A STRUCTURAL ECONOMETRIC MODEL OF PRICE DISCRIMINATION IN THE FRENCH MORTGAGE LENDING INDUSTRY*

Robert J. Gary-Bobo** and Sophie Larribeau***
University of Cergy-Pontoise, THEMA
33, boulevard du Port, 95011 Cergy-Pontoise,Cedex, France
Email: robert.gary-bobo@eco.u-cergy.fr

January 2002; revised October 2002; revised July 2003, final revision

Abstract.
We propose a model of discrimination in the market for mortgages. The model explains *accepted* loan applications and simultaneously determines loan sizes and interest rates. A competitive and a discriminating monopoly version of the model are proposed. Offered interest rates and loan sizes are a function of observable borrower characteristics. The competitive model rests on a marginal condition, reflecting contract optimality, to which a zero-profit condition is added. In contrast, the discriminating monopoly maximizes profits under a borrower participation constraint, reflecting the availability of a rental market as an outside option. Each version of the model is a bivariate, nonlinear model, and is estimated by standard maximum likelihood methods. The data used for estimation is a sample of clients of a French network of mortgage lenders. We show the presence of "social discrimination" in the data, the loan conditions depending, not only on the borrower's wage and down payment, but also on the borrower's occupational status. Abnormally high risk premia in the competitive version of the model suggest the presence of market power, justifying an attempt at estimating its monopolistic version. The discriminating monopoly model estimates show that the borrowers' price-elasticity of demand for housing varies with occupational status, and is inversely related with the lender's interest rate markups. This confirms that the lender exploits structural differences in the preferences to discriminate, as predicted by standard theories.
Keywords: Mortgage Loans, Price Discrimination, Discriminating Monopoly.

* We thank Richard Arnott, Marc Ivaldi, and Daniel McFadden for their advice during a preliminary stage of the present research, the participants and organizers of the 12th CREST-NBER Franco-American Seminar on Economics and 1rst CEPR Workshop on Applied I.O., held in Toulouse, October 1999, the participants and organizers of the CEPR Conference "Discrimination and Unequal Outcomes", held in Le Mans, France, January 2002, Florian Heider, Carmen Matutes, Peter Norman, Nathalie Picard, Jean-Charles Rochet and Nancy Wallace, for comments and helpful discussions. The first author thanks John Quigley for his help and hospitality in June 2000. We finally thank Steve Martin and two anonymous referees of IJIO for helpful remarks. We are responsible for remaining errors.

** University of Cergy-Pontoise, THEMA, France, and CEPR. *** University of Cergy-Pontoise, THEMA, France.

1. Introduction

To the best of our knowledge, only a handful of contributions have attempted to estimate a structural econometric model of price discrimination, and none have studied the related phenomenon of interest rate discrimination, which can be viewed as a particular case, with such methods[1]. In the present paper, we propose a structural econometric model of a market for mortgages, and estimate it on micro-data.

The model determines interest rates and loan sizes simultaneously as a function of observable borrower characteristics. It can thus be viewed as describing a case of first-degree price discrimination in which the lender imposes differing price-quantity pairs on different types of borrowers. In addition, our data set shows that the lender exerts a form of market power. We have therefore specified and estimated two versions of the model, a competitive and a discriminating monopoly version.

The competitive variant of our model rests on the idea of "competition in contracts": in equilibrium, any additional entry of a lender with loan offers cannot simultaneously attract borrowers and make strictly positive profits. Two equations then describe the equilibrium: the first is a zero-profit condition; the second is a necessary condition for contract optimality, expressing the efficiency of trading between lenders and borrowers. These two equations, once solved, give the amount lent and the interest rate charged as functions of borrower characteristics. The discriminating monopoly variant of the model rests on the idea of surplus-extraction by a monopolist: the lender maximizes profit subject to the participation constraint of borrowers. Each borrower has an outside option, which is to rent a house instead of buying one. The surplus-extraction or zero-surplus equation says that in monopolistic equilibrium, borrowers are indifferent between renting a house or accepting a mortgage contract offered by the lender. A second equation is a necessary condition for contract optimality, as in the competitive version. This is consistent with the classic result that a first-degree discriminating monopoly will propose efficient trades. These two equations again determine the loan size and interest rate as functions of borrower characteristics.

These assumptions give rise to nonlinear, bivariate econometric models which are estimated by standard maximum likelihood methods. The estimated risk-premia applied to borrowers in the competitive version are much too high to be reasonable. This indicates the presence of market power, because the estimated risk-premia are, in fact, interest-rate markups imposed on borrowers. The discriminating monopoly version of the model, in spite of its added complexity, gives much more reasonable estimated values of the risk-premium function.

We also find that the amount lent and the interest rate charged vary significantly with the borrower's income and down payment. There is a clear indication that differences in treatment depend on the occupational status of borrowers: for instance, executives will pay less to reimburse their loans than blue collar workers, everything else being equal. More precisely, the mere fact of being identified as an executive by the lender would result

[1] In contrast, there is an important strand of empirical literature on racial discrimination in mortgage lending in the U.S., which is discussed below.

in better loan conditions, even if the executive's income were in the range of a blue-collar worker's wage. Estimations also show that categories of borrowers have significantly different preference parameters, and these differences are exploited by the monopolist, as suggested by standard theory. For instance, workers have a smaller price-elasticity of demand for housing than executives. The borrower's price-elasticity of demand is inversely related with the interest-rate markup charged by the lender.

To the best of our knowledge, this type of nonlinear, structural approach had not been attempted before, and the estimations obtained show that the approach could be applied to exploit richer data sets in other countries. We also think that our model could easily be adapted to test for the presence of racial discrimination in mortgage lending, in the U.S. (see the discussion in the next to last section of this paper).

Recent structural econometric approaches to nonlinear pricing or price discrimination include the pionneering work of Ivaldi and Martimort (1994), using data on energy provision to French dairy producers, and Bousquet and Ivaldi (1997), on telephone pricing. Third-degree price discrimination on the European car market has been studied by Goldberg and Verboven (2001). Clerydes (2001) uses data on book sales to study discriminatory pricing of paperbacks and hardcovers. Leslie (2001) studies price discrimination in the sales of tickets for a Broadway show. Cohen (2000) shows that package sizes are used to price discriminate in the U.S paper towel market. McManus (2001) tests for the presence of second-degree discrimination, and therefore product design distortions, in the price-quality menus offered by coffee shops surrounding the University of Virginia. Miravete (2001) studies nonlinear tariffs and consumer choice in a menu of optional calling plans proposed by the Bell telephone company in Kentucky. Finally, Verboven (2002) uses differing driver preferences for gasoline and diesel cars in Europe to estimate the extent of price discrimination by manufacturers. Many of the above quoted studies use a form of discrete choice model of product differentiation to represent the behaviour of heterogeneous consumers[2].

Relationships with the theoretical literature on price discrimination and credit rationing

The theoretical literature on credit and banking has emphasized screening under adverse selection, proposing a theory which has the same formal structure as Rothschild and Stiglitz's (1976) model of competitive insurance markets. Among contributions to this topic, see Milde and Riley (1988), in which loan size is used as a screening instrument, and Bester (1985), establishing that collateral is a screening instrument, under asymmetric information. Calem and Stutzer (1995) have addressed the problem of racial discrimination with a theoretical screening model in which the probability of rejection of mortgage loan applications is used as a screening device: clients choose from a *menu of contracts* in which higher interest rates are traded off against higher probabilities of acceptance. This idea cannot be applied in the following, since we use data on accepted loan files only. Our theoretical model aims at modelling loans conditionally on acceptance. Brueckner (1994), and Stanton and Wallace (1998) address the delicate problem of informational asymmetries about lender mobility, and the associated risks of premature repayment and mortgage

[2] On the theory of discrete choice models and its application to oligopoly theory, see Anderson et al. (1992).

renegotiation. They construct a separating equilibrium in which borrowers with differing mobility select fixed rate mortgages with different combinations of rates and *points*[3]. This approach, which provides a good explanation for some of the observed mortgage "menus" in the United States, cannot be applied here, because the lender does not make use of points in our data set. It is also quite certain that the French provincial housing markets are characterized by much less mobility than the U.S. markets (this is an important cultural difference between the two countries which has not been studied, as far as we know). It follows that the interest-rate risk generated by prepayment on loans is apparently negligible in France, at least as a first approximation.

In the theoretical model of Brueckner (2000), borrowers self-select by choosing different initial loan-to-value ratios, high interest rates being associated with high LTV, and the unobservable borrower characteristic driving self-selection is the level of personal default costs. But Brueckner's model is a refinement of the typically American strategic default or default-option theory, which is not applicable to French data, because the French borrower's liability is not limited to the value of his (her) house. The initial LTV ratio (in fact, the related "down payment ratio") plays a role in our empirical analysis; it has a statistically significant effect as an argument of our estimated risk-premium function. We find that the risk-premium is a decreasing function of the down payment ratio. But this could simply be the mechanical result of a better collateralization of the loan. The model presented below maps the down payment, the income, and other exogenous characteristics of borrowers into (interest-rate, loan size) pairs, and the LTV ratio is endogenous. Our French lenders propose a very elementary kind of "menu", in which the interest rate varies with the loan term (a twenty year loan will typically bear a higher rate than a ten year one, everything else being equal). Thus, in principle at least, borrowers could self-select according to some unobservable characteristic or preference parameter. The question is then to discern which of these characteristics exactly is screened by the choice of the loan term. In theory, it could be many things, such as the degree of risk aversion, the rate of time preference, the probability of default, etc., and possibly a combination of several aspects simultaneously. The most common (and probably most reasonable) view is that the lower the borrower's income, the longer the chosen loan life, because, due to liquidity constraints, poorer borrowers will simply try to spread reimbursement over a longer period. The borrower's income and social status being observed by the lender, it is difficult to identify another dimension of the space of consumer characteristics along which borrowers would significantly self-select while choosing various loan durations, and which is not at the same time already observed by the banker. We have carefully taken the loan term into account, but we treat it as exogenous, *i.e.*, as if it were a characteristic of the borrower: this is clearly a simplification, but a more sophisticated treatment seemed to be out of reach.

To sum up, our French lenders, who are certainly somewhat old-fashioned, as compared to US professionals, are also working in a different legal environment, which makes it difficult to test some existing theories based on the idea of self-selection. This is an impor-

[3] The idea of separation by mobility, when borrowers are better informed than lenders about their probability of moving, has been first modelled by Chari and Jagannathan (1989), but in a model in which the interest rate is constant.

tant reason for modelling our lender as a first-degree discriminating monopolist: given our data set, there are no banking practices, and no compelling dimensions of both the borrower characteristic and the credit contract spaces along which to construct a reasonable model of self-selection or second-degree discrimination[4].

At this point, it should be noted that none of the structural econometric approaches to price discrimination cited above went as far as to use a condition that the choice of product qualities maximizes profit to help identify their model's parameters: product lines are exogenously given. It is difficult to capture taste heterogeneity in a model of demand for differentiated products and to derive a measure of the extent of price discrimination or product-quality distortions, and the assumption that a producer's product line is optimal does not seem to have been tested or used as an identification restriction in the literature. Given this difficulty, a contribution of the present paper consists in the use of assumptions about the seller's profit-maximizing behaviour (i.e., our surplus-extraction equation) to help identify structural parameters in a price-discrimination problem.

Another difficulty, pointed out above and by recent theoretical work, is that a realistic model of second-degree discrimination would be likely to entail several dimensions of uncertainty about the borrower's characteristics. For instance, at least two dimensions could be considered: the borrower's marginal willingness to pay for a larger loan *and* a parameter determining the utility level of his (her) outside option on the housing market. A good model would also probably involve several screening instruments: loan size, loan life, down payment, points, prepayment penalties, are possible instruments. Thus, a multidimensional discrimination model would be required, as studied in the work of Rochet and Choné (1998). This is known to lead — apart from hard technical problems — to much less separation of consumer types, and thus much less discrimination power than in the classic, one-dimensional model. Rochet and Choné (1998) show that bunching is a robust feature of optimal solutions in the multidimensional screening problem. The optimal solution cannot be explicitly computed in general, and with the exception of Armstrong (1999), not much has been published on the approximation of the optimal discrimination policy by simple pricing rules. A consensus on the form of the appropriate model has not yet emerged. Our approach, which is to model data as if they reflected first-degree discrimination based on observed client characteristics, is therefore justified, at least as a first step.

In the following, section 2 is devoted to a description of the data and to the results of a preliminary linear econometric investigation. Section 3 presents the two versions of the model and section 4 their econometric estimation. Section 5 is devoted to a discussion of the empirical literature on racial discrimination in mortgage lending, and section 6 contains concluding remarks.

---

[4] Why not a model of third-degree price discrimination then? This is essentially a question of terminology. The received definition of third-degree price discrimination, is linear pricing (*i.e.*, constant unit prices) combined with market segmentation with respect to observable consumer characteristics. Our model is more general than this, because it represents nonlinear pricing. On the other hand, first degree (*i.e.*, perfect) discrimination exists only in pure theory, because consumer characteristics are never perfectly observed in practice, implying that some consumers with unobservable differences are equally treated by the seller. It follows that our approach could be called "nonlinear third-degree price discrimination", instead of first-degree price discrimination.

2. The data

2.1. Description

We obtained a sample of observations on the clients of a French mortgage lender, the Crédit Hypothécaire de France (a nickname), hereafter CHF. The CHF is in fact a network of building societies, scattered on the French territory, the BSs. These local BSs have independent application screening and interest rate policies; they own in common a financial institution, which borrows money on national and international bond markets, and provides funds to the BSs. The CHF is a prudent and profitable institution, with a long history and a solid reputation. The BSs do not securitize their loans. Rating agencies have granted a very high note (AA+) to the CHF, so that the institution's cost of funds is well approximated by, and closely parallels, the long-term rate on French state bonds (the "OAT" rate), with an almost constant difference of a few base points. In the absence of sufficiently precise information on the cost of funds, we used the French OAT rate directly in the estimations. Although the CHF has a special legal status, it is fair to describe the behavior of the local BSs as profit maximization. Until 1995, when the French government reformed its housing policy, the CHF had the privilege of distributing a particular kind of state subsidized home loan. This privilege has disappeared today, since all commercial banks can now initiate the same subsidized loans, but the CHF network has developed a strong expertise in mortgage lending to the working class, and goodwill in accordance with this specialization. Its clientele is composed of a vast majority of modest income employees and workers. It is likely that many of the CHF clients would see their applications rejected elsewhere.

On top of distributing state subsidized loans, the characteristics of which are tightly regulated, the CHF also supplies the so-called "free loans", which are unregulated, ordinary mortgages. Until recently, the vast majority of these mortgages have been classic, fixed rate, fixed repayment mortgages. French mortgage law is in a sense simpler than that of the U.S., since the borrowers' liability is *not* limited to the value of the house (lenders can pursue other borrower assets to mitigate default-related losses). In addition, house prices have not decreased very much in the provincial regions, which are the geographical origin of the sampled borrowers, during the observation period. It follows that strategic default (or the exercise of the default option) is not empirically relevant in the sample. In practice, mortgage defaults seem to be triggered by consumer insolvency, mostly due to loss of income. A form of unemployment insurance of mortgage loans does indeed exist, but it is not compulsory, it is expensive, and limited in scope. Informational asymmetries and moral hazard provide an explanation for the weakness of unemployment insurance of mortgages in France (on this topic, see Chiappori and Pinquet (1999)). These loans can in principle be renegotiated, the prepayment penalty being around 3% of the principal's remaining value in all BSs.

For the econometric investigations below, we have used a sample of 2610 observations on *accepted* free loans, originated from various BSs across France between 1989 and 1994. We have eliminated the subsidized loans[5]. There is no information on rejected applications, and no observations of default or of repayment "incidents".

[5] The interest rate, the amount lent, and the borrower characteristics of subsidized loans are strictly regulated,

Each observation corresponds to a file, including, 1^o) the amount of the loan, 2^o) the loan interest rate (including insurance), 3^o) the down payment (savings used to buy the house by the borrower), 4^o) the starting year, 5^o) the loan term, 6^o) the borrower's yearly wage, 7^o) the age of the borrower, 8^o) the family size, 9^o) the borrower's occupational status, falling into 6 categories, and 10^o) the geographical location of the BS granting the loan. We kept only four of the occupational status categories, blue-collar workers (1180 observations), white-collar employees (908 observations), so-called intermediate professions (363 observations), and executives (159 observations). We also do not use the geographical location variable in the estimations, and all regions have been pooled[6].

The model presented below has been constructed to be estimated with this limited set of information. It can of course easily be adapted to use more explanatory variables.

Table 1 provides descriptive statistics on the data. The amounts lent are not very high. The down payment ratios, that is, (down payment / loan + down payment) are small, around 17%. The loan terms are distributed between 1 year and 20 years (the empirical distribution of loan terms is depicted on Figure 9). The nominal interest rates are very high with a mean value of 11,7%, but the real interest rates were also very high at the beginning of the nineties in France, the inflation rate being already quite low (around 2% per year). The interesting aspect of the data is the substantial variance of the loan rates. This will allow the estimation of a risk-premium function, and of an interest-rate elasticity of the demand for housing, in spite of the fact that the observation period is short. Another striking fact is the markup on state bond rates, which is on average equal to 3%. The lenders seem to exert a form of market power.

2.2. An exploratory linear model

To gain some understanding of our data set's content, we have estimated a simple simultaneous equation model, explaining loan sizes and rates. More precisely, we started from the point of view that the relevant endogenous variable for this study is not the interest rate itself, but the constant repayment annuity, denoted p, and defined as follows. A household who borrows for T years at continuous-time rate r, will repay the amount, $p(r,T) = r/(1-e^{-rT})$ at each instant of time to reimburse the loan completely, interest and principal, per euro borrowed (see below, subsection 3.1). The discrete-time equivalent of p, *i.e.,* the amount repaid each year, which is denoted P, is

$$P(\bar{r},T) = \frac{1-\frac{1}{1+\bar{r}}}{1-(\frac{1}{1+\bar{r}})^T},$$

limiting the banker's freedom considerably. They therefore constitute a very bad basis to test for a theory of discrimination.

[6] It would have been interesting to use geographical data but, (i), the available data is the location of the BS, not that of the borrower's house, and this regional information is coarse (we do not even know if the borrower buys a house in the countryside, in the suburbs, or in the city center); (ii), it is hard to find good data on local house prices in France; (iii) we have introduced regional dummies in preliminary explorations using linear econometric models: some of these dummies had significant coefficients, others did not; no interpretable pattern seemed to emerge; the introduction of dummies did not change the other, important coefficients.

where $\bar{r} = e^r - 1$ is the annual interest rate. If i denotes the lender's cost of funds, define $C = P(i, T)$. This is a measure of the lender's cost, which is commensurate with P. If borrowers are liquidity-constrained, P is a relevant variable, otherwise, the interest rate r would be the only important one. (On liquidity constraints, see Deaton (1992)). Let then m denote the amount lent and a the household's down payment, so that $H = m + a$ is the value of the house. Let finally w denote the household's yearly wage, and n the family size. Using the subset of 1180 unregulated loans granted to blue-collar workers to estimate a linear model explaining $log(H)$ and $log(P)$ simultaneously, we obtained the following results[7]:

$$\begin{aligned} log(H) = \underset{(23.15)}{6.239} - \underset{(-20.75)}{1.551}\ log(P) \\ + \underset{(4.187)}{0.102}\ log(w) + \underset{(19.60)}{0.193}\ log(a) + \underset{(5.823)}{0.028}\ n, \end{aligned} \qquad (A)$$

and

$$log(P) = \underset{(-19.39)}{-0.4787} + \underset{(62.82)}{0.7239}\, log(C) - \underset{(-6.96)}{0.01862}\, log\left(\frac{a}{H}\right). \qquad (B)$$

Note that a/H, the *down payment ratio*, has a significant negative impact on P, and thus on the loan rate, as expected. But the striking fact is merely the significance of $log(P)$ in equation A, which determines house value. The coefficient of $log(P)$ can be interpreted as a form of price-elasticity of housing demand. We estimated the same model separately with the subsets of white collar employees, intermediate professions, and executives; the 3SLS estimated values of the coefficient on $log(P)$ are -1.347 for white collars, -1.629 for intermediates, and -1.648 for executives, all highly significant. While white and blue collars are not significantly different from each other, white collars are significantly different from intermediate professions and executives[8]. These findings suggest that the price-elasticity of house size varies with occupational status and even seems to increase (in absolute value) with the level in the social hierarchy. In the following, we try to provide an explanation for this phenomenon, and to relate it with observed interest-rate markups.

[7] Equations A and B have been estimated with the 3SLS method, to take care of simultaneity problems, *i.e.,* to estimate simultaneous equations systems with this classic method, the instruments used are the model's exogenous variables; in this case, $log(C)$, $log(a)$, n and $log(w)$. The Student t-ratios are in parentheses. The adjusted R^2 is .42 for equation A and .77 for equation B; the estimated correlation of A and B's error terms is .244.

[8] We have estimated other variants of this system, all leading to the same conclusions. For a detailed account of an econometric study of our data set with the help of standard linear econometric methods, see Gary-Bobo and Larribeau (2003).

3. A model of mortgage lending

To describe the model, we first define the demand side, developing a continuous-time model of an expected utility maximizing, infinitely-lived and liquidity-constrained consumer-borrower. We then turn to the banking firm's profit function. This preliminary modelling work allows the presentation of two variants of the theory, one competitive, one monopolistic, corresponding to two assumptions about competition on mortgage lending markets.

3.1. The borrowers

Let α denote a household's vector of characteristics. Each household of type α is represented by an infinitely-lived, expected-utility maximizing consumer. The instantaneous utility of consumer α, denoted u, is defined on the set of bundles (c, h), where c is instantaneous aggregate consumption and h is housing, measured in constant-quality square meters. To sum up, $u = u(c, h; \alpha)$. Time t varies continuously, and it is assumed that consumer α's utility for a *certain* consumption path $t \to (c(t), h(t))$ is

$$\delta \int_0^{+\infty} e^{-\delta t} u(c(t), h(t); \alpha) dt, \tag{1}$$

where δ is a positive discount factor[9]. Now, each consumer is subject to a liquidity constraint of an extreme form: she cannot borrow against future income, and can only obtain a home loan with the entire house as a collateral. For simplicity, it is assumed that each consumer is a worker with a constant wage w. Yet, the worker is subject to a risk of default, which can be interpreted as loss of income, or as a layoff. To this, we add the simplifying assumption (with a European flavor), that once laid off, the worker remains unemployed forever. Default (or unemployment) randomly occurs at time t^*. The probability of defaulting between time 0 (the starting point of the loan here) and time t is assumed to be exponential and its cumulative distribution function is denoted F, that is,

$$F(t^* \leq t) = 1 - e^{-\theta t}, \tag{2}$$

where θ is a nonnegative parameter. The density of F is $f(t) = \theta e^{-\theta t}$. The only event triggering mortgage defaults is unemployment, and it implies a complete loss of the house, due to mortgage foreclosure by the bank, for simplicity.

We consider classical, fixed interest, direct mortgage loans, with constant repayment (or self-amortizing annuities). Mortgage loans are characterized by an amount lent m, a fixed (continuous-time) interest rate r, a term T, and a starting date T^*; they can be described by the array (r, m, T, T^*). The borrower is endowed with a down payment a; this represents money that has been accumulated in the past, and is assumed to be entirely invested in the purchase of the house. The consumer rents a house if he or she does not borrow to buy one, and the accumulated sum of money a is then deposited in some bank and yields an interest i_0. To simplify the model, we assume that $i_0 = 0$ in the following.

[9] The expected utility in equation (1) is multiplied by δ only for convenience and to simplify computations.

Now, if the house price per square meter is denoted π, then the house size (or housing consumption level) of a borrower is simply

$$h = \frac{m+a}{\pi}. \tag{3}$$

We define the *down payment ratio* of a consumer as $a/(m+a)$.

The continuous-time constant amortizing repayment of a m euros loan of term T, originated at time $T^* = 0$, is $p(r,T)m$, and satisfies by definition,

$$me^{rT} = \int_0^T p(r,T)me^{rt}dt.$$

Straightforward integration then yields,

$$p(r,T) = \frac{r}{1-e^{-rT}}. \tag{4}$$

For convenience, we call $p(r,T)$ the *price* of the loan, and denote it simply by p. Note that all continuous-time variables can easily be transformed into their discrete-time counterparts, with the appropriate formulas[10].

Then, because of the assumed liquidity constraint, the borrower faces a budget constraint at each time t. Taking consumption as a numeraire, the budget constraint has the simple form,

$$c = w - pm. \tag{5}$$

Definitions (3) and (5) show that, given w, a and π, there is a one-to-one relationship between (p,m) and (c,h). Given T, the interest rate r can be retrieved from p by inverting (4). It follows from this that a mortgage (r,m,T,T^*) is equivalently characterized by the vector (c,h,T,T^*), given w, a and π. In the following, it will be more convenient to reason in the (c,h) consumption-bundle plane, instead of the (r,m) plane.

The expected utility of a borrower

We define different levels of instantaneous utility as follows.

1^o) When consumer α repays a mortgage with price p and enjoys a house of size h, her utility level is $u = u(w-pm,h;\alpha)$.

2^o) If the consumer defaults, she loses her income and her house. In this case, her utility level is assumed to be independent of p and h; it is denoted by $u = u_0(\alpha)$.

3^o) If the consumer never defaulted, and has completely repaid the loan, then $u = u(w,h;\alpha)$.

4^o) If the consumer loses her income after the loan is completely repaid, she does not lose her house, and $u = u(0,h;\alpha)$.

5^o) Finally, the consumer can rent a house. Let ρ denote the rent per square meter. If the consumer rents, she will choose $(c,h) \geq 0$ so as to maximize $u(c,h;\alpha)$ subject to the

10 For instance, if T is expressed in years, and $\hat{r}$ denotes the annual interest rate, then $r{=}log(1{+}\hat{r})$.

constraint $c + \rho h = w$. The solution of this standard maximization problem is the pair of demand functions $(c^*(\rho, w; \alpha), h^*(\rho, w; \alpha))$. The instantaneous indirect utility of a renter is thus,

$$v^r(\rho, w; \alpha) = u[c^*(\rho, w; \alpha), h^*(\rho, w; \alpha); \alpha]. \tag{6}$$

Let now $v(t, t^*, T^*, T; \alpha)$ denote the instantaneous utility of a consumer of type α at time t, given that default occurs at time t^*, under a mortgage contract (r, m, T, T^*). The expression for v can be computed with the utility levels defined above, since by assumption, the consumer rents her house from $t = 0$ until T^*, loses her income at t^*, and loses her house at t^* if $T^* < t^* < T + T^*$. We can now define the discounted sum of utilities for the mortgage (r, m, T, T^*) as

$$U(t^*, T^*, T; \alpha) = \delta \int_0^{+\infty} v(t, t^*, T^*, T; \alpha) e^{-\delta t} dt. \tag{7}$$

Using the assumption that the date of default is exponentially distributed, we define the expected utility of a loan of term T, originated at T^*, as follows,

$$V(T^*, T; \alpha) = \int_0^{\infty} \theta e^{-\theta t^*} U(t^*, T^*, T; \alpha) dt^*. \tag{8}$$

It is now possible to compute V. After some cumbersome, but straightforward computations, we obtain the following result[11].

Result 1

$$\begin{aligned}
V(T^*, T; \alpha) = {} & v^r(\rho, w; \alpha)[1 - e^{-(\theta+\delta)T^*}]\Big(\frac{\delta}{\theta + \delta}\Big) \\
& + u_0(\alpha)[1 - e^{-(\theta+\delta)(T+T^*)}]\Big(\frac{\theta}{\theta + \delta}\Big) \\
& + u(w - pm, h; \alpha)[e^{-(\theta+\delta)T^*} - e^{-(\theta+\delta)(T+T^*)}]\Big(\frac{\delta}{\theta + \delta}\Big) \\
& + u(w, h; \alpha)\Big(\frac{\delta}{\theta + \delta}\Big)e^{-(\theta+\delta)(T+T^*)} \\
& + u(0, h; \alpha)\Big(\frac{\theta}{\theta + \delta}\Big)e^{-(\theta+\delta)(T+T^*)}.
\end{aligned} \tag{9}$$

From this result, one can derive the consumer's tenure choice conditions, with static expectations, that is, on the assumption that a mortgage characterized by (r, m, T) will always be available in the future. These tenure choice conditions are in fact equivalent to the borrower's *participation constraint*, from the point of view of the banker. Define $V^r(\rho, w; \alpha)$ as the expected utility of a consumer who rents a house forever. Clearly, with the help of (9), one easily checks that

$$V^r(\rho, w; \alpha) = \lim_{T^* \to +\infty} V(T^*, T; \alpha) = v^r(\rho, w; \alpha)\Big(\frac{\delta}{\theta + \delta}\Big) + u_0(\alpha)\Big(\frac{\theta}{\theta + \delta}\Big). \tag{10}$$

[11] Detailed computations are relegated to the supplementary material web page.

Now, consumer α will accept a loan $(r, m, T, 0)$ at time 0 if and only if, for all $T^* \geq 0$,

$$V(0, T; \alpha) \geq V(T^*, T; \alpha). \tag{11}$$

This condition has a remarkably simple equivalent formulation.

Result 2
Condition (11) is equivalent to

$$V(0, T; \alpha) \geq V^r(\rho, w; \alpha). \tag{12}$$

To prove[12] this result, it is sufficient to show that V is decreasing with respect to T^*, if and only if condition (12) holds.

From now on, and to clarify notation, we denote

$$\bar{V}(c, h, T; \alpha) = V(0, T; \alpha), \tag{13}$$

letting c and h explicitly appear as arguments of the expected utility. Parameter α must be viewed as a vector of characteristics including θ, w, a, δ.

3.2. The lender

We now construct the bank's expected profit function from a mortgage loan to a borrower of type α, starting at time $T^* = 0$. From now on, the reader must understand that all mortgages start at date $T^* = 0$.

We assume that the bank's liability structure can be modelled as if it sold bonds or certificates of deposit in exchange for an interest rate denoted i to finance its loans. The bank is also assumed to avoid any form of interest rate or refinancing risk. The timing of payments on its debt, interest and principal, matches the timing of revenues from its mortgage portfolio (amortization schedules are parallel). The loans are not securitized, and kept on the asset side of the bank's balance sheet (this assumption corresponds to common French practice). The interest rate i is used to discount the bank's future cash flows. If a client defaults at date t^*, for the sake of simplicity, we assume that the liquidation value of the house is zero. More realistic representations of mortgage foreclosure lead to complex nonlinear expressions of expected profit, and are intractable with our data (see however Appendix A, in which we sketch a more sophisticated model).

We also simplify the model by assuming that administrative costs are zero. Given that we do not have observations of the banker's costs, and that this implies difficulties to estimate cost parameters, our choice is a reasonable approximation[13]. (This point is

[12] For a proof of this result, see the supplementary material web page.

[13] A realistic view is that there is a fixed administrative cost per loan file, because there are no obvious reasons why a larger mortgage should be more costly to manage than a small one, at least as a first approximation. These costs will be small relative to the size of loans, with a small impact on the interest rate charged in equilibrium. The presence of fixed costs per branch is an additional problem, albeit not a serious one. They affect the competitive version of the model, but only the expression of the zero-profit condition in this competitive version, since the necessary conditions for profit maximization do not depend on fixed costs. These costs are likely to bias upwards slightly (*i.e.*, to add a small constant) to parameter θ in equation (17) below. In contrast, fixed costs do not intervene at all in the equations of the monopolistic version of the model: this is a clear advantage of this latter version.

discussed further in Appendix A).

With our simplifying assumptions, the date 0 net present value of a loan, conditional on t^*, denoted $b(t^*)$, is

$$b(t^*) = -m + m\int_0^{t^*} p(r,T)e^{-it}dt. \tag{14}$$

The expected profit from a loan can therefore be computed as,

$$\Pi(c,h,T;\alpha) = \int_0^T \theta e^{-\theta t^*} b(t^*)dt^* + e^{-\theta T}b(T). \tag{15}$$

Straightforward computations yield the following formula[14],

$$\Pi(c,h,T;\alpha) = -m + \frac{p(r,T)m}{p(\theta+i,T)}. \tag{16}$$

Given (16), the zero-profit condition $\Pi(c,h,T;\alpha) = 0$ leads to

$$p = p(\theta+i,T), \tag{17}$$

meaning that interest rates charged can be expressed as i plus a risk-premium θ, in competitive equilibrium.

With this formulation, a simple change of variables in (16), using (3) and (5), yields the expression,

$$\Pi(c,h,T;\alpha) = \Big(\frac{w}{p(\theta+i,T)} + a\Big) - \pi h - \frac{c}{p(\theta+i,T)}, \tag{18}$$

which is linear with respect to (c,h). The banker's marginal rate of substitution between c and h is therefore constant for each given type of borrower. More precisely,

$$\frac{\frac{\partial\Pi}{\partial c}}{\frac{\partial\Pi}{\partial h}} = \frac{1}{\pi p(\theta+i,T)}. \tag{19}$$

3.3. Competitive equilibrium in mortgage contracts

We assume here, *(i)*, that a large number of banks compete to supply loans to consumers, and *(ii)*, that a consumer's type α is observable by every bank. Banks "compete in contracts". For given T, the banker who offers the most advantageous contract $(c(T,\alpha),h(T,\alpha),T)$ will attract all the clients of type α. It follows that competition will drive profits to zero for each pair (T,α). Formally, in equilibrium, if the contract $(c(T,\alpha),h(T,\alpha),T)$ is traded, $\Pi[c(T,\alpha),h(T,\alpha),T;\alpha] = 0$. In addition, each contract $(c(T,\alpha),h(T,\alpha),T)$ in the menu offered in equilibrium, will satisfy the familiar necessary condition for optimality,

$$\frac{\frac{\partial\Pi}{\partial c}}{\frac{\partial\Pi}{\partial h}} = \frac{\frac{\partial\bar{V}}{\partial c}}{\frac{\partial\bar{V}}{\partial h}}, \tag{20}$$

for otherwise, a competitor could offer a more advantageous contract to consumers of type α, given T.

[14] For a proof of expression (16), see the supplementary material web page.

INSERT FIGURE 1 HERE

The competitive equilibrium is represented on figure 1(a). For any given T, the equilibrium contract $(c^*, h^*) = (c(T, \alpha), h(T, \alpha))$ simply maximizes $\bar{V}(c, h, T; \alpha)$ under the zero-profit constraint $\Pi[c, h, T; \alpha] = 0$, which is linear here.

3.4. The discriminating monopolist's menu of contracts

If banks are endowed with a form of market power, due to concentration in the mortgage lending sector, or due to the fact that competitors consistently reject the applications of some consumer types, then, a more appropriate model will be that of a first-degree discriminating monopolist. Perfect discrimination requires that, for each type pair (T, α), for which trade occurs, the contract $(c(T, \alpha), h(T, \alpha), T)$ should maximize $\Pi(c, h, T; \alpha)$ subject to the borrower's participation constraint (12), for given T. The necessary condition for optimality (20) will then again be satisfied for all (T, α). But the zero-profit condition should be replaced by the zero-surplus condition,

$$\bar{V}(c(T, \alpha), h(T, \alpha), T; \alpha) = V^r(\rho, w; \alpha), \tag{21}$$

which says that the entire consumer surplus is extracted by the monopolist. This situation is represented on figure 1(b): for each (T, α), the monopolist maximizes profit Π, with respect to (c, h), subject to the constraint that $\bar{V}(c, h, T; \alpha) \geq V^r(\rho, w; \alpha)$. Expected profit is decreasing with respect to c and h: the bankers always prefer to reduce house size h (lend less) for a fixed reimbursement flow pm, and they always prefer to increase the rate, that is, reduce consumption c, for a given loan size m. The optimal contract is therefore located on the lowest iso-profit line compatible with borrower participation.

4. Estimation of the models

We will first propose a way of estimating the competitive model described above, and discuss the possibilities of testing for the presence of discrimination. We then present some estimation results. The same will later be done with the discriminating monopolist model.

4.1. Specification and estimation of the competitive model

In order to obtain a reasonably tractable formulation we assume that the instantaneous utility u is quasi-linear, with the particular parametric form,

$$u(c, h) = u_0 + c + \gamma \frac{h^\epsilon}{\epsilon}, \tag{22}$$

where γ is a function of individual characteristics and ϵ, an individual preference parameter[15]. The parameter β defined as

$$\beta = \frac{1}{1 - \epsilon},$$

[15] The specification $u = \gamma_1 c + \gamma_2 \frac{h^\epsilon}{\epsilon}$ is of course equivalent to, and cannot be distinguished empirically from, (22). The ratio $\gamma = \gamma_1 / \gamma_2$ only is identifiable, jointly with ϵ. The constant u_0 plays no role in the econometric specification. These things become intuitively clear below.

with $0 < \epsilon < 1$, can be interpreted as the "price-elasticity of the demand for housing", as will be shown below. With the above specification of u, using (9) with $T^* = 0$, one gets,

$$\bar{V}(c, h, T; \alpha) = u_0 + \frac{w\delta}{\delta + \theta} e^{-(\delta+\theta)T} + \frac{c\delta}{\delta + \theta}(1 - e^{-(\delta+\theta)T}) + \frac{\gamma}{\delta + \theta}\left[\delta + \theta e^{-(\delta+\theta)T}\right]\frac{h^{\epsilon}}{\epsilon}. \qquad (23)$$

From (23), it is easy to compute the marginal rate of substitution of a borrower,

$$\frac{\partial\bar{V}/\partial c}{\partial\bar{V}/\partial h} = \frac{h^{\frac{1}{\beta}}}{\gamma K(\theta, \delta, T)}, \qquad (24)$$

where by definition,

$$K(\theta, \delta, T) = \frac{1 + (\theta/\delta)e^{-(\theta+\delta)T}}{1 - e^{-(\theta+\delta)T}}. \qquad (25)$$

We remark that if the borrower becomes very impatient, then $lim_{\delta\to+\infty}K(\theta, \delta, T) = 1$ and if the borrower becomes extremely patient, $lim_{\delta\to 0}K(\theta, \delta, T) = +\infty$.

The contract optimality condition (20) can then be rewritten,

$$h = \left(\frac{\gamma K(\theta, \delta, T)}{\pi p(\theta + i, T)}\right)^{\beta}. \qquad (26)$$

Using then the change of variable $h = (m + a)/\pi$ and taking logs yields the expression,

$$log(m + a) = \beta log(\gamma) + (1 - \beta)log(\pi) + \beta log[K] - \beta log[p(\theta + i, T)]. \qquad (27)$$

Expression (27) is very close to being a simple "demand for housing" function, in the expression of which $p(\theta + i, T)$ plays the role of a housing price.

Recall that the zero-profit condition (17) writes $p = p(\theta + i, T)$. If we assume that both equations (27) and (17) are satisfied up to a random error term, (27) and (17) become a bivariate nonlinear system.

Econometric specification

For the purpose of estimation, our bivariate model (17)-(27) should be fully specified. We consider (a, w, T) as exogenous variables. The discrete variable T could in principle be treated as endogenous, but this would lead us to a much more complicated model with three endogenous variables, and the equation determining T would be very difficult to estimate. In addition the empirical distribution of T is very much concentrated on 10, 15 and 20 years (see figure 9). This is why we treat T as exogenous. In addition, there are other exogenous observations on the households, such as family size, age, and occupational status, that can be used to explain differences in preferences for housing and default risk. Let X denote the vector of all exogenous variables, including, a, w, π, T and i. The vector

$(\beta, \gamma, \delta, \theta)$ fully characterizes a consumer. A possibility is then to assume that γ, β and θ are themselves functions of X, and to specify functional forms $\gamma(X)$, $\theta(X)$ and $\beta(X)$, with parameters to be estimated (δ could also, in principle, be specified that way).

We have been able to estimate a somewhat simplified version of (17)-(27). To limit the number of parameters, we have set $K = 1$: this is tantamount to assuming that consumers are very impatient[16]. We obtain the following model.

$$log(m + a) = \beta(X)log(\gamma(X)) + (1 - \beta(X))log(\pi) - \beta(X)log[p(\theta(X) + i, T)] + e_1, \quad (28)$$
$$log(p) = log[p(\theta(X) + i, T)] + e_2. \quad (29)$$

The perturbation vector $e = (e_1, e_2)$ is assumed normally distributed with mean zero and covariance matrix Ω. These error terms can be viewed as a mixture of several random sources: unobserved borrower characteristics, measurement errors, and possibly optimization errors on the part of economic agents. The random effect of unobserved agent characteristics on p and $m + a$ is likely to be correlated: the covariance matrix Ω is therefore *not* constrained to be diagonal, and we'll obtain an estimate of e_1 and e_2's covariance [17].

Of course, in the above equation, it must be understood that $p = p(r, T)$. (Recall that r and T are observed). The endogenous variables are simply $log(m + a)$, the logarithm of the total house price (in euros), and $log(p)$, the logarithm of the continuous-time constant reimbursement annuity. The function $log(\gamma(X))$ is a linear function of the constant, $log(w)$, the age of the borrower (in logarithms), and family size (in logarithms). The parameter β varies with occupational status dummies only (white collar, intermediate professions and executives), blue-collar workers being the reference group. Let $Exec$, $Interm$, $Whitecol$, $Bluecol$ respectively denote the executive, intermediate profession, white collar and blue collar dummies. Finally, θ is specified as follows:

$$\begin{aligned} log(\theta(X)) = {} & \theta_0 + \theta_1 log(w) + \theta_2 log(w).(Exec) \\ & + \theta_3 log(w).(Interm) + \theta_4 log(w).(Whitecol) + \theta_5 log\Big(\frac{m + a}{a}\Big), \end{aligned} \quad (30)$$

a specification in which the wage (in logs) interacts with occupational status. The model has been estimated by the maximum likelihood method. It is less simple than it seems, since $(m + a)$ appears in $\theta(X)$ and is endogenous; this gives rise to complicated Jacobian terms in the likelihood expression[18]. The price of houses π has been calibrated: we set its value at 428 euros per square meter (this corresponds roughly to \$45 per square foot), a reasonable figure for the French provinces.

Finally, information on the starting date of mortgage contracts is taken into account through the cost of funds i, which varies with time, but it is not included directly as an

---

[16] An estimate of δ has been obtained in a variant of the monopoly model (see Appendix B), but at the cost of simplifying the utility function (we then set β=1), and at the cost of estimating a simplified version of the θ function. Parameter δ is difficult to estimate.

[17] A model is structural when it is derived from first principles (which is the case here), but the error terms should ideally also have a clear interpretation. On the interpretation of error terms in structural econometric models, see J. Rust (1994).

[18] The likelihood function for this model is derived in the supplementary material web page.

explanatory variable[19]. Borrower age and family size variables do not work very well with our data set. In principle, other specifications could be estimated, such as including the occupational status variables in γ, or including them independently of the wage variable in the θ function. We have tried various specifications of the β, γ and θ functions which do not always work well[20].

Estimation results

The results are presented in Table 2. All parameters are significant, except the white-collar dummy coefficient in $\beta(X)$, the family size coefficient in $\gamma(X)$, the intermediate-professions$\times$wage, and the executives$\times$wage interaction coefficients in $\theta(X)$. All parameter estimates also have the expected sign. We will concentrate our comments on β and on the estimated risk premium function.

First, the "price-elasticity of demand for housing" β varies significantly with occupational status: for executives, $\beta = 2.1766$; for intermediate professions $\beta = 2.1475$; for white collars $\beta = 2.1231$; and for blue collars, we find $\beta = 2.1166$. That is, the lower in the professional hierarchy, the less price-elastic is the demand for housing[21]. The β of a white collar is not significantly different from that of a blue collar, but executives and intermediates do differ from blue collars in that respect.

Second, the estimated θ function shows that wages act as a discrimination device, as expected: the higher a borrower's wage, the lower the price charged. The coefficients of wage$\times$status interaction terms in the θ function are not all significant: this is disappointing, but the estimates of these coefficients will become significant, with a nice pattern, in the monopolistic version of the model below. At the same time, the inverse of the down payment ratio has a positive coefficient, as expected.

The presence of market power

The mean value of θ in the blue-collar category is $\theta(Bluecol) = 0.0398$; it is $\theta(Whitecol) = 0.0425$ for white-collars; $\theta(Interm) = 0.0367$ for intermediate professions, and $\theta(Exec) = 0.0330$ for executives. But the order of magnitude of these risk-premia is too high. Since the probability of default before term is $F(T) = 1 - e^{-\theta T}$, we can estimate $F(15)$ and the figures are 0.464, 0.469, 0.380 and 0.350 for the blue collars, white collars, intermediates and executives, respectively. These probabilities of default do not correspond to what is known *a priori* about mortgage default rates in France. A reasonable figure would be something like 3%, not 35% or 46%! The screening process in the local BSs is a kind of "old-style" job made by some *comités de crédit*. The local officers are very likely to

[19] A referee sugested that we could use this starting date information to capture shifts in the risk-premium function due to changes in the unemployment level. This is a good idea in principle, but it yields poor results in practice, presumably because changes in the unemployment rate are too slow during the estimation period

[20] With the data and the estimation methods employed here, it is reasonable to use a somewhat parsimonious specification: we should not demand too much from the data or try to estimate too many parameters.

[21] The reader could have expected the reverse ranking: we provide an explanation below for this result.

use local information on markets, borrowers[22], etc... The chief executives of the CHF claim that their network does a very good screening job and that $F(15) \approx 0.01$. This is a sufficient indication for the presence of market power: the values of θ should be interpreted as markups, not as risk-premia. The fact that these markups are inversely related to the corresponding price-elasticities of demand for housing, the βs, is an indication that price discrimination is taking place. Workers are less price-elastic than executives, they therefore pay more in (monopolistic) equilibrium. This is why we estimate a monopolistic version of the model in the next subsection.

Figures 2 to 4 represent numerical simulations of the model with the estimated parameters. Figures 2 and 3 depict the interest rate charged to borrowers as a function of the wage and of the down payment ratio, respectively. To draw these figures, the values of exogenous variables are set equal to their overall sample mean, with the exception of the variable appearing on the x-axis. The schedules show differences of treatment between social categories, everything else being equal. The mere fact of being an executive leads to a reduction of the interest rate of approximately half a percentage point, relative to white collars. Figure 4 shows the house size schedule as a function of the wage. The same type of social discrimination appears clearly.

4.2. The Monopoly Model

To establish the equations of the monopoly model, it is sufficient to replace the zero-profit condition (17) with the zero-surplus condition (21). In order to obtain the analytical expression of (21), we first compute V^r, the indirect utility of a household on the rental market. The rental demand for housing is obtained by maximizing $u(c,h)$, as specified by (22) above, with respect to (c,h), subject to the budget constraint $c + \rho h = w$, where ρ is the rent per square meter. This easily yields the instantaneous indirect utility of renting,

$$v^r(\rho, w) = u_0 + w + \frac{1}{\beta - 1}\gamma^\beta \rho^{1-\beta}, \tag{31}$$

from which we derive the expected utility of renting forever, defined by (10) above:

$$V^r(\rho, w) = v^r(\rho, w)\frac{\delta}{\theta + \delta} + u_0\frac{\theta}{\delta + \theta}. \tag{32}$$

Now, we equate $\bar{V}$, given by (23), with V^r, to obtain the analytical form of (21), and simple computations yield the following expression.

$$-pm + \frac{\beta\gamma}{\beta - 1}K(\theta, \delta, T)h^{(\beta-1)/\beta} = \frac{\gamma^\beta \rho^{1-\beta}}{(\beta - 1)(1 - e^{-(\theta+\delta)T})}. \tag{33}$$

Note that u_0 has disappeared from (33). The model to be estimated is (27) and (33). As before, we set $\delta = +\infty$, and thus $K = 1$ (see Appendix B, for an estimation of δ), and get the following bivariate model.

$$log(m + a) = \beta(X)log(\gamma(X)) + (1 - \beta(X))log(\pi) - \beta(X)log[p(\theta(X) + i, T)] + e_1, \tag{34}$$

[22] Our information on real defaults at the CHF is vague: we cannot provide a statistical analysis of defaulting borrowers in the CHF network.

$$pm = \frac{\beta(X)\gamma(X)}{\beta(X)-1}\Big(\frac{m+a}{\pi}\Big)^{\frac{\beta(X)-1}{\beta(X)}} - \frac{\gamma(X)^{\beta(X)}\rho^{1-\beta(X)}}{\beta(X)-1} + e_2, \tag{35}$$

where e_1 and e_2 are zero-mean random error terms, with a normal distribution, $y_1 = log(m+a)$ and $y_2 = pm$ are the endogenous variables, and X is the vector of exogenous variables. Of course, $p = p(r,T)$. The specifications of $\gamma(X)$ and $\theta(X)$ are exactly the same as before (θ being defined by (30); γ being a loglinear function of the constant, w, age and family size). For this estimation, we set $\beta(X) = exp[b_0 + b_1(Exec) + b_2(Interm) + b_3(Whitecol)]$. The price of a square meter and the yearly rent per square meter are calibrated at reasonable values, respectively $\pi = 428$ and $\rho = 46$, in euros.

The model is again estimated by standard, full information maximum likelihood methods[23]. Estimation results are summarized by Table 3. All coefficients are significant, except the coefficients of the intermediate professions dummy in β, of family size and age in γ, and of the white-collar×wage interaction variable in θ.

Again, the values of β can be ranked according to occupational status, estimated values being $\beta = 1.6117$ for blue collars, $\beta = 1.6002$ for white collars, and $\beta = 1.6308$ for executives.

The estimated mean values of θ for each occupational status reveal the expected ranking of risk premia, from the lowest, the executives, to the highest, the white and blue collars (see bottom of Table 3). "Social discrimination" is present, since wage-status interaction terms are significant for executives and intermediate professions in the risk-premium function θ. Workers are discriminated against, just because they are workers, not simply because their wage is low. A negative coefficient on the wage in the θ function, *which varies with occupational status*, means that the way the wage is taken into account to assess default risk depends on status. According to the banker, the richest borrowers are the less risky. The extent of the interest rate reduction which is granted for a given increase in the wage is higher, the higher the status.

The number of variables introduced in the risk-premium function is too small to guarantee that the estimates reflect prejudice against the working class (or a favorable prejudice for executives). The observed discrimination could be a form of statistical discrimination, that is, the occupational status variables are likely to act as proxy variables for unobserved factors correlated with default risk (see the discussion in section 5 below).

Finally, these estimates of θ correspond to much more reasonable probabilities of default than the competitive estimates obtained above. Computing $F(15) = 1 - e^{-15\bar{\theta}}$ with the estimated average values of the risk premium, $\bar{\theta}$, in each category, yields $F = .169$ for blue collars, $F = .119$ for white collars, $F = .055$ for intermediates and $F = .018$ for executives: the expected ranking.

We conclude that housing demand elasticities (*i.e.*, the βs) are overestimated by the competitive model, because risk-premia are also overestimated. Since equation (29) (*i.e.*, the competitive pricing equation) is likely to be an incorrect representation, upward biased risk-premia are transformed into upward biased housing demand elasticities while estimating equation (28) (*i.e.*, the house size equation) simultaneously with (29). We probably get better estimates of structural parameters with the monopoly version of the model, in

[23] The likelihood function for this model is derived in the present paper's supplementary material section.

spite of the added complexity of the zero-surplus equation (35).

We also conclude that workers seem to be discriminated against by banks, the origin of the discrimination being mostly due to differing elasticities of demand, and secondarily to differences in perceived default risks. These differences in elasticities β could simply capture the fact that blue-collar workers are more likely to see their loan applications rejected by other commercial banks, since the CHF is specialized in "social loans." Nevertheless, the results show that consumer heterogeneity is exploited by bankers to make more profit, as illustrated by the simulations.

Figures 5 to 8 represent numerical simulations of the model with estimated values of the parameters. Figures 5 to 7 illustrate the same kind of phenomena as Figures 2 to 4, but under the assumption of monopolistic behavior. Figure 8 represents the estimated reservation utility or participation levels of the social categories, as defined by (21) above, in the form of indifference curves. It is easy to see that executives consume more than the workers for any given size of the house. To compute each of these indifference curves, we used the sample mean values of the exogenous variables, except for the wage, the mean of which is evaluated in each occupational status sub-sample.

In Appendix B, we propose a variant of the monopoly model, in which the impatience parameter δ and function K are estimated. This has been done, as explained above, at the cost of drastic simplifications of the risk-premium and utility functions, to permit identification. More work could be done in this direction, but we have chosen to emphasize the links between interest-rate markups and the elasticity of the demand for housing[24].

5. Further remarks on racial discrimination in credit markets

Discrimination in credit markets has recently attracted considerable attention, and the question of deciding whether or not — and why — lenders discriminate against minority groups is a hotly debated topic among economists. The importance of the question is amplified by the fact that racial discrimination in mortgage lending has been made illegal, in the United States, by the Equal Credit Opportunity Act of 1974, and by the availability of new data sources, allowing for new econometric tests. The recent literature on this question is mostly empirical, and has concentrated on racial or sexual discrimination problems. Empirical studies of discrimination in mortgage lending have developed with the debate triggered by the contributions of Shafer and Ladd (1981) and Munnell *et al.* (1996). Several other important contributions to this literature are commented on in Ladd's (1998) survey article, and in the recent book of Ross and Yinger (2002).

A difficult problem in most empirical studies is to detect the presence of discrimination in the sense of G. Becker (1957). More precisely, a lender, or seller, is said to discriminate in the sense of Becker, if she is ready to forego profits just because of her prejudices. This form of discrimination is based on a particular "taste for discrimination" of the sellers, and is not usually considered by standard I.O. theories of price discrimination.

In contrast, a lender might treat a minority group differently, because racial or ethnic characteristics are correlated with some variables, important in the determination of credit

[24] The estimation of time preference parameters is known to be very difficult in various fields of applied econometrics, such as macroeconomics or finance.

worthiness and default risk, and which remain unobserved. This latter form of behavior is called *statistical discrimination* in the sense of Arrow (1973) and Phelps (1972). For a more recent treatment of this subject, see Loury (2002). An important difficulty stems from the fact that econometricians can never be sure of having introduced enough explanatory variables to control for possible risk differences in their estimation of default probabilities. It follows that a significant coefficient on race in regressions could only indicate that statistical discrimination is taking place.

Another difficulty is the need to separate the effect of risk from that of market power in the formation of interest rates[25]. The risk premium charged on some borrowers can also be interpreted as a standard monopolistic markup. To be more precise, it might be that the borrowers' preference parameters are correlated with their social, racial or ethnic group, because individual preferences depend on a group's particular economic conditions. Then, if market power is present, standard price discrimination can in turn become the explanation for differential treatment, without necessarily reflecting the presence of prejudice. (This type of approach, however, is not without its dangers, which would be to attribute the bulk of observed differences in treatment to taste differences correlated with race.) However, competition should tend to eliminate discrimination in the sense of Becker, since prejudiced lenders would lose business in favor of the unprejudiced. It follows that market power and discrimination in this sense must be closely interrelated.

Finally, the most difficult problem in detecting the presence of prejudice is that discrimination in the market for mortgages might reflect the existence of discrimination in other markets, such as the labour and housing markets, and thus be purely statistical in nature. Some minority groups would pay higher mortgage rates because they have higher probabilities of losing their jobs, and this, in turn, could simply be a consequence of their employer's behavior.

Much of the published work on discrimination in the mortgage market, to the best of our knowledge, has been devoted to the study of default and of credit denial rates (again see H. Ladd (1998), Ross and Yinger (2002)). In contrast, the model presented above aims at explaining the structure of *accepted* loan applications and determines loan sizes and interest rates simultaneously. This model could be used to test for the presence of discrimination in the sense of Becker, paying attention to the role of local market conditions, of preference heterogeneity and of differing default risks. In theory, the model allows one to separate, in the interest rate and in loan size differences, what can be attributed to prejudice, from the impact of differences in preferences and default risks. Variations in β, γ and δ can reflect differing preferences while θ reflects default risk differences as perceived by the banker. An observable characteristic can significantly change (β, γ, δ), and thus lead to changes in treatment by the banker.

If X contains enough information (enough control variables) to estimate a risk premium reasonably, the function $\theta(X)$ should not significantly depend on race. If it indeed does depend on race significantly, then, discrimination in the sense of Becker is taking place. Of course, if the information contained in X is not sufficient to control for differ-

[25] On the role of concentration in the explanation of interest rates, see Cavaluzzo and Cavaluzzo (1998). On the empirical relationship between concentration and discrimination, see Berkovec *et al.* (1998). Graddy (1995) finds a form of third-degree, race-based price discrimination on the New York fish market

ences in riskiness, then, statistical discrimination in the sense of Arrow-Phelps can be the explanation for a significant coefficient on race in function θ.

The model presented above allows for separation of these effects from plain discrimination effects based on observable differences in preferences. There is a danger here, however, which would be to attribute the bulk of differences in treatment to differences in preferences: minority consumers would have relatively smaller or bad quality houses just because they happen not to like nice housing! To avoid ambiguities of this type, it seems to us that race should not be introduced as a variable in the specifications of β, γ and δ.

6. Conclusion

The present contribution proposes a model of the mortgage lending market. The model can be used to test for the presence of discrimination, using only information on accepted loan applications. It rests on the idea of discrimination by the lender, based on observable attributes of the borrower. It explains the interest rate and the loan size of accepted loan applications simultaneously. We study a competitive equilibrium variant and a discriminating monopolist variant of the model, in order to take phenomena related to market power into account.

The model has been estimated with a sample of loan files originating from branches of a French mortgage lender. We reject the competitive model because estimated interest rate markups are too high to reflect default risks only. The monopolistic model gives a better account of the discrimination phenomena at work in the data. We conclude that "social discrimination" is present, in the sense that *ceteris paribus*, a member of the working class would pay higher interest rates than an (equally rich) executive, just because he or she is identified as a blue collar. Part of the differences in interest rates must be attributed to observable differences in preferences, since blue-collar workers have a significantly smaller price-elasticity of demand for housing than executives. The model also shows how borrower characteristics and interest rates affect the size of granted loans.

References

Anderson, S. de Palma, A. and J. Thisse (1992), *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge.

Armstrong, M. (1999), "Price Discrimination by a Many-Product Firm", *Review of Economic Studies*, 66, 151-168.

Arrow, J. K. (1973), "The Theory of Discrimination", in A.O. Ashenfelter and A. Ries, eds. *Discrimination in Labour Markets*, Princeton University Press, Princeton, New Jersey.

Becker, G. (1957), The Economics of Discrimination, University of Chicago Press, Chicago, Illinois.

Berkovec, J. A., Canner, G. B., Gabriel, S. A. and T. H. Hannan (1998), "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending", *Review of Economics and Statistics*, 80, 241-250.

Bester, H. (1985), "Screening vs. Rationing in Credit Markets with Imperfect Information," *American Economic Review*, 75, 850-855.

Bousquet, A. and M. Ivaldi (1997), "Optimal Pricing of Telephone Usage: an Econometric Implementation", *Information Economics and Policy*, 9, 219-239.

Brueckner, J. K. (1994), "Borrower Mobility, Adverse Selection, and Mortgage Points", *Journal of Financial Intermediation*, 3, 416-441.

Brueckner, J. K. (2000), "Mortgage Default with Asymmetric Information", *Journal of Real Estate Finance and Economics*, 20, 251-274.

Calem, P. and M. Stutzer (1995), "The Simple Analytics of Observed Discrimination in Credit Markets", *Journal of Financial Intermediation*, 4, 189-212.

Cavaluzzo, K.S. and L.C. Cavaluzzo (1998), "Market Structure and Discrimination: The Case of Small Businesses", *Journal of Money, Credit and Banking*, 30, 771-792.

Chari, V. V., and R. Jagannathan (1989), "Adverse Selection in a Model of Real Estate Lending", *Journal of Finance*, 36, 471-483.

Chiappori, P. A. and J. Pinquet, (1999), "L'assurance-chômage des emprunteurs", *Revue française d'économie*,14, 91-115.

Clerydes, S. K., (2001), "Product Selection with Multi-Peaked Preferences in Book Publishing", University of Cyprus, Nicosia, Cyprus, mimeo.

Cohen, A. (2000), "Package Size and Price Discrimination in Paper Towels", University of Virginia, mimeo.

Deaton, A. (1992), *Understanding Consumption*, Clarendon Lectures in Economics, Oxford University Press, Oxford.

Gary-Bobo, R.J., and S. Larribeau (2003), "The Bank's Market Power and the Interest-Rate Elasticity of Demand for Housing: An Econometric Study of Discrimination on French Mortgage Data", *Annales d'Economie et de Statistique*, special issue, forthcoming.

Goldberg, P.K., and F. Verboven (2001), "The Evolution of Price Dispersion in the European Car Market", *Review of Economic Studies*, 68, 811-848.

Graddy, K. (1995), "Testing for Imperfect Competition at the Fulton Fish Market", *Rand Journal of Economics*, 26, 75-92.

Ivaldi, M. and D. Martimort (1994), "Competition under Nonlinear Pricing", *Annales d'Economie et de Statistique* , 34, 72-114.

Ladd H. F. (1998), "Evidence on Discrimination in Mortgage Lending", *Journal of Economic Perspectives*, 12, 41-62.

Leslie, P.J.(2001), "Price Discrimination in Broadway Theatre", UCLA, mimeo.

Loury, G.C. (2002), *The Anatomy of Racial Inequality*, Cambridge University Press, Cambridge, Massachusetts.

McManus, B. (2001), "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee", Washington University-St Louis, Olin School of Business, mimeo.

Milde H. and J. Riley (1988), "Signaling in Credit Markets", *Quarterly Journal of Economics*, 103, 101-130.

Miravete, E. (2001), "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans", University of Pennsylvania, mimeo.

Munnell, A.H., G.M.B. Tootell, L.E. Browne and J. McEneaney (1996), "Mortgage Lending in Boston: Interpreting HMDA Data", *American Economic Review*, 86, 25-53.

Phelps, E. S. (1972), "The Statistical Theory of Racism and Sexism", *American Economic Review*, 62, 659-661.

Rochet, J-C. and P. Choné (1998), "Ironing, Sweeping, and Multidimensional Screening", *Econometrica*, 66, 783-826.

Ross, S.L. and J.M. Yinger (2002), *Color of Credit: Mortgage Discrimination, Research Methods, and Fair Lending Enforcement*, MIT Press, Cambridge, Mass.

Rothschild M. and J. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *Quarterly Journal of Economics*, 90, 630-649.

Rust, J. (1994), "Structural Estimation of Markov Decision Processes", Chapter 51 in R.F. Engle and D.L. McFadden eds., *Handbook of Econometrics, Volume IV*, North Holland Elsevier, Amsterdam.

Shafer, R. and H.F. Ladd (1981), *Discrimination in Mortgage Lending*, MIT Press, Cambridge, Mass.

Stanton, R. and N. Wallace (1997), "Mortgage Choice: What's the Point?", *Real Estate Economics*, 26, 173-205.

Verboven, F. (2002), "Quality-based Price Discrimination and Tax Incidence: Evidence from Gasoline and Diesel Cars", Catholic University of Leuven, Belgium, mimeo.

Appendix A. The Expected Value of Foreclosures

A more complex version of the expected profit function derived as expression (16) above takes the expected value of foreclosures into account. Let $Z(t^*, m, T; \alpha)$ denote the liquidation value of borrower α's house, given that default occurs at date t^*. Function Z describes the expected net revenues of foreclosure. For convenience, we assume that $Z = 0$ for all $t^* \geq T$. We can then reformulate the banker's profit function, conditional on t^*, as

$$b(t^*, c, h, T; \alpha) = -m + m \int_0^{t^*} p(r, T)e^{-it}dt + e^{-it^*} Z(t^*, m, T; \alpha).$$

Taking the expectation of b with respect to t^*, we get the expected profit function,

$$\Pi(c, h, T; \alpha) = -m + \frac{pm}{p(\theta + i, T)} + \bar{Z}(m, T; \alpha)$$

where,

$$\bar{Z}(m, T; \alpha) = \int_0^{t^*} \theta e^{(\theta+i)t^*} Z(t^*, m, T; \alpha)dt^*,$$

is the expected present value of foreclosure. Now, Z itself should be specified more precisely. Let $\zeta(m; \alpha)$ denote the liquidation value of the house. Then, a reasonable specification of Z can be written,

$$Z(t^*, m, T; \alpha) = Min\Big[mp\frac{(1 - e^{-i(T-t^*)})}{i}, \zeta(m; \alpha)\Big],$$

where $(mp/i)(1 - e^{-i(T-t^*)}) = pm \int_{t^*}^{T} e^{-i(t-t^*)}dt$ is the value of the remaining debt (principal and interest), when default occurs at date t^*. Now, to evaluate Z accurately, some knowledge of future house prices, conditional on α and t^* is required. To be more precise, ζ should depend on the banker's house price expectations. This kind of information is not present in our data set; it is always a problem to model expectations, and the expression of $\bar{Z}$ is not simple. If we accept the simplifying assumption of constant house prices, and if ζ is never high enough to ensure the banker completely against losses in case of foreclosure (*i.e.*, if the banker does not believe that house prices will be high enough to recoup all losses in case of default), the expression of $\bar{Z}$ can be studied numerically. According to our simulations, $\bar{Z}$ happens to be approximately linear with respect to $(m + a)$. Let us then use a linear approximation and assume,

$$\bar{Z} = z_1 + z_2(m + a)$$

where z_1 and z_2 are parameters (to be estimated). We these assumptions, it easy to see that the zero-profit condition becomes,

$$pm = (1 - z_2)p(\theta + i, T)\Big[(m + a) - \Big(\frac{a + z_1}{1 - z_2}\Big)\Big],$$

and that the contract optimality condition becomes,

$$\frac{h^{1/\beta}}{\gamma(\alpha)} = \frac{1}{\pi(1-z_2)p(\theta+i,T)}.$$

There are problems to estimate these two equations. To get an intuition of the difficulties, assume to simplify that $a = 0$ and $z_1 = 0$, then, an identification problem will arise, because the respective magnitudes of $(1-z_2)$ and of the constant in the θ function cannot be determined.

Assume now that the value of foreclosures is zero, but that there is a fixed administrative cost k per loan. We easily obtain the zero-profit constraint, in this case, by setting $z_2 = 0$ and $z_1 = -k$. Equation (17) can then be rewritten as,

$$p = p(\theta+i,T)\Big(1+\frac{k}{m}\Big),$$

and the contract optimality condition is still given by (27). Again, there is a problem to estimate the value of k in this version of the model, and it is easy to see that, k/m being small in general, the above formulation is not very different from (17).

Assume now that the value of foreclosures is zero, that the cost per loan file is zero, but that there are nonzero fixed costs in each branch, denoted A. This changes the zero-profit condition (17) again. To get an intuition of these changes, and to simplify the discussion, assume that there is only one risk class of consumers (the discussion becomes more involved with several types of borrower, but leads essentially to the same conclusions). The identical clients borrow m euros for T years and pay the price p. The profit of a BS can be expressed as,

$$\Pi = -A - Nm + \frac{Npm}{p(\theta+i,T)},$$

where N is the number of identical borrowers. The zero-profit condition is then,

$$p = p(\theta+i,T)\Big(1+\frac{A}{Nm}\Big).$$

In the absence of information on A and N, the multiplicative term $(1 + A/Nm)$ cannot be easily estimated. This term is also likely to be close to one, for Nm is a huge sum of money. In the competitive version of the model, fixed costs will bias the estimation of the constant in the θ function upwards.

To sum up, we have chosen a reasonable and tractable approach, which works well numerically: the joint estimation of (17) and (27). This approach permits one to estimate a risk-premium function θ jointly with preference parameters (in the β and γ functions). The chosen approach is such that estimated parameters of the θ function probably reflect the expected liquidation value of the house, and the quality of the mortgage as a collateral, as well as the personal characteristics of the borrower. In the competitive version of the model, estimated parameters also reflect administrative costs. Ideally, we would have preferred to separate completely, on the one hand, the impact of borrower characteristics on default rates from that of expected property prices and of the down payment ratio (both contributing to the quality of the mortgage), on the other hand.

But this is not the only difficulty: in the data set, the effect of expected liquidation values is entangled with substantial market power phenomena, *i.e.* with the markup element in the θ function, and therefore difficult to identify. We think that the markup element dominates in the data, and that the role played by expected liquidation values is of secondary importance.

Appendix B. A Variant of the Monopoly Model

We also estimated a variant of the monopoly model, with a more parsimonious specification of utility, that is,

$$u(c,h) = u_0 + c + \gamma log(h), \tag{36}$$

but we tried to estimate the impatience parameter δ. With this specification, we find that the instantaneous reservation utility of borrowers writes,

$$v^r(\rho, w) = u_0 + w - \gamma + \gamma log\Big(\frac{\gamma}{\rho}\Big). \tag{37}$$

The necessary condition for optimality (20) now writes,

$$h = \frac{\gamma K(\theta, \delta, T)}{\pi p(\theta + i, T)}, \tag{38}$$

where K is still defined by (25). With the above specification of u, using (9) with $T^* = 0$, one gets,

$$\bar{V}(c,h,T;\alpha) = \frac{\delta c}{\delta + \theta}(1 - e^{-(\delta+\theta)T}) + \frac{\gamma log(h)}{\delta + \theta}\Big[\delta + \theta e^{-(\delta+\theta)T}\Big] \\ + \frac{w\delta}{\delta + \theta} e^{-(\delta+\theta)T} + u_0. \tag{39}$$

Then, equating (39) with $V^r(\rho, w; \alpha)$, using (37), yields the specific form of the zero-surplus condition (21), that is, after some computations,

$$(c - w)(1 - e^{-(\delta+\theta)T}) + \gamma log(h)(1 + (\theta/\delta)e^{-(\delta+\theta)T}) = \gamma[log(\gamma/\rho) - 1]. \tag{40}$$

The system to be estimated is (38) and (40), with $c - w = -pm$ and $h = (m + a)/\pi$. Taking logs on both sides of (38), we find the system,

$$log(m + a) = log(\gamma(X)) + log[K(\theta, \delta, T)] - log[p(\theta + i, T)] + e_1, \tag{41}$$

$$pm = -\gamma(X)\frac{[1 + log(\gamma(X)/\rho)]}{[1 - e^{-(\delta+\theta)T}]} + \gamma(X) log\Big(\frac{m + a}{\pi}\Big) K(\theta, \delta, T) + e_2, \tag{42}$$

where e_1 and e_2 are zero-mean random error terms, with a normal distribution, $y_1 = log(m + a)$ and $y_2 = pm$ are the endogenous variables, and X are exogenous variables.

We assume that for each individual j, θ can be expressed as $\theta_j = -(1/T_j)log(1 - F)$ where F is a common probability of default. The parameters to estimate are $D = log(\delta)$, F, ρ and the parameters of function γ, which is specified as follows,

$$log(\gamma(X)) = \gamma_0 + \gamma_1 log(w) + \gamma_2 log(age) + \gamma_3 log(T) \\ + \gamma_4(Exec) + \gamma_5(Interm) + \gamma_6(Whitecol). \tag{43}$$

This model was difficult to estimate. We estimated $\delta = e^D$ and the γs by standard maximum likelihood for fixed ρ and F, and then estimated ρ and F by a grid-search procedure. The results are presented in Table 4. All the γ parameters are significantly different from zero, except the coefficient on the white-collar dummy, meaning that white collars can be merged with the blue collars. The estimated value of δ is 0.158, a reasonable figure. Given that we assume liquidity constrained borrowers, they must be impatient enough to be willing to borrow, so that it is reassuring to find δ above the maximal interest rates in the sample. The estimated rent ρ, which is FFrancs 294 per square meter and per year is also reasonable, given that our observations are outside the very expensive Paris area (this makes a rent of \$4.41 per square foot and per year, or an approximate monthly rent of \$202 for a two-rooms flat of 550 square feet). Finally, the estimated value of F is 0.011, very close to the expected 1% value. This, again, corresponds to a much more realistic average estimated value, $\bar{\theta} = 0.00073$, than those obtained with the competitive model. In fact θ seems to be of the order of magnitude of 1 to 10 base points.