, E. (2008). *Love & loans: The effect of beauty and personal characteristics in credit markets*. SSRN. http://dx.doi.org/10.2139/ssrn.1107307

Raya, J. M. (2018). The determinants of foreclosures: Evidence from the Spanish case. *Papers in Regional Science*, *97*(4), 957–970. https://doi.org/10.1111/pirs.12316

Saengchote, K. (2013). *Soft information in the subprime mortgage market* [Paper presentation]. 26th Australasian Finance and Banking Conference.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., & Seligman, M. E. P. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One*, *8*, e73791. https://doi.org/10.1371/journal.pone.0073791

Sumner, C., Byers, A., & Shearing, M. (2011). Determining personality traits & privacy concerns from facebook activity. *Black Hat Briefings*, *11*, 197–221.
Turkyilmaz, C. A., Erdem, S., & Uslu, A. (2015). The effects of personality traits and website quality on online impulse buying. *Procedia - Social and Behavioral Sciences*, *175*, 98–105. https://doi.org/10.1016/j.sbspro.2015.01.1179
Wanzare, N., & Scheule, H. (2010). *The impact of hard and soft information on mortgage default* [Unpubl MA thesis, Univ].
Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*, 189–202. https://doi.org/10.1177/106591290605900202
Wong, T., Fong, T., Li, K., & Choi, H. (2011). *Loan-to-value ratio as a macroprudential tool-Hong Kong's experience and cross-country evidence*. Syst. Risk, Basel III, Financ. Stab Regul.
Woolcock, M. (1998). Social capital and economic development: Toward a theoretical synthesis and policy framework. *Theory and Society*, *27*, 151–208. https://doi.org/10.1023/A:1006884930135

Appendix 1. Text analysis variables as predictors of LTT, LTA and ATT

Bandyopadhyay (2021) provides a novel framework for machine learning models to incorporate quantified nontraditional variables during the life of a loan. Additionally, Netzer et al. (2019)[29] and Gao et al. (2018) point out that the text that borrowers write in their loan request can significantly help predict loan default even when accounting for financial and demographic measures. This result is consistent with the idea that people who differ in the way they think and feel also differ in what they say and write about those thoughts and feelings (Fast & Funder, 2008). As a final exercise, we compiled text information from social networks using the Linguistic Inquiry and Word Count (LIWC) engine. LIWC, a popular tool in psychology, reveals our thoughts, feelings, personality, and motivations from words used in social networks. Basically, it reads a given text and counts the percentage of words that reflect different emotions, thinking styles, social concerns, and even parts of speech. Using LIWC, Sumner et al. (2011) explore the extent to which it is possible to determine personality traits and privacy concerns based on Facebook use. Additionally, Schwartz et al. (2013) apply LIWC to a dataset of over 15.4 million Facebook messages collected from 75 thousand volunteers to find language features that distinguish demographic and psychological attributes.

Table A.1. Determinants of the LTT and LTA. Model without social capital variables.

Variables	LTT	LTA	ATT	LTT	LTA	ATT
Coastal	9.145	5.075	4.096	1.627	0.149	0.763
Spread	18.89**	6.266	11.51***	6.570***	2.058*	4.206***
Education						
Secondary (noncompulsory)	3.180	0.686	1.646	−0.951	0.964	−3.255
University degree	−6.482	3.778	−1.193	−2.137	0.911	−4.059*
Age	0.0553	−2.777	2.094	−0.148	0.0592	−0.401
Age^2	−0.00421	0.0413	−0.0341	−0.00249	−0.00260	0.00266
Number of Holders						
Two	6.921**	3.202	4.224	5.492***	3.827***	0.169
Three	8.945**	6.081*	2.711	7.938***	7.385***	−0.299
Spanish	−5.224*	−2.537	−2.912*	−4.286***	−1.538**	−2.174**
Married	0.195	1.840	−2.784	1.647	1.400	−0.732
Public sector employee	−3.324	−2.633	0.953	−2.589	−2.158	−0.972
Unemployed	−7.221	−1.507	−11.27	−2.226*	−0.0227	−3.921
Permanent	−2.089	0.387	−4.088*	−3.304	0.478	−2.991*
Income in real terms (thousand €)	0.543	−0.747	1.676	1.830*	0.368	1.459
Euribor	−21.73**	−1.797	−22.89***	−9.347***	−2.762***	−6.776***
Constant	112.6***	95.8**	125.6***	119.7***	79.79***	155.2***
Bank, Year and Location controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	512	512	512	3,307	3,307	3,307
Adjusted R-squared	0.19	0.19	0.28	0.21	0.23	0.31

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A.2. Determinants of LTT, LTA and ATT, including personality variables.

Variables	LTT	LTA	ATT
Coastal	1.814382	−8.427668	13.8378*
Spread	14.46586*	3.234433	9.610315
Education			
Secondary (noncompulsory)	5.262762	−0.963420	5.036156
University degree	6.740875	−5.589021	6.859027
Age	−0.3617292*	−0.153557	−0.1886835
Number of Holders			
Two	−5.258003	−3.85881	−1.957763
Three	−0.5237927	13.08109	−10.0028
Spanish	−0.5926987	3.164296	−3.07260
Public sector employee	9.811015*	6.010578	6.62104
Unemployed	20.53622*	22.55828*	−6.00025
Permanent	0.953179	−0.936629	2.53272
Income in real terms (thousand €)	2.64524	−0.123478	2.883234
Euribor	−10.2021	0.902224	−9.59948
Social Capital Variables			
Attractiveness	5.941562**	4.290032	1.83237
Trustworthiness	−2.99439	0.746590	−2.4443
Creditworthiness	−11.19218**	−8.22209	−7.25078
Professionalism	7.408148**	4.588118	3.396505
Facebook likes	−0.001369	−0.001309*	0.000222
Personality			
Analytic Thinking	−0.046566***	−0.0320374**	−0.0172055
Clout	−0.1795581	−0.1556448	−0.1202104
Authenticity	0.3799114 **	0.3229069	0.0389359
Emotional tone	−0.0329363	−0.1751197	0.0922938
Constant	127.9842 ***	91.2375 ***	158.0275***
Bank, Year and Location controls	Yes	Yes	Yes
Observations	121	121	121
Adjusted R-squared	0.422	0.410	0.508

Using LIWC, we compiled information for a subsample of the 512 mortgagors for whom we have social network information. In particular, 121[30] individuals have enough words to obtain indicators about their personality. The engine counts I-words (I, me, my), social words, positive and negative emotions and cognitive processes and compares the counts with the average in the social network. From this information, four personality variables are computed – analytic thinking, clout, authenticity and social tone – and are again compared with the social network average. We have introduced this information in our model in Table A.2. Even when controlling for demographic, financial and social capital variables, being analytic reduces LTT and LTA, and being authentic increases LTT. Personality traits explain loan decisions. Berg et al. (2018) document that variables that serve as proxies for character and reputation are also significantly related to future payment behaviour. This result is consistent with marketing research documenting the importance of personality traits in impulse shopping (Turkyilmaz et al., 2015).

Appendix 2. Social Capital variables as predictors of foreclosure

We also performed an exercise using the 323 homes[31] (out of the total of 3,307) sold by the real estate company that were financed by the bank for which we had data on the housing stock.

Both datasets included the exact address of the home. Therefore, by merging them, we were able to determine whether each mortgage had ended in foreclosure. In all, 20.43% of these homes ended in foreclosure (66). This figure seems quite high. However, it is worth noting that in this case, the denominator is new mortgages during peak years of the boom. The first exercise we conducted is just a robustness check. We have used LTV as a proxy of future loan performance. Table A.3 shows the marginal effect of LTT on the probability of foreclosure. One percentage point more of LTT increases the probability of foreclosure by 0.0086 (significant at the 5% level). That is,

Table A.3. Marginal effect of LTT on the probability of foreclosure. Linear probability model.

Variables	Foreclosure
LTT	0.0085**
Constant	−0.7359
Observations	234
Adjusted R-squared	0.03

Robust standard errors in parentheses.

Table A.4. Mean comparison test among mortgages that ended and did not end in foreclosure.

	Difference
Number of leisure photos	−4.28
Number of photos from restaurants	−0.62
LinkedIn contacts	−1.17
Facebook likes	−6.74
Attractiveness	−0.45*
Trustworthiness	−0.12
Creditworthiness	−0.40*
Professionalism	−0.33*
Predictions of Estimated Models	
With social capital variables	–3.90*
Without social capital variables	−2.19*

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

an increase of 10 percentage points in LTT (for example, from 70 to 80) increases the probability of foreclosure by 0.086. As stated, the mean probability of foreclosure in this sample is 0.20. Therefore, 10 percentage points of LTT increased the probability of foreclosure by 43%.

Finally, for only 48 homes (12 ended in foreclosure)[32], we also know the social capital information of mortgagor. Table A.4 presents a mean comparison test of the social capital variables between mortgages that ended and did not end in foreclosure. Although due to the small sample size, only attractiveness, professionalism and creditworthiness are statistically significant, in all cases, mortgagors of homes that ended in foreclosure show lower mean values on these variables. The final two rows show the same mean comparison test using the forecast of the LTT estimated model with and without social capital variables. Again, in both cases, the value of LTT is higher for the homes that ended in foreclosure (both significant at the 10% level). In addition, this difference is higher in the case of the models including social network variables, which can be interpreted as additional evidence that these models offer better predictions of future loan performance.