Asymmetric information in loan contracts: New evidence from Italian big data

Francesco Benvenuti*, Monica Billio†, Michele Costola‡, Marco LiCalzi§

August 29, 2023

Abstract

We provide new empirical evidence for an open problem regarding asymmetric information in loan contracts: the effect of higher collateral requirements on the interest rates applied by banks to borrowers is not clear under such asymmetry. Previous literature has argued both for positive and negative links, based on different models and econometric analyses. Our purpose is to examine recent Italian data, analyzing for the first time big data for thousands of borrowers collected by means of a European Central Bank project. First, we apply an unsupervised analysis to this unique microeconomic data set with a focus on the link between interest rate and loan to value. We do not find evidence of a strong link between those variables. Then, we develop a game-theoretic model, which supports the empirical results, based on the principal-agent problem adapted to the specific case of loans. Finally, we analyze loan data in a supervised setting, controlling for different borrowers' categorical variables related to the interest rate determination. We conclude that the interest rate in loan contracts is influenced by asymmetric information and that higher collateral is not necessarily associated with a lower interest rate.

JEL Classification: C55; C57; C72; D12; H81.

Keywords: Collateral; interest rate; large data sets: modeling and analysis; principal-agent models; principal components.

*CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. Corresponding author's email: fraben@econ.au.dk (F. Benvenuti).

†Università Ca' Foscari Venezia, Department of Economics.

‡Università Ca' Foscari Venezia, Department of Economics.

§Università Ca' Foscari Venezia, Department of Management.

1 Introduction

Does collateral increase or decrease interest rates of loan contracts? The answer to this apparently simple question is not straightforward, and the topic has been broadly discussed both in the economic and applied econometric literature. However, depending on the particular assumptions or on the analyzed data, contradictory conclusions have been reached. We provide an overview of that in Section 2, where our extensive literature review sums up how different game theoretical models imply either a positive or a negative link between interest rates and collateral. This also holds true for the empirical studies that have tried to sort out the problem, because their results ended up being contradictory, depending on the considered data.

The present work[1] concentrates on this problem, examining what recent Italian data suggests, and compared to the past empirical literature on the subject, we base our findings on a relevant amount of microeconomic data. The database that we analyze contains millions of data (i.e. millions of borrowers' data) associated with a residential mortgage-backed security (RMBS). This unique data is available for the first time thanks to the huge collection of loan contracts stored in the European DataWarehouse (ED) database.[2] Therefore, the loans considered in this work are contracts underlying Asset-Backed Securities (ABS); the recent storage of this loan data is motivated by the ABS loan level initiative, which aims to improve the transparency of ABS markets, by making the market participants able to access this information. This initiative was conceived by the European Central Bank.

Our unique data contains both numerical and specific categorical variables, such as the repayment method or the borrower employment status (see Section 3 for a complete list and description of them), which clearly influence the decision maker (e.g. the bank) on the interest rate to apply. In order to extract information from the numerical data, we first look at some of its descriptive features, then we apply an unsupervised analysis to those variables, in particular to the interest rate and the loan to value of each contract. When splitting the data into clusters, we do not find strong evidence in favor of a negative link between interest rates and collateral. We additionally employ the principal component analysis, which exploits the correlation structure between different borrowers' features, namely the amount of collateral he can pledge and the sum he can borrow. We collect the results in biplots, which are easily interpretable. We show how a higher primary income is associated with higher collateral and shorter loans, while borrowers who receive higher amounts have a better income and provide more collateral, as expected.

In order to justify our empirical findings, we build a game-theoretic model which shows the a priori possibility of an ambiguous effect of collateral requirements on the interest rates of loan contracts due to the influence of asymmetric information. Our model is based on the principal-agent theory, and it deals with three different scenarios: perfect information, moral hazard, and adverse

[1] Part of this work and ideas have been developed during my MSc thesis (unpublished).

[2] The ED "*is the first centralised platform in Europe which collects, stores and distributes standardised ABS loan level data*". See the ED website for a detailed description: www.eurodw.eu.

selection. We show how the link between collateral and interest rates can be inconsistent, since it depends on the setting considered. In particular, interest rates can be higher even with more collateral under moral hazard, while the common intuition of a lower interest rate associated to more collateral holds true under perfect information and adverse selection. This is a consequence of including a cost function in the overall payoff of the borrower, related to his effort at investing the sum borrowed, which makes it possible for the principal to use higher collateral as a tool to direct the borrower's behavior, rather than only as a guarantee for the sum lent.

Compared with past literature, our model is based on few financial and game-theoretic principles, rather than on specific aspects of loan contracts. In this sense, our approach is fairly intuitive.

Our previous findings are supported by the final supervised analysis, which includes the numerical and the categorical variables in a regression model and considers three different specifications driven by the amount of data available. We find that the pricing of interest rate is influenced by key borrowers' features, and that higher collateral is not always a guarantee for the borrower to obtain a better interest rate. Overall, we argue that the open question posed in the beginning does not have a straightforward answer: information asymmetries can always influence loan contracts, with an effect on the collateral-interest rate relation. Our paper adds a contribution to the empirical literature on this problem and justifies the different findings of previous studies.

The paper is organized as follows. In Section 2, we present a comprehensive literature review. In Section 3, we describe the data set and we perform the unsupervised analysis. In Section 4, we build a game theory model to support our findings, while in Section 5 we apply a supervised analysis to our loan data. In Section 6, we conclude.

2 Literature review

In this section, we present a thorough literature review on the open problem studied in this paper. Given the abundance of results on the topic, a comprehensive overview is needed as a reference point for our discussion, since we will claim that the effect of collateral on interest rates is not unambiguous. Our model of Section 4 is more intuitive and less specific than most of the works mentioned below, since it takes into account only a few specific features of loan contracts, and it develops the discussion on simple intuitive principles. In this sense, it incorporates different conclusions reached by previous literature; we refer to the game-theoretic section below for details. However, it shares with past literature the main point: asymmetric information affects financial contracts and specifically loan contracts.

A seminal work regarding the effect of imperfect information in the credit market and the role of collateral and interest rates is Stiglitz and Weiss [1981], where borrowers with higher wealth and who can provide more collateral, are also prone to invest in high risk projects, decreasing the advantage of banks. This paper has not gone without criticism and it has triggered off a large amount of literature on the relationship between collateral and risk in loan contracts. A different view is indeed proposed in Bester [1985, 1987] and Wang [2010], where the possibility of different combinations of interest rate and collateral to sort borrowers is recovered. More recently, Su and Zhang [2017] argue that collateral is largely employed in loan contracts, hence the credit rationing as described by Stiglitz and Weiss hardly ever occurs in practice, and Flatnes and Carter [2019] demonstrate that moral hazard is heavily reduced through collateral requirements.

The Stiglitz and Weiss' model has been reconsidered in Coco [1999] to account for different risk attitudes of borrowers, in Broll and Gilroy [1986] who explain the dynamics of the credit market under asymmetric information with a greater focus on collateral requirements rather than on the interest rate, and in Bieta, Broll and Siebe [2008] who conclude that collateral cannot serve to reduce the typical asymmetric information of loan contracts.

On the other side, some parts of the literature have discussed different assumptions compared to the Stiglitz and Weiss' model: Chan and Thakor [1987] examine the case when all rents accrue to borrowers rather than to depositors, while De Meza and Webb [1987] account for different expected returns between projects. With a slightly different approach, Chan and Kanatas [1985] investigate whether the existence of collateral is justified outside of a moral hazard framework, examining a situation of asymmetric payoff valuation between borrowers and lenders, concluding that better borrowers should pledge a higher amount of collateral in order to take advantage of a lower interest rate. On the contrary, according to De Meza and Southey [1996] higher collateral is required from high-risk borrowers, as in Rajan and Winton [1995], who propose a model where collateral requirements increase when the borrower experiences financial difficulties.

Another point of the discussion has been the substantial difference in the collateral value for borrowers and for banks, constituting a disincentive to request collateral [Besanko and Thakor, 1987; Booth, Thakor and Udell, 1991]. A related discussion [Benjamin, 1978] concentrates on the additional costs for stipulating secured loan contracts. Despite these costs, it is shown that collateral can be useful in order to enforce debt contracts, as previously discussed in Barro [1976].

Due to this extensive amount of contrasting theoretical conclusions, other authors have turned to an empirical analysis to examine if any hypothesis is supported by real data. However, their results are also contradictory, depending for instance on the different methods, countries, periods, and models specified.

We mention just a few relevant examples. Berger and Udell [1990] present a cross-section analysis using a sample of over one million commercial loans from the Federal Reserve's Survey of Terms of Bank Lending. They regress the loan risk premium on collateral and other control variables at different periods, finding more frequently a positive coefficient for the former regressor. A similar conclusion is reached by Leeth and Scott [1989]. Taking a sample of 1,000 U.S. small business loans, they associate higher default probability and greater loan size and loan maturity, all risk indicators, to secured loans. Later, Angbazo, Mei and Saunders [1998] find empirical support for this conclusion, using over 4,000 loan transactions registered on Loan Pricing Corporation's database, between 1987 and 1994. Their conclusion is that collateral is generally related with riskier loans and riskier borrowers. A confirmation of this result is obtained in Jiménez and Saurina [2004]. In analyzing the determinants of the default probability (PD) of bank loans, they discuss extensively the role of collateral in this context. Their empirical results, using data from the Credit Register of the Bank of Spain, suggest a higher PD for collateralized loans.

However, an opposite conclusion is reached by Degryse and Van Cayseele [2000], who use data from Belgian banks and find a negative relationship between interest rates and collateral, even if the latter is decreasing as the duration of the bank–firm relationship increases. This is supported by Capra, Fernandez, Ramirez-Comeig et al. [2005]. They use a sample of small and medium-sized firms to test the role of collateral, and their analysis supports the hypothesis of a negative relationship between collateral and interest rates, which is the contract chosen by lower risk borrowers. Nevertheless, moral hazard is shown to affect the initial choice of the contract, weakening the obtained link.

Breit and Arano [2008] consider the determinants of the interest rate applied to small businesses, and hypothesize a lower risk premium when collateral is required. They find that the interest rate is lower when collateral secures the loan. Similarly Agarwal and Hauswald [2010], using 2002 and 2003 U.S. data for SME, find a negative and significant coefficient for collateral regressed on the offered loan rate. On the contrary Lehmann and Neuberger [2001], treating collateral as a dummy variable, analyze 1988 U.S. data for 174 lines of credit, finding a positive coefficient.

Collateral is shown to reduce credit risk also in Thailand, through an empirical analysis carried out by Menkhoff, Neuberger and Suwanaporn [2006]. Furthermore, its incidence is shown to be higher than in mature markets. Booth and Booth [2006] reach the same conclusion, analyzing data from the Securities and Exchange Commission (SEC) of loans contracts stipulated from 1987 to 1989.

Godlewski and Weill [2011] highlight the conflicting literature on the theme, explaining the dissimilar conclusions provided by the authors as a different degree of asymmetric information among countries, using a sample of 4,940 loans from 31 countries. In particular, they analyze the link between loan risk premium and collateral: a simple OLS regression is employed, using as independent variables collateral, the degree of information asymmetries for each country and some other variables. Financial, accounting standards and economic development indicators are used as proxy for the degree of information asymmetries. The results show a significant positive coefficient for the collateral variable, apparently not supporting its use as a device to solve adverse selection issues. However, a further analysis shows that this positive link is weaker when information asymmetries increase. Berger, Frame and Ioannidou [2016] conduct an empirical study on commercial loans of Bolivian financial institutions and find a negative relationship between collateral amount and risk-premium, which is explained through a lower loss for banks in case of borrowers' default. A similar conclusion, but with a different sample, is given in Blazy and Weill [2006], analyzing the role of collateral in French banks and discussing how guarantees reduce losses when default happens. They conclude that collateral could help to solve adverse selection problems.

In the next section, and in Section 5, we add to this literature by analyzing recently collected loan data which are higher dimensional than most of the works mentioned here. Moreover, we consider specific categorical variables which are very informative about borrowers' features. We provide new evidence on the ambiguous relationship between collateral and interest rate.

3 Unsupervised analysis

3.1 Data description

In this paper, we contribute to the previous literature by analyzing a unique big data set of loan contracts at microeconomic level. This is provided by the European DataWarehouse (ED), the first centralized platform in Europe which collects ABS loan data, motivated by the ABS loan level initiative directly conceived by the European central bank. The mechanism is simple: *"servicers, trustees or other designated entities"* upload this data periodically, following the ECB recommendations.[3] In fact, the data set analyzed is very peculiar compared with the previous literature on the subject. It is a huge collection of microeconomic data for millions of Italian loan contracts, collected for each single borrower associated with a residential mortgage-backed security; the dimension of the available data is $1,147,311$ prior to cleaning.

The variables are listed in Table 1. Our goal is to understand what evidence this collection of big data supplies.

Table 1: Variables available for each loan contract.

Variable type	*Variables*	*Description (only for qualitative variables)*
Quantitative	Interest rate	
	LTV	
	Debt to income	
	Original balance	
	Loan term	
	Borrower's income	
Qualitative	Borrower type	Individual or commercial.
	Employment status	Employed, unemployed, self-employed, etc.
	Resident	Resident in the country (more or less than 3 years) or not resident.
	Repayment method	Type of principal repayment.
	Payment frequency	Number of months between payments.
	Payment type	Annuity, linear, etc.
	Lien	Seniority on the liquidation of the property.
	Interest rate type	Floating, fixed.
	Property type	Residential or for commercial use.

Note. The table reports the variables employed in the analysis, with a short description for the categorical variables.

We begin by considering the numerical variables available for each borrower, and we apply to them an unsupervised analysis. Those are: the interest rate margin of the loan contract, the primary

[3] Source: European DataWarehouse website, see footnote 2.

income of the borrower, the loan term, the original balance of the loan and the original loan to value ratio defined as:

$$\text{LTV} = \frac{\text{Loan Amount}}{\text{Appraised Property Value}}.$$

An RMBS is essentially a bond backed by the interest coming from a mortgage, hence the property value is used as collateral. Therefore, the results for this variable will provide a direct empirical answer to the question posed at the beginning: it is inversely proportional to the collateral amount.

We exclude from the unsupervised analysis the debt to income ratio, due to the lack of data (see Table 4 below); in fact, the information provided by this index is embedded in the borrower's income, which is included.

3.2 Empirical analysis

In order to get a better understanding of the interplay between the quantitative variables of Table 1, which is relevant for the interest rate decision, we apply to this subset of data two unsupervised learning methods: the K-means clustering and the principal component analysis. In particular, this analysis allows to inspect the link between collateral and interest rates.

To begin with, we analyze the data without any time splitting (i.e. without separating it according to its origination date, see below). The first step consists of cleaning the data: essentially, there are three anomalous cases. One is missing values, this happens for instance when the bank does not upload it as it should. The second is similar, some values are instead recorded as zeros, which must be removed. Finally, we remove outliers: for instance, when values are incorrectly recorded. Regarding the last point, we delete the values which differ from the median more than three times the scaled median absolute deviation. Table 2 and Table 3 collect the descriptive statistics of the data after cleaning. loan term refers to month units, the other values are numerical, in particular interest rate and LTV are in percentages. Note that the average LTV is less than 1, as expected, and that on average loan contracts have a term of 20/30 years. In Table 3 it is easy to see how many different contracts deviate from the average, through an immediate descriptive quartiles representation.

Table 2: Descriptive statistics.

Variables	Mean	Median	Std. Deviation	Skewness	Kurtosis
Interest rate	1.94	1.76	1.0491	0.8242	3.298
LTV	63.28	67.94	20.576	-0.40838	2.5251
Original balance	1.1212e+05	1.05e+05	47,954	0.61774	3.0792
Loan term	274.27	288	79.981	-0.1625	2.2303
Borrower's income	27,272	24,000	12,508	0.77687	2.9537

Note. The table reports the descriptive statistics for the numerical variables.

Panel A and Panel B of Figure 1 represent, respectively, the histogram of the loan to value and the interest rate margin variables, which are the most relevant variables in our analysis; see also

Table 3: Quartiles for the numerical variables.

Variables	$\alpha = 25\%$	$\alpha = 50\%$	$\alpha = 75\%$
Interest rate	1.15	1.76	2.50
LTV	48.27	67.94	78.70
Original balance	76,000	105,000	140,000
Loan term	238	288	360
Borrower's income	18,037	24,000	35,012

Note. The table reports the quartiles for the numerical variables.

Figure 2 which represents instead the two variables jointly. Note that the loan to value is frequently lower than 1, and that interest rates are of course positive and usually close to their average. The relation between interest rate and collateral can be examined by looking at the cluster analysis of the loan to value - interest rate variables.

Panel A: Interest Rate: Histogram. Panel B: Loan to Value: Histogram.

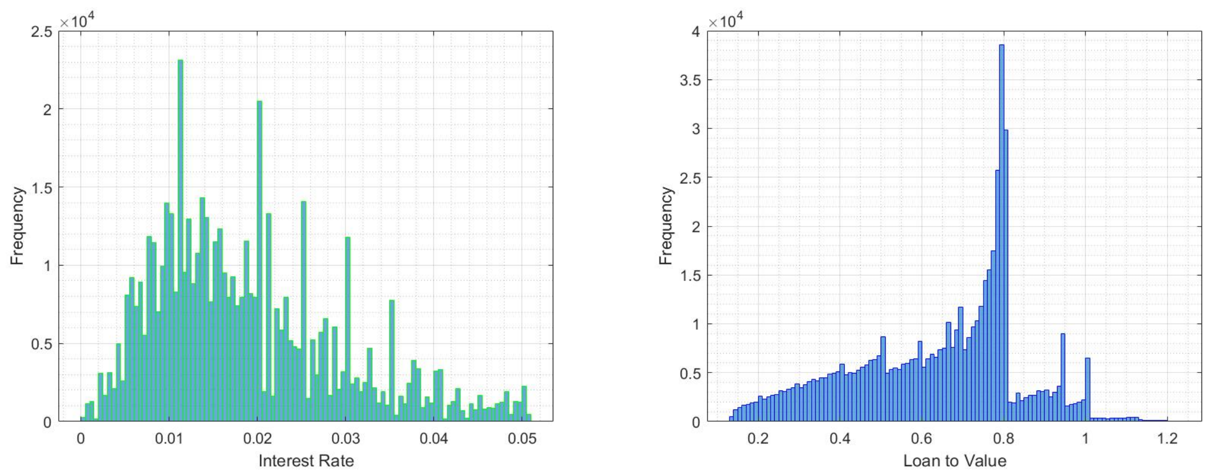


Figure 1: Panel A: The histogram of the interest rate. The strictly positive values are concentrated mostly in the interval $[0.01, 0.04]$. Panel B: The most common LTV ratio is around 0.8; some loans have a LTV ratio higher than 1, i.e. the amount of loan is higher than the appraised property value.

The K-means clustering partitions the data set into K disjoint sets of related data, where the relation is quantified by the distance from the so-called centroid, the cluster center; when the point-to-cluster distances are minimized, each observation gets assigned to its optimal cluster [see Lloyd, 1982; Arthur and Vassilvitskii, 2007]. This is an iterative algorithm, and the number of clusters is chosen a priori. We start by looking at a 2-means clustering (Figure 3, Panel A), to verify if the observations with higher interest rates are associated to the cluster with a higher LTV, using as metric the L^1 distance. When partitioning the observations in two clusters, it seems that there is a positive link between the loan to value and the interest margin: the second centroid is associated with slightly higher interest rates and higher loan to value. Nevertheless, this prevailing

Panel A: Loan to Value and Interest Rate.

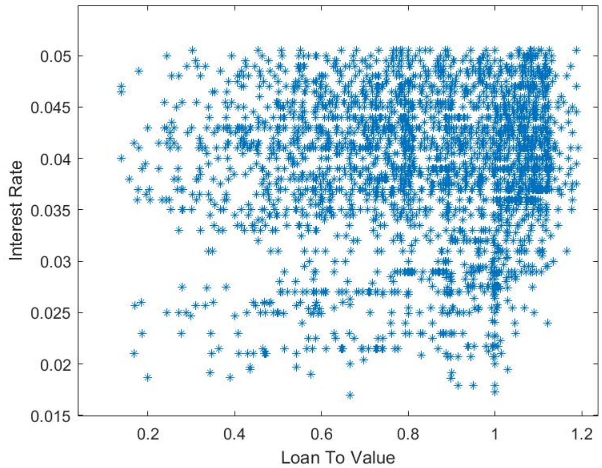


Figure 2: A simple Cartesian plot shows how there is a slightly higher concentration of points when both the Loan to Value and the interest rate are higher. However, there is not a clear trend, as we show in the cluster analysis.

effect reflects a general trend and it is not a strong evidence: we justify this theoretically with a model in Section 4. Indeed, it is clear how higher and lower interest rates are basically distributed almost uniformly on both clusters. These results justify the contrasting conclusions present in the economic literature and assess the possibility to find contrasting evidence when different data sets are analyzed.

When introducing a third cluster (see Figure 3, Panel B) the conclusions do not change, as expected. Since there is not a clear partition of the data set, increasing the clusters does not provide different evidence, and actually it confirms that the interest rate-LTV link is unclear, consistently with Section 2. Looking at the correlation matrix, represented with a heatmap in Figure 4, we notice a slightly positive correlation between the variables LTV and the interest rate. This is consistent with the cluster analysis. Moreover, the correlation of the primary income with the loan term, as well as with the loan to value, is almost zero; the other correlations are positive, except for the ones related to the interest rate.

The higher correlation of the primary income with the original balance suggests that borrowers with higher income borrow more money on average, as expected. Instead, the negative correlations in the interest rate column sets this variable aside from the others. In fact, this is the dependent variable of our supervised analysis, namely the one decided by the lender given all the others.

Now we provide more insights by applying the principal component analysis (see the Appendix for a short description of the methodology and terminology used). The ratio behind the PCA is to reduce the number of variables while retaining the maximum amount of information. Indeed, the key feature of the PCA, which is strictly linked to its construction, is that usually only a subset of vectors of the score matrix can be retained without losing too much information compared to the

Panel A: Cluster Analysis With 2 Centroids. Panel B: Cluster Analysis With 3 Centroids.

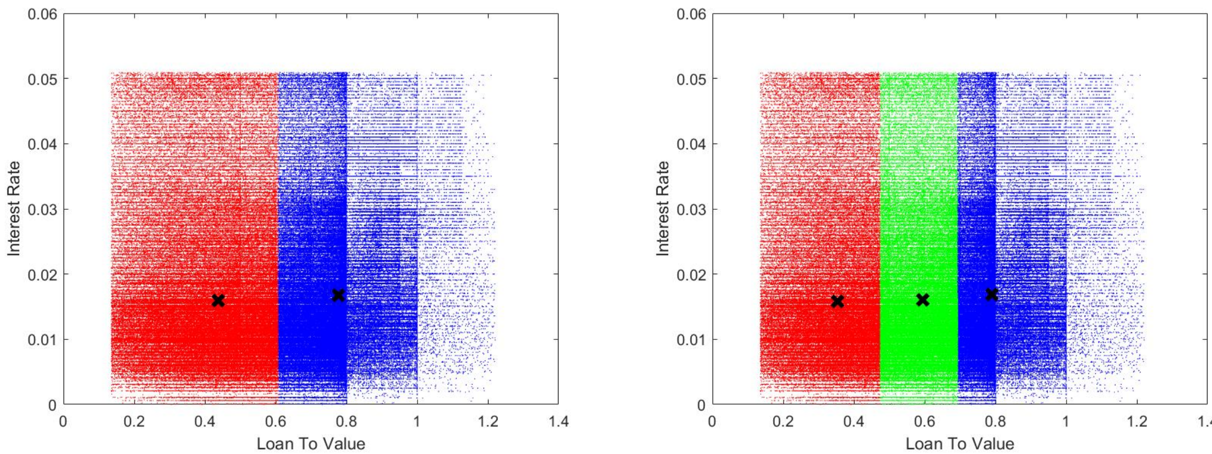


Figure 3: Panel A: Cluster analysis applied to the variables Interest Rate and LTV, with a 2-cluster splitting. Panel B: Compared to Panel A, there is an additional centroid, partitioning the data into 3 clusters.

Panel A: Correlation Matrix.

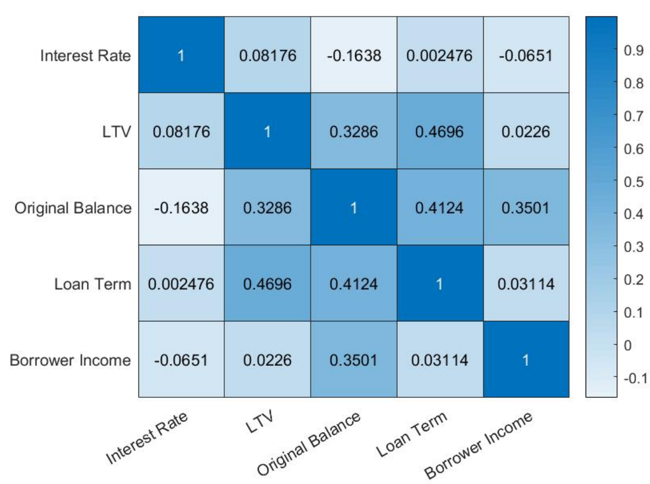


Figure 4: Correlation among the variables, represented through an intuitive heatmap.

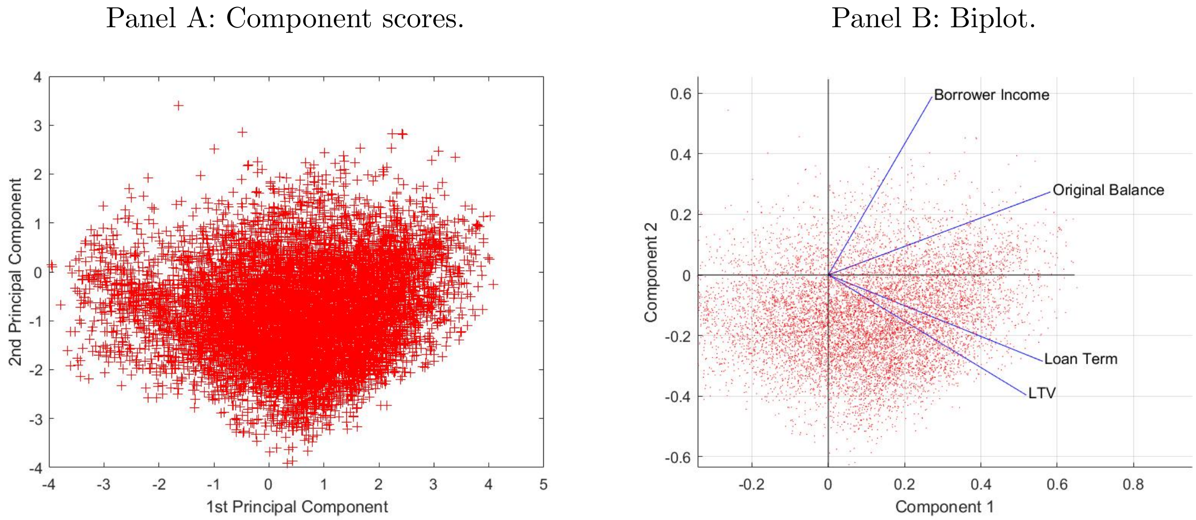


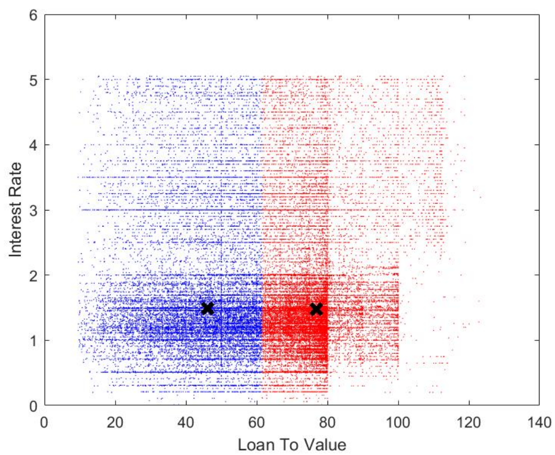
Figure 5: Panel A: The Cartesian plot of the first two principal components. The red points represent the data. Panel B: Geometrical interpretation of the first two principal components: the first two principal axes are both positive for the primary income and the original balance, while the loan term and the LTV have a negative second component.

original data. The latter is measured as data variation.

In Panel A of Figure 5 we retain the first two components, and we plot the scores associated to them. The data are evenly distributed, therefore we look at the biplot in Panel B in order to provide a graphical interpretation of the principal components in terms of the original data. There is a clear pattern. The first principal component, which explains almost 40% of the data variation, reflects the overall level of the loan variables. This is a common result of PCA: it simply shows that the greatest source of variation is the magnitude of the variables. This was expected, since the entries of the correlation matrices are all positive for the four variables considered. The second principal component is more interesting. It contrasts two groups: {primary income, original balance} and {loan term, loan to value}. This captures some of the results of the previous correlation analysis, and suggests that a higher primary income leads both to the possibility to pledge more collateral and to ask for a loan for a shorter time. The last consideration is supported by the possibility of a more rapid repayment. The other variable, the original balance, is positively related to the primary income, which means that loans of higher amounts are associated with a higher borrower's income. Finally, the contrast between the original balance and the loan to value ratio suggests that an increase in the collateral value is more than proportional than an increase of the original balance. This supports the conjecture of a (slightly) positive link between the amount granted and the collateral required.

Now we examine if the results of the previous analysis change when we take into account a possible structural break, the economic crisis. We conduct this robustness check by splitting the data in two shorter periods, based on the loan origination date: before 2006 (before the crisis, higher

Panel A: 2 Clusters, Before 2006.



Panel B: 3 Clusters, Before 2006.

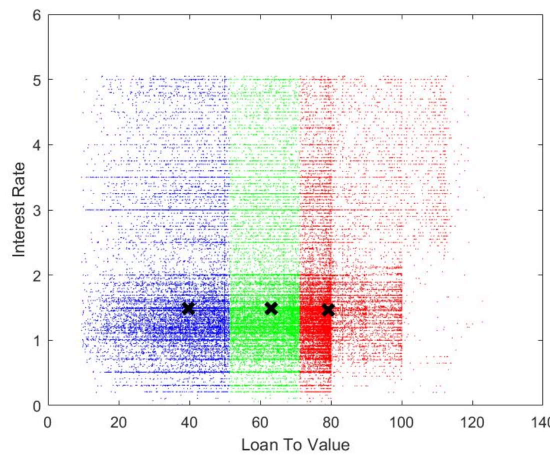


Figure 6: Panel A: Cluster analysis applied to the variables interest rate and LTV before 2006, with a 2-cluster splitting. Panel B: Compared to Panel A, there is an additional centroid, partitioning the data into 3 clusters.

margins) and after 2011 (post-crisis, lower margins). This choice can be motivated by the different level of the Euribor interest rate during the two periods.

Essentially, the analysis performed above is unchanged. Figure 6 and Figure 7 represent the cluster analysis on the data sampled, respectively, before and after the crisis. The difference with the previous results is negligible. The same holds for the correlation matrix (see Figure 8): the only changed sign is the correlation of the interest rate with the primary income; this should be interpreted as a reflection of the positive correlation between the borrower's income and the original balance which is negatively correlated with the interest rate. Finally, the interpretation of the biplots is unchanged too (compare Figure 9 with Figure 5).

4 The model

4.1 Perfect information

In this section we provide a theoretical justification for our results, namely building contrasting — but equally plausible — models in order to show the possibility of ambiguous relations between collateral and interest rates of loan contracts, as the empirical analysis points out. This also helps to explain the contradictory results presented in Section 2.

The starting point of our discussion is to consider a principal-agent game in the context of a loan contract. Indeed, a bank can be seen as a principal which gives an agent (the borrower) a sum L and expects a net payoff which depends on the interest rate i. Therefore, the bank decides the contractual conditions, and the agent receives the sum borrowed and is required to put an effort $e \in E$ to invest the borrowed capital; this effort e is costly for the borrower. Assume that $e \geq 0$

Panel A: 2 Clusters, After 2011. Panel B: 3 Clusters, After 2011.

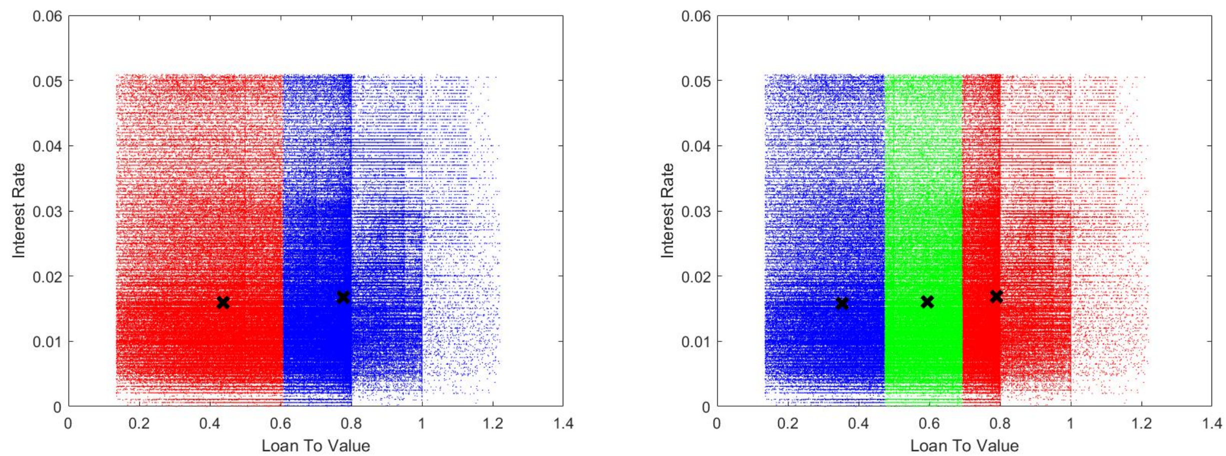


Figure 7: Panel A: Cluster Analysis for the Italian data, after 2011, with two centroids. Panel B: Cluster Analysis, same as in Panel A, but with three centroids.

Panel A: Heatmap, before 2006. Panel B: Heatmap, after 2011.

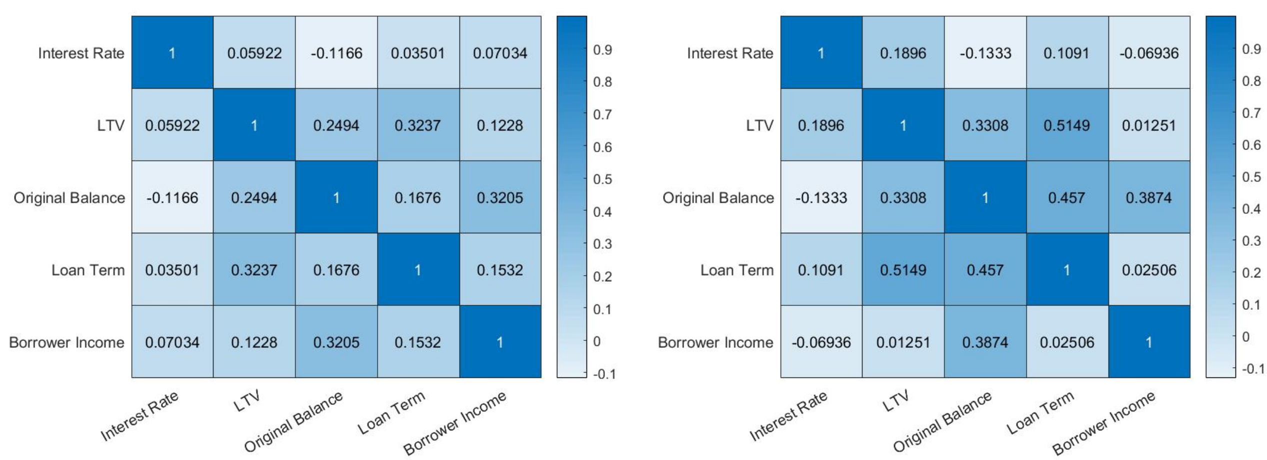


Figure 8: Panel A: Correlation among the variables before 2006. Panel B: Correlation among the variables after 2011.

Panel A: Biplot, before 2006.

Panel B: Biplot, after 2011.

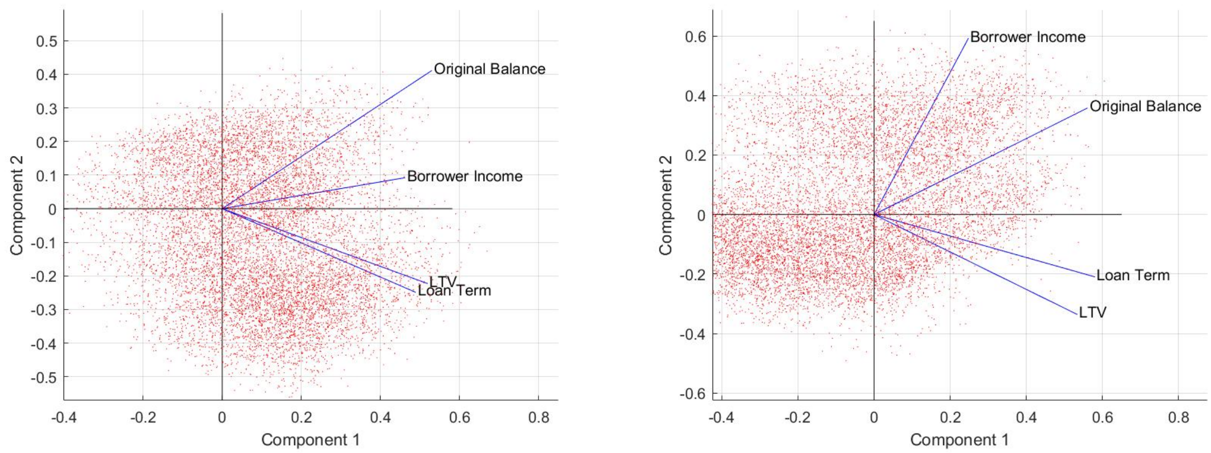


Figure 9: Panel A: Biplot of the first two principal components before 2006. Panel B: Biplot after 2011; overall, the pattern does not change over the two different periods.

belongs to a compact set E of possible efforts.

We define the critical effort $\tilde{e}$ as the minimum effort e required for the borrower to avoid default, and consider two possible states of the world $s_k \in S$, $k = 1, 2$, related to a loan contract (we rule out partial default):

$$S = \begin{cases} s_1 = \text{"Default"} \\ s_2 = \text{"Solvency"}. \end{cases}$$

The bank has two possible related final payoffs: one for the state of the world s_1 and another for the complementary case. The realization of S depends on $\tilde{e}$, but since the critical effort is defined *ceteris paribus*, there are other factors which can influence the realization of a specific state of the world. These quantities are stochastic, thus they influence the conditional probability $\mathbb{P}(S = s_k|\tilde{e})$. Under the previous notation, the optimization problem for the principal becomes

$$\max \sum_{k=1}^{2} \mathbb{P}(S = s_k|\tilde{e}) \cdot U_1(\pi_{c_k}), \tag{1}$$

where π_k is the final payoff associated with the specific state of the world k, which depends also on the effort required to the borrower. The probability $\mathbb{P}(S = s_k|\tilde{e})$ must be estimated by the principal in order to compute his maximum expected payoff and U_1 is an appropriate utility function associated with the lender. Since the effort is costly for the borrower, we assume that the borrower's cost function $c : e \longmapsto c(e)$ is increasing. Hence, taking into account this function, the previous maximization problem is subject to the restriction:[4]

$$\sum_{k=1}^{2} \mathbb{P}(S = s_k|\tilde{e}) \cdot U_2(W) - c(e) \geq A, \tag{2}$$

where U_2 is the utility function associated with the borrower. The term A is the so-called reservation utility, and intuitively represents the agent's expected utility of alternative investment opportunities, hence it has to be lower than the left hand side of (2). This restriction is assumed to be both necessary and sufficient to ensure that the contract is signed by the borrower. The argument W of the function U_2 includes both the sum borrowed L and the payoff of any external investment made with this sum, or any quantifiable additional utility obtained by the borrower using the sum L, clearly depending on s_k.

Under this framework we prove that the presence of collateral implies a lower interest rate.

The effect of either the absence or the presence of collateral, say of value C, on the bank's final (gross) payoff π — or equivalently, except for a sign factor, on the borrower's payment — can be represented through piecewise-defined functions. Equivalently, for loans which are both secured but differ for their collateral value, for instance when $C_1 > C_0$, the representation is:

$$\pi_c = \begin{cases} C_0 & \text{if } S = s_1 \\ L(1+i) & \text{if } S = s_2, \end{cases}$$

$$\pi_{c'} = \begin{cases} C_1 & \text{if } S = s_1 \\ L(1+i') & \text{if } S = s_2. \end{cases}$$

The determination of the link between the different rates i, i' is the aim of our discussion. We base our analysis on two established principles of finance: the risk-return trade-off and the risk-aversion of economic agents.

Considering the risk-return trade-off, from a theoretical point of view, an injective relation between risk and return[5] can be assumed, but in practice some problems occur. In fact, risk is not observable: but it can be estimated, for example using risk classes. Moreover, the relation risk-return is a function only when considering the average return for each risk class. Hence, when we consider an increasing function which maps risk into return, we interpret it as derived using an estimation of risk and for an average of many loan contracts: then it is acceptable to assume that when the risk increases, the return increases too.

Moving from the previous considerations, we notice that the presence of collateral (or of collateral with a higher value) decreases the expected loss of the bank if default occurs, or equivalently it increases its expected return. Thus, given a fixed expected maximum return without collateral, the same value can be obtained applying lower interest rates if collateral is provided by borrowers. Similarly, considering a contract whose collateral value is C_1, the interest rate should be lower in

[4]This is satisfied a priori, otherwise the game does not exist.

[5]Various definitions for the meaning of these two terms exist: let us consider, respectively, the variance and the expected value of payments.

order to maintain the principal's expected maximum return constant, that is if $C_1 > C_0$, then $i' < i \Leftrightarrow L(1+i') < L(1+i)$, being $L > 0$. Moreover, the borrower's risk is not changed by the presence of collateral under perfect information, because in this context the risk is independent from collateral. In fact if risk with or without collateral is the same, because borrowers are the same and there are not adverse selection problems, collateral and interest rates can be considered as substitutes in order to price the risk suffered by the bank. Therefore, when risk is fixed, if the presence of collateral increases the expected return, a contradiction of the risk-return principle arises. Thus, in order to obtain the same return, the interest rate applied must be lowered.

From this discussion it follows immediately that if $C \in (0, \infty)$ is the value of collateral and i the interest rate of a loan contract, when collateral is required to each borrower, an inverse relationship between i and C holds. Indeed, under perfect information, collateral is not used to solve moral hazard or adverse selection problems. Therefore, as pointed out above, the risk from each borrower does not change if he provides C. Since g is bijective, the expected total return r must be unique for a given risk. Since r is increasing in C, then the interest rate i must decrease when C increases.

It is important to underline that all these conclusions are derived *ceteris paribus*, that is fixing all the other variables that could affect the interest rate decision. In other words, banks prefer higher collateral requirements and higher margins, but in an efficient and ideal market, with no free-lunches and perfect competition, this is not feasible. A bank is rewarded only for the risk it assumes, because risk is priced, contrary to its subjective decision about the collateral/interest proportion (similarly to the choice of the debt-to-equity ratio in private companies).

On the other hand, considering the situation from the agent's point of view, when the principal requires additional collateral with value C, this term enters in utility function of equation (2) with a negative sign. But this can change the inequality: when (2) becomes

$$\sum_{k=1}^{2} \mathbb{P}(S = s_k|\tilde{e}) \cdot U_2(W - C) - c(e) < A, \tag{3}$$

the borrower does not sign the contract. Since the risk of the borrower is supposed to be constant, this condition is not acceptable because the bank would give up on the expected payoff that it would have obtained without any collateral. In fact, collateral should increase the expected payoff of banks. At the same time, under perfect competition, the agent could sign the contract with other banks so that (2) is satisfied. These banks exist as far as their remuneration is in line with the market risk premium associated with the specific risk of the borrower.[6] The only way that a bank has to avoid this situation, and then satisfy (2), is by means of applying a lower interest rate. Note that interest rates are included in equation (2), because we consider the net borrowed capital, which includes the interest to be paid, and it is higher when interests are lower. Again, the bank cannot simultaneously achieve a higher i and a higher C for a given borrower.

[6] This conclusion follows from the well-known implications of perfect competition.

4.2 Asymmetric information

When e is not a fixed quantity, i.e. the agent actively chooses the effort, moral hazard comes into play. In this case the agent maximizes his utility function by taking e into consideration:

$$\arg\max_{e \in E} \sum_{k=1}^{2} \mathbb{P}(S = s_k|\tilde{e}) \cdot U_2(W) - c(e). \tag{4}$$

If I is the value of any investment made by the borrower with the sum L, which is the loan balance, when $e < \tilde{e}$, the argument of the utility function U_2 contains $L + I_d$, where I_d is the value of I in the case of default (recall the definition of $\tilde{e}$). The key point is the following: in this scenario a collateral requirement can be used by the bank (the principal) to induce the borrower to increase e.

We consider the case when the effort can be either high or low: $e \in \{e^H, e^L\}$; then the presence of collateral changes the maximization problem into:

$$\arg\max_{e \in E} \sum_{k=1}^{2} \mathbb{P}(S = s_k|\tilde{e}) \cdot U_2(W') - c(e), \tag{5}$$

where W' includes the quantity $-C$:

$$W' = \begin{cases} L - C + k & \text{if } S = s_1 \\ L + I + k' & \text{if } S = s_2. \end{cases}$$

There, k, k' denote respectively the other profits and costs. Now, $c(e^H) > c(e^L)$, given the monotonicity of $c(e)$. The principal prefers higher effort, e^H, and computes the probabilities associated with each result, for a given effort: $\mathbb{P}(S = s_k|e^H)$ and $\mathbb{P}(S = s_k|e^L)$. Clearly, a lower effort is more likely to produce default (D):

$$\mathbb{P}(D|e^L) > \mathbb{P}(D|e^H). \tag{6}$$

This explains the principal's preference for the higher effort. If the agent maximizes his expected net utility by choosing a lower unverifiable effort, the principal should design the contract in order to shift the agent's choice from e^L to e^H. However, the value I must be taken into consideration too. Therefore the inequality:

$$\sum_{k=1}^{2} \mathbb{P}(S = s_k|e^H) \cdot U_2(W') - c(e^H) > \sum_{k=1}^{2} \mathbb{P}(S = s_k|e^L) \cdot U_2(W') - c(e^L) \tag{7}$$

is not always verified and depends on the additional payoff given by I — which is, in probability, higher when $e = e^H$ — compared with the higher cost associated with e^H. In all cases where (7) holds, the equilibrium will be at e^H. More specifically, in every situation where the higher cost associated with default is greater than the higher cost of choosing e^H, an agent is more likely to choose e^H.

This explains why, under moral hazard, collateral can be associated with a higher interest. As a matter of fact, higher interest rates and higher collateral requirements are compatible, though not always necessary, if they are both considered as a penalization for a riskier borrower who prefers e^L

to e^H.

The effect of moral hazard on collateral has been already analyzed, but with a different derivation, in Booth, Thakor and Udell [1991] and Flatnes and Carter [2019]. The main difference of our model is its generality: we rely on few intuitive principles of game theory. Another difference is that, by considering the totality of contracts rather than a single loan, our discussion develops in terms of frequency. Nevertheless, our results are consistent with these papers.

In addition to moral hazard, the effect of adverse selection may also change the conclusions. Consider the situation where each borrower has more information than the principal regarding his own quality q, which means he is either a *good* (G) or a *bad* (B) borrower. A good borrower is an agent who is less likely to default. On the contrary, a bad borrower is an agent associated with a higher probability of default. We assume that the quality q is not influenced by the effort e. In this case different agents should get different contractual conditions, and in particular better agents should obtain better conditions in terms of the interest rate. But this is not easily achievable when information is not perfect, because the bank does not know who are the good borrowers. On the other hand, bad borrowers try to take advantage of better contractual conditions, obfuscating their true category. At the same time, an adverse-selection effect takes place if banks raise the interest rate, because this leads to exclude lower risk borrowers and to retain those who are riskier.

In this case banks could design the contract in order to discern and select between good and bad borrowers, by requiring a signal: if an agent proves his good quality to the principal, then he can take advantage of its features, otherwise he achieves a lower utility. The main point here is that borrowers are not interested in revealing their true quality if they can obtain better conditions by cheating on it, and they are prone to reveal this information if they can take advantage of it. In this case, if an agent regards himself as a good borrower ($q = G$), in order to compute his expected payoff, he uses a strictly lower weight $\mathbb{P}(D|q = G)$ for the sum associated with the event *"Default"* compared to a bad borrower: this results in $\mathbb{P}(D|q = B) > \mathbb{P}(D|q = G)$. On the other hand, from the principal's point of view, collateral is used in order to distinguish better borrowers and to apply a lower interest rate: *ceteris paribus*, when collateral is provided, interest should be lower. Thus, collateral can be seen as a tool that mitigates the adverse selection problem.

When a lower interest rate does not compensate for the borrower's higher expected cost of providing collateral, assuming a common cost function $c(e)$ for all borrowers:

$$\sum_{k=1}^{2} \mathbb{P}(S = s_k|q = B) \cdot U_2(W') - c(e) < \sum_{k=1}^{2} \mathbb{P}(S = s_k|q = G) \cdot U_2(W') - c(e). \tag{8}$$

As in the perfect information case, the presence of collateral is associated with a lower interest rate, but here collateral allows the bank to discriminate between good and bad borrowers. This result is consistent with Bester [1985], Bester [1987], and Chan and Thakor [1987], but derived more straightforwardly, since our model is based on fewer assumptions. Moreover, our derivation can be adapted to include more specific hypotheses. For example, taking into consideration the difference in the collateral value for borrowers and for banks, Besanko and Thakor [1987] suggest the same conclusion. Indeed, their result can be immediately justified under our framework, because even when the bank evaluates collateral less than the borrower, say $\beta \cdot C$ for $\beta \in (0,1)$, in our model collateral has a direct effect only on the agent's decision. Therefore, since the value remains C for the borrower and the β coefficient affects only the bank decision, the incentive or the signal effects remain in place in our model.

Finally, when moral hazard and adverse selection are both present, the effect is not clear if we consider the totality of contracts. This conclusion is consistent with Booth, Thakor and Udell [1991], who cannot derive a straightforward relation when they are jointly present. As a matter of fact, we have discussed how collateral can be used as a tool in both cases, but it is associated with, respectively, higher and lower interest rate. Both e and the borrowers' quality q influence the probability of default (D). In fact, we can write the probability of default as $\mathbb{P}(S = D) = f(e, q, \cdot)$, where $\cdot$ denotes all the other variables possibly affecting this probability. In this case,

$$\mathbb{P}(S = D | q = B, e = e^L) > \mathbb{P}(S = D | q = B), \tag{9}$$

because we have assumed that a lower effort increases the default probability. These quantities are unknown to the lender, since q and e are unobservable under asymmetric information. Therefore, the decision on the interest rate cannot be easily associated with collateral requirements. In fact, the bank could achieve both a signal and an incentive effect by raising at a proper level collateral, but this would generate a negative outcome at the same time, that is a loss of surplus. This loss is due to the contracts associated with those borrowers who consider the cost of additional collateral too high, and therefore do not enter the game. For example, these borrowers could belong to the group (e^H, B). Indeed, without the second assumption on their quality, some of them would have signed the contract under moral hazard, evaluating it profitable as discussed above. This loss of lender's profit is intuitive, since the bank should lose some payoff in a situation where it is penalized by lack of information.

5 Supervised analysis

We now perform a supervised analysis, which includes the categorical variables excluded in the analysis of Section 3, to model the interest rates on the considered data using a classical linear regression model. In this respect, we specify the following model:

$$\textit{interest rate}_{i,t} = c + \boldsymbol{\beta} \cdot \textit{Borrower Controls}_i + \boldsymbol{\gamma} \cdot \textit{Loan Controls}_i + \varepsilon_{i,t}, \tag{10}$$

where the interest rate represents the rate charged to the loan i at the origination date t. We define the *Borrower Controls*$_i$ category that includes the logarithm of: the borrower's income, resident, borrower type, debt to income, and the employment status. The *Loan Controls*$_i$ encompasses the logarithm of the original balance, loan term, lien, repayment method, payment type, payment

Table 4: Data availability.

Variable	**% Available**
Interest rate (dependent)	100.00
Loan to value	83.00
Debt to income	24.05
Original balance	71.11
Loan term	78.78
B Borrower's income	79.81
Borrower type	71.50
Employment status	95.44
Resident	22.17
Repayment method	97.91
Payment frequency	96.01
Payment type	95.97
Lien	7.68
Interest rate type	95.51
Property type	97.40
Origination year	100
EURIBOR 3M	99.99
Total loans	**724,541**

Note. The table reports the percentage of available data among the considered variables over the total loans.

frequency, interest rate type, property type, and the origination year. Additionally, we include the 3 months EURIBOR which is considered a reference for the offered interest rate in the European Interbank market. As it can be viewed in Table 4, the data availability is heterogeneous across the considered variables. For instance, the availability for the debt to income ratio is available less than one loan out of four while for resident and lien is less than 23% and 8%, respectively.

Therefore, the number of observations in the estimation is completely driven by the variable in the selection with the higher value of missing data. To cope with this heterogeneity, we consider three specifications of the models. The first considers all the variables except the debt to income, resident, and lien. This choice is equivalent to impose a threshold above 70% in the variables selection criterion. The second specification involves all the variables included in the previous one plus the debt to income and resident (threshold above 20%). The last specification considers all the variables including lien.

Results are reported in Table 5. Following the data availability, the fixed effects (FE) included in all three specifications involve the borrower type, the employment status, the repayment method, the payment frequency and type, the interest rate type, the property type, and the origination year. The resident fixed effect is included in the second and third specifications while lien is included only in the last specification. The first specification involves 433,786 observations.

As we have seen in the previous sections, the presence of information asymmetries makes the determination of the LTV sign uncertain a priori: in this particular specification a higher collateral requirement is associated with a lower interest rate, as in the cases of perfect information or adverse selection. Indeed, loan to value is significant and positively related to the interest rate: in this case high ratios are considered riskier loans. The original balance is significant and negative related to the interest rate, indicating that a higher amount of loan corresponds to a lower interest rate. Analogously, the loan term is negatively related to the loan terms and this is quite a surprising result since usually the duration of the loan increases the lending interest rate. The same applies to the borrower's income which is positively related to the interest rate. As expected, the Euribor 3 months interest rate is significant and positively related to the loan's interest rate in all three specifications since it represents the basis for the pricing of several European debt financial instruments including loans. With the inclusion of the debt to income ratio, the number of observations decreases by more than 87%. In this case, we have a change in a sign for the LTV and the loan term. The debt to income ratio results negatively related to the interest rate. In the last specification, the reduction of the sample is more than 98%. In this case, all the variables are included. The debt to income ratio becomes not significant. Different from the other specifications and as expected, the borrower's income results are negatively associated with the interest rate while the loan terms are significant and positively related to the interest rate. The loan to value remains significant and negative as in the second specification. The moral hazard model of Section 4 explains the latter sign, while overall those results can be understood as follows. First, missing data on fundamental credit risk drivers such as debt to income ratio introduce a strong restriction from the original sample. Consequently, the resulting limited sample could reflect systematic errors, rather than random ones, on the data, and this could bias the estimated coefficients. A solution would be to extend the analysis to a multi-country level (e.g. Europe). Second, if those results are confirmed, they clearly indicate mispricing in the interest rate.

6 Conclusion

In this work, we have studied the open problem regarding the effect of collateral on interest rates of loan contracts, which is strictly connected to the theory of asymmetric information. We have dealt with this topic both empirically and theoretically. The key message of our discussion is: the link collateral-interest rate in loan contracts is ambiguous, and it cannot be decided a priori.

This is clear when examining past literature on the topic: there are many contradictory findings. In this paper, we have broadly summarized these previous findings, both for theoretical and empirical literature. In our analysis of Italian loan data, we have provided some new econometric evidence using a unique data set of numerical and categorical variables, for thousands of borrowers, stored in the European DataWarehouse database. We applied cluster analysis, PCA and a regression model to it. Our results show that the link interest rate - collateral is in fact unclear. As a byproduct, we discussed different insights on other variables which influence the determination of interest rates.

In our model, we have studied what is the consequence of higher collateral on interest rates, based

Table 5: Results of the regression analysis.

Interest rate	(1)	(2)	(3)
Loan to value ratio	0.00003***	-0.00002***	-0.00003***
	[0.00000]	[0.00000]	[0.00001]
Original balance	-0.00381***	-0.00300***	-0.00009
	[0.00004]	[0.00012]	[0.00021]
Loan term	0.00000***	-0.00000**	0.00001***
	[0.00000]	[0.00000]	[0.00000]
Borrower's income	0.00030***	0.00090***	-0.00066***
	[0.00003]	[0.00010]	[0.00015]
Debt to income		-0.00000***	0.00006
		[0.00000]	[0.00004]
EURIBOR 3M	0.00086***	0.00264***	0.00143***
	[0.00007]	[0.00017]	[0.00027]
Borrower type FE	Yes	Yes	Yes
Employment status FE	Yes	Yes	Yes
Resident	No	Yes	Yes
Repayment method FE	Yes	Yes	Yes
Payment frequency FE	Yes	Yes	Yes
Payment type FE	Yes	Yes	Yes
Interest rate type FE	Yes	Yes	Yes
Lien	No	No	Yes
Property type FE	Yes	Yes	Yes
Origination year FE	Yes	Yes	Yes
SE	Rob.	Rob.	Rob.
Observations	433,786	55,883	7,949
R-squared	0.51021	0.54609	0.81818

Note. Estimate results according to the model described in Equation 10. Specification (1) includes all the listed variables in Table 4 except for the debt to income, resident, and lien. Specification (2) includes all the variables except for resident and lien. Specification (3) includes all the variables.

on asymmetric information effects. We proved that higher interest rates can be associated with more collateral under moral hazard, because collateral requirements can be used as an incentive to increase the effort. Lower interest rates are compatible with more collateral under perfect information and adverse selection.

As far as we know this is the first study which examines this data through the methods of Section 3 and Section 5, and it is one of the few works in the literature which analyzes such a large amount of specific microeconomic data. The game theory discussion embeds past and more detailed models; we obtained the same conclusions based on few intuitive principles. We leave to future research the extension of this research to a multi-country level.

A Appendix

We detail the principal component analysis of Section 3 and the terminology used therein. A comprehensive treatment of the PCA can be found in Jolliffe [2002]. Our matrix of data $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{n \times v}$, $n, v \in \mathbb{N}$, is the matrix whose j-th column, $j = 1, 2, \ldots, v$, is the column vector of observations for the j-th variable and whose i-th row, $i = 1, 2, \ldots, n$, refers to the row vector of variables for the i-th observation, denoted by $\mathbf{x}_i$. We consider $\mathbf{X}$ = [primary income, original balance, loan term, LTV]. Then $\boldsymbol{\Sigma}$ is the covariance square matrix for the random variables $\mathbf{x}$, and the PCA associated optimization problem is:

$$\underset{\boldsymbol{\alpha}_1}{\operatorname{argmax}}\, \boldsymbol{\alpha}_1' \boldsymbol{\Sigma} \boldsymbol{\alpha}_1 \quad \text{s.t.} \quad \boldsymbol{\alpha}_1' \boldsymbol{\alpha}_1 = 1,$$

where $\boldsymbol{\alpha}_1$ is a vector of v constants. The choice of this constraint is common, but arbitrary. This procedure is iterative, in the sense that it consists of successive maximizations of the quantity $\boldsymbol{\alpha}_j' \boldsymbol{\Sigma} \boldsymbol{\alpha}_j$ by imposing the orthogonality condition $\boldsymbol{\alpha}_k^T \boldsymbol{\Sigma} \boldsymbol{\alpha}_j = 0$ with $k < j$, $j = 2, \ldots, v$. The principal components (PCs) are the linear combinations $\boldsymbol{\alpha}_g^{*T} \boldsymbol{X}_{n,\cdot} = \sum_{j=1}^{d} \alpha_{g,j} x_j$ for $g = 1, 2, \ldots, d$ and each n, where $\boldsymbol{\alpha}_j'$ are called loadings. The quantity $\tilde{z} = \boldsymbol{\alpha}_j' \boldsymbol{x}_i$ is called the score for the i-th observation on the j-th PC.

Given the relevant difference in the unit of measurement of our data (see the difference in the standard deviation, Table 3, and Figure 10) we employ the weighted PCA. The weighted PCA follows the PCA model just described, but it assigns a different weight to the data [see Jolliffe, 2002, Section 14.2.1]. In other terms, the analysis is performed on standardized variables in order to correct for data size. The weights are inversely proportional to the variance of the variables. The loadings associated with the weighted PCA are not orthonormal, therefore after performing the weighted PCA we multiply the coefficient matrix for an appropriate transformation of the weights, to recover orthogonal loadings.

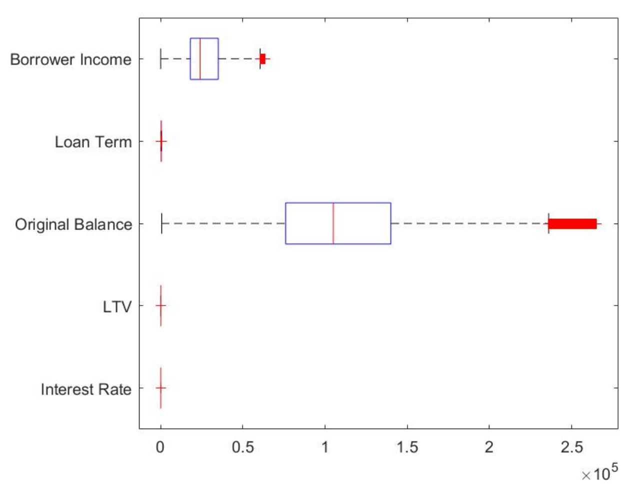


Figure 10: Boxplot of the numerical variables available for the analysis. There is a clear difference in the unit of measurement.

References

Agarwal, S. and Hauswald, R. [2010], 'Distance and private information in lending', *Review of Financial Studies* **23** (7), pp. 2757–2788.

Angbazo, L. A., Mei, J. and Saunders, A. [1998], 'Credit spreads in the market for highly leveraged transaction loans', *Journal of Banking and Finance* **22** (10), pp. 1249–1282.

Arthur, D. and Vassilvitskii, S. [2007], 'K-means++: The advantages of careful seeding', *in* 'Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms', Society for Industrial and Applied Mathematics, pp. 1027–1035.

Barro, R. J. [1976], 'The loan market, collateral, and rates of interest', *Journal of Money, Credit and Banking* **8** (4), pp. 439–456.

Benjamin, D. K. [1978], 'The use of collateral to enforce debt contracts', *Economic Inquiry* **16** (3), pp. 333–359.

Berger, A. N., Frame, W. S. and Ioannidou, V. [2016], 'Reexamining the empirical relation between loan risk and collateral: The roles of collateral liquidity and types', *Journal of Financial Intermediation* **26**(c), 28–46.

Berger, A. N. and Udell, G. F. [1990], 'Collateral, loan quality and bank risk', *Journal of Monetary Economics* **25** (1), pp. 21–42.

Besanko, D. and Thakor, A. V. [1987], 'Collateral and rationing: Sorting equilibria in monopolistic and competitive credit markets', *International Economic Review* **28** (3), pp. 671–689.

Bester, H. [1985], 'Screening vs. rationing in credit markets with imperfect information', *The American Economic Review* **75** (4), pp. 850–855.

Bester, H. [1987], 'The role of collateral in credit markets with imperfect information', *European Economic Review* **31** (4), pp. 887–899.

Bieta, V., Broll, U. and Siebe, W. [2008], 'The banking firm: The role of signaling with collaterals'. Dresden Discussion Paper in Economics.

Blazy, R. and Weill, L. [2006], 'Le rôle des garanties dans les prêts des banques françaises', *Revue d'Économie Politique* **116** (4), pp. 501–522.

Booth, A., Thakor, A. V. and Udell, G. F. [1991], 'Secured lending and default risk: Equilibrium analysis, policy implications and empirical results', *The Economic Journal* **101** (406), pp. 458–472.

Booth, J. R. and Booth, L. C. [2006], 'Loan collateral decisions and corporate borrowing costs', *Journal of Money, Credit and Banking* **38** (1), pp. 67–90.

Breit, E. and Arano, K. [2008], 'Determinants of loan interest rates: Evidence from the survey of small business finances (SSBF)', *Available at SSRN: 1140739* .

Broll, U. and Gilroy, M. B. [1986], 'Collateral in banking policy and adverse selection', *The Manchester School* **54** (4), pp. 357–366.

Capra, C. M., Fernandez, M., Ramirez-Comeig, I. et al. [2005], 'Moral hazard and collateral as screening device: Empirical and experimental evidence'. Working Paper, Emory University.

Chan, Y.-S. and Kanatas, G. [1985], 'Asymmetric valuations and the role of collateral in loan agreements', *Journal of Money, Credit and Banking* **17** (1), pp. 84–95.

Chan, Y.-S. and Thakor, A. V. [1987], 'Collateral and competitive equilibria with moral hazard and private information', *The Journal of Finance* **42** (2), pp. 345–363.

Coco, G. [1999], 'Collateral, heterogeneity in risk attitude and the credit market equilibrium', *European Economic Review* **43** (3), pp. 559–574.

De Meza, D. and Southey, C. [1996], 'The borrower's curse: Optimism, finance and entrepreneurship', *The Economic Journal* **106** (435), pp. 375–386.

De Meza, D. and Webb, D. C. [1987], 'Too much investment: A problem of asymmetric information', *The Quarterly Journal of Economics* **102** (2), pp. 281–292.

Degryse, H. and Van Cayseele, P. [2000], 'Relationship lending within a bank-based system: Evidence from European small business data', *Journal of Financial Intermediation* **9** (1), pp. 90–109.

Flatnes, J. E. and Carter, M. R. [2019], 'A little skin in the game: Reducing moral hazard in joint liability lending through a mandatory collateral requirement', *Journal of Economic Behavior and Organization* **164**, 199–214.

Godlewski, C. J. and Weill, L. [2011], 'Does collateral help mitigate adverse selection? A cross-country analysis', *Journal of Financial Services Research* **40**(1), 49–78.

Jiménez, G. and Saurina, J. [2004], 'Collateral, type of lender and relationship banking as determinants of credit risk', *Journal of Banking and Finance* **28** (9), pp. 2191–2212.

Jolliffe, I. [2002], *Principal component analysis*, Springer, 2nd Edition.

Leeth, J. D. and Scott, J. A. [1989], 'The incidence of secured debt: Evidence from the small business community', *Journal of Financial and Quantitative Analysis* **24** (3), pp. 379–394.

Lehmann, E. and Neuberger, D. [2001], 'Do lending relationships matter? Evidence from bank survey data in Germany', *Journal of Economic Behavior and Organization* **45** (4), pp. 339–359.

Lloyd, S. [1982], 'Least squares quantization in PCM', *IEEE transactions on information theory* **28**(2), 129–137.

Menkhoff, L., Neuberger, D. and Suwanaporn, C. [2006], 'Collateral-based lending in emerging markets: Evidence from Thailand', *Journal of Banking and Finance* **30** (1), pp. 1–21.

Rajan, R. and Winton, A. [1995], 'Covenants and collateral as incentives to monitor', *The Journal of Finance* **50** (4), pp. 1113–1146.

Stiglitz, J. E. and Weiss, A. [1981], 'Credit rationing in markets with imperfect information', *The American Economic Review* **71** (3), pp. 393–410.

Su, X. and Zhang, L. [2017], 'A reexamination of credit rationing in the Stiglitz and Weiss model', *Journal of Money, Credit and Banking* **49**(5), 1059–1072.

Wang, Y.-L. [2010], 'Does collateral cause inefficient resource allocation?', *Journal of Economics and Business* **62** (3), pp. 220–233.