A FRAMEWORK FOR ESTIMATING STRUCTURAL MODELS OF MORTGAGE DEBTORS' BEHAVIOR

JUAN ESTEBAN CARRANZA

ABSTRACT. Empirical techniques are discussed to estimate structural models of mortgage holders' behavior. The discussed methodologies yield estimates of the primitives of the model that allow computation of default probabilities and consistent simulation of counterfactual equilibria. Techniques to estimate static version of the model, as well as dynamic ones are discussed. It is shown that popular multinomial techniques impose severe structural restrictions on the underlying behavioral model. The framework is general enough to allow for multiple modelling variations and is a potential contribution to the empirical literature on mortgage default and pricing. No results are shown yet.

1. INTRODUCTION

This paper discusses alternative approaches to estimate models of behavior of mortgage debtors. Specifically, the focus is on the understanding of default decisions. The goal is the development of empirical techniques that yield estimates of the structural parameters of the model that generates observed behavior. Such estimation would allow the computation of default probabilities and the evaluation of counterfactual equilibria in a manner that is consistent with an underlying economic model.

The literature on the behavior of homeowners with mortgages has relied on contingent claims models *a la* Black and Scholes, treating mortgages and houses as any other financial asset. On the empirical side, it is very common the use of multinomial qualitative regressions, such as proportional hazard models, to correlate behavior with the theoretically relevant variables. Such models are difficult to tie to an underlying behavioral model that incorporates the specificities of mortgage financing; therefore the empirical models and the obtained estimates are difficult to interpret as more than empirical regularities. For example, it will be shown below that multinomial choice models impose structural restrictions on the underlying behavioral model, that most of the times would be difficult to justify.

This approach also ignores many of the singular features of home financing, including the fact that in most cases debtors have a specific preference match with their own house which is difficult to trade in the market; therefore, it is often found that debtors don't default on their mortgages when it's apparently in their benefit to do so (i.e. when the default option is "in the money"). Instead of justifying such behavior on the grounds of irrationality or ignorance (or "woodheadness" as its often called), the approach of this paper treats such deviations from theoretically

Date: June, 2005.
This is a very preliminary and unquotable draft.

correct behavior as the reflection of underlying states that affect debtors' choices but that are unobserved by the econometrician.

Much of the discussion below borrows from the empirical IO literature. The problem of debtors is treated as a dynamic choice problem under uncertainty: each period, the debtor decides whether to default on the debt or keep on paying to retain the option value of defaulting in the following period. Alternatively, a static version of the model is considered with debtors abstaining from defaulting on their loan as long as static payoffs are positive. The implications and limitations of each specification are discussed.

Despite the general applicability of its main ideas, the model discussed below is framed for its application with data from the Colombian mortgage market during the last years of the 1990's. During this time (and similarly in other Latin American Economies) the Colombian economy suffered a severe credit crunch that inhibited significatively the sale of new mortgages. Therefore, the model abstracts, form example, from the issue of new mortgage buyers; it also assumes that every debtor faces the same interest rate, which was a particular feature of the Colombian housing financing system. Nevertheless, the proposed techniques can be accommodated to allow for more general conditions (more on this later).

Identification and estimation of the model depends on the amount and quality of the available data. In an ideal scenario individual mortgage holders, their characteristics and the characteristics of the involved real estate assets are observed over time. Estimation alternatives for different data availability are discussed. The basic idea is to generate a structural model that yields specific predictions and to find the vector of parameters that best matches the predictions of the model to observed data according to a pre-specified metric. It will also be discussed how outside information (i.e. from household surveys) can be incorporated.

The order of the discussion is as follows: first, a somewhat general behavioral model of mortgage holders' behavior is described. Then, different approaches to estimate the primitives of the model are discussed: on one side, it is shown how static versions of the model can be estimated. Additionally, the general framework for estimating the dynamic model is discussed. The discussion is general enough to accommodate modelling variations, and specifics of the estimation will depend on the features of the available data and the focus of the study.

2. The modelling of mortgage debtors behavior

Consider a debtor i with an outstanding mortgage of size K_{it} at time t on an asset ("home") with characteristics x_i. He will continue to pay as long as the utility obtained from consuming the asset, net of payments $R(.)$, plus the expected continuation value V' is bigger than the utility from "defaulting" on the debt:

$$u(x_i, \theta_i) - \alpha(Y_{i,t} - R(K_{i,t}, r_t)) + \beta EV' > L(P_{i,t} - K_{i,t}, .) + W_t \tag{2.1}$$

In the preceding equation Y_i is the debtor's income; payments $R(.)$ depend on the size of the debt K_{it} and the interest rate rt, which is assumed to be constant across debtors. The value of "default" is given by a function of the difference between the perceived price of the asset P_{it} and the outstanding debt plus the value W_t associated with the search of a new home. In a general setup, W_t, which is presumably negative, is determined endogenously from the general house search problem, but for now we will abstract from its determination. The parametrization of the utility function will be discussed later.

On the other side, this continuation value is determined endogenously and can be obtained from the solution of the value function of the described problem:

$$V_i(Y_{i,t}, K_{it}, r_t, P_{i,t}, S_t) = max\{u(x_i, \theta_i) - \alpha(Y_i - R(K_{it}, r_t)) + \beta EV_i(Y_{i,t+1}, K_{i,t+1}, r_{t+1}, P_{i,t+1}, S_{t+1}), [L(P_{i,t} - K_{i,t}, S) + W_t]\} \tag{2.2}$$

where β is the discount rate and S is a set of yet unspecified state variables. Notice that given a parameterized version of $u(.)$ and for a given set of parameters, we could compute (2.2), if we knew the transition of the state variables and if we observed all the relevant characteristics of the asset (including its price).

Notice additionally that the model above does not necessarily assume that "default" implies that the home is somehow forfeited immediately, leaving the debtor with the difference between the price and the outstanding debt. This may or may not be true, depending on the specific contractual conditions of the mortgage and the institutional environment or the periodicity of the data. For example if default only implies that the home is forfeited with some probability $F(.)$, then we could specify $L(.) = F(.)[P_{i,t} - K_{i,t}] + (1 - F(.))[\beta EV_i(Y_{i,t+1}, K_{i,t+1}, r_t + 1, P_{i,t+1}, .)]$. The probability $F(.)$ would depend on a set of state variables that would have to be incorporated in S. As can be seen, the specification can be modified accordingly to incorporate the specifics of the contractual arrangements or the focus of the study.

3. Estimation

Before discussing how the model can be identified and estimated, we must specify the source of randomness in the model. As is usual in similar models, we will assume that there is an additive choice- and individual-specific *iid* random disturbance associated with each choice $\epsilon_{i,t}$. This randomness is an unobserved state variable that corresponds to interactions between unobserved (by the econometrician) characteristics of the given house and individual preferences. More intuitively, this shock quantifies the unobserved preference match between the debtor and her home.

There are two types of estimation strategies, both of which are discussed below. On one hand, if we have micro-level data with matching information of houses and debtors characteristics, then the single randomness mentioned above should be enough to generate a model that can be estimated via maximum likelihood. On the other hand,if default decisions are known only at the aggregate level and over time, then an additional randomness must be specified –common to all individuals with the same choice.

In addition to these two slightly different estimation approaches, we can also generate a static version of the model, which is a particular case of the general model, but is easier to estimate. The discussion below follows all this possible cases, beginning with the simplest case.

3.1. A particular case: a static model with and without micro-level data.

Suppose we observe a sample of individual information on the debtors, the involved real estate and matching information on their default behavior. Assume that debtors optimize miopically, disregarding the dynamic value of their problem (i.e. they disregard the option value of defaulting in the future). By adding an additive random term $\epsilon_{i,t}$, we can reformulate (2.1):

$$u(x_i, \theta_i) - \alpha(Y_{i,t} - R(K_{i,t}, r_t)) - [L(P_{i,t} - K_{i,t}, S) + W_t] + \epsilon_{i,t} > 0 \tag{3.1}$$

Notice that this model generates a conventional logit or probit model. Specifically, suppose that $u(.)$ is linear on the characteristics of the "house" and that its constant term is also a linear function of the characteristics of the debtor; assume also that $L(.)+W$ is a linear function of the observed states. Depending on whether $\epsilon_{i,t}$ is assumed to be a extreme value or standard normal deviate, these assumptions generate a simple linear probit/logit model. Debtor i does not default as long as the following condition holds:

$$\gamma_0 + \gamma_D D_{i,t} + \gamma_x x_j - \alpha(Y_{i,t} - R(K_{i,t}, r_t)) - [L(P_{i,t} - K_{i,t}, S) + W_t] + \epsilon_{i,t} > 0 \tag{3.2}$$

where $D_{i,t}$ is a vector of individual characteristics of debtor i at time t. Notice that eventhough the parameters of $L(.)$ are not identified separately from the preference parameters, its estimates are structurally stable.

The preceding equation implicitly assumes not only that debtors have no dynamic concerns, but also that debtors' utilities are identical up to a constant term, which may be too strong an assumption, specially if we have reasons to believe that the sensitivity of individual preferences to changes in mortgage payments is correlated with the debtors demographics, or that default behavior is correlated with assets characteristics.

We can generalize slightly the preceding model to allow for fully heterogeneous debtors by letting preference parameters depend on individual characteristics. Default behavior depends on:

$$\gamma_0 + \gamma_i(D_{i,t}, \sigma_\gamma)x_j - \alpha_i(D_{i,t}, \sigma_\alpha)(Y_{i,t} - R(K_{i,t}, r_t)) - [L(P_{i,t} - K_{i,t}, S) + W_t] + \epsilon_{i,t} > 0 \tag{3.3}$$

We would have to specify a parametric distribution for γ_i and α that would depend on demographics and parameters σ_γ and σ_α, that are to be estimated. This would in general yield a non-linear probit/logit model, depending on the assumed distribution of the unobservables. As usual, estimation of this model would proceed via numerical maximum likelihood. For example, in the logit case, the likelihood function would be based on the predicted non-default probability:

$$Pr[u(x_i, \theta_i) - \alpha_i(Y_{i,t} - R_{i,t} - L_{i,t}) > 0] = \frac{exp(u(x_i, \theta_i) - \alpha_i(Y_{i,t} - R) - L_{i,t})}{1 + exp(u(x_i, \theta_i) - \alpha_i(Y_{i,t} - R_{i,t} - L_{i,t}))} \tag{3.4}$$

The discussion above assumed that detailed information on individual behavior and characteristics was available. If no detailed information is available on individual default behavior, we can still estimate the model by computing the default probability and matching it to its sample analog, the overall default rate. In the logit case, the predicted no-default probability at time t is:

$$\int \frac{exp(u(x_i; \theta_i) - \alpha_i(Y_{i,t} - R_{i,t}) - L_{i,t})}{1 + exp(u(x_i, \theta_i) - \alpha_i(Y_{i,t} - R_{i,t} - L_{i,t}))} dF_{\theta,\alpha,Y} \equiv \wp_t \tag{3.5}$$

The integral is taken with respect to the assumed distribution of θ and α which would also depend on demographics. Numeric techniques can be used to compute $\wp$ if all parameters are known; in addition, information on the joint distribution of demographics and product characteristics can be incorporated into the computation by using simulation techniques.

Notice that (3.5) above yields the predicted aggregate default behavior and if the model is correctly specified this prediction must be *identical* to the observed default rate, i.e. $\wp_t = p_t$ where p_t is the observed proportion of debtors who don't default on their loans. Therefore, to avoid overfitting of the model, an additional random component must be specified. This random component can be a measurement error or an unobserved state variable that is common to all debtors.

The good news is that advantage can be taken of the identity of the prediction and the data to solve directly for this randomness, given any value of the parameters that are to be estimated. This error can then be interacted with a set of instruments to construct moment conditions, that allow the numeric estimation of the parameters. The precise interpretation of this randomness is crucial for the construction of the moment condition, since valid instruments should be independent of this error component.

3.2. **Estimating the dynamic model of default behavior with micro data.** The discussion above abstracted from two issues. On one hand, it assumed away any dynamic concerns by setting the continuation value V' and W in (2.1) equal to zero (or a constant). As a result of this assumption, obtained estimates of α –a parameter of central importance – would presumably be biased downwards, since debtors should in principle be willing to withstand transitorily negative static payoffs in face of transitory income shocks, just to hold on to their homes.

On the other hand, by setting the reservation value equal to zero (the right hand side of (2.1)), the model is assuming away any wealth effects induced by changes in the price of the real estate and the size of the debt. The inclusion of the interest rate and the size of the debt on the linear and non-linear models discussed above may correct the problem to some extent. It must be emphasized, though, that the price to be included in such estimation should be the perceived *current* price of the asset (based on carefully computed price indices), not the price at which it was bought, specially if there is cross-sectional variation in asset price variation rates.

In any case, in non-stationary environments the use of the static models discussed before doesn't yield the structural parameters of the model, which are needed if we're interested in analyzing counterfactual equilibria or doing policy analysis. Whether such feature is empirically relevant in specific cases is a matter of judgement. Now we turn to the estimation of the complete dynamic model.

Suppose for now that we have micro data. The dynamic model of individual behavior can be described by the following Bellman equation:

$$\begin{aligned} V_j(S_{i,t}) = max\{&u(x_i,\theta_i) - \alpha(Y_i - R(K_{it}, r_t)) + \epsilon_{i,t} \\ &-L(P_{i,t} - K_{i,t}) - W_t + \beta EV_j(S_{i,t+1}), 0\} \end{aligned} \tag{3.6}$$

where $S_{i,t}$ stands for the set of space variables of consumer i at time t; it will be assumed that ϵ is an extreme value deviation that has the same interpretation as above. Notice that we can take expectations of the equation above to obtain the function $EV(S)$ which, given a set of parameters and known transitions of the state variables, can be computed using numeric techniques, provided that the state space is compact.

Let $\delta_{i,t} \equiv u(x_j,\theta_i) - \alpha_i(Y_{i,t} - R(K_{i,t}, r_t)) - L(P_{j,t} - K_{i,t}) - W_t$. The implied non-default probability of each debtor is obtained similarly as above:

$$Pr[\delta_{i,t} + \beta EV_i(S_{i,t+1}) > 0] = \frac{exp(\delta_{i,t} + \beta EV_j(S_{i,t+1}))}{1 + exp(\delta_{i,t} + \beta EV_i(S_{i,t+1}))} \tag{3.7}$$

We can assume that consumers are identical up to a constant term (the dynamic counterpart of (3.2) above) or we can correlate θ_i with a set of demographics as in (3.3). In both cases this probability would serve as the basis for a maximum likelihood estimation. Numerically, the vector of parameters must be found that maximizes the likelihood of the sample. Notice that each evaluation of the likelihood function requires the computation of a fixed point to obtain the function $EV_i(S_{i,t+1})$.

This technique requires that we estimate beforehand the transition of the relevant state variables. Specifically, the intertemporal probability of individual income, interest rate and real estate prices must be estimated. We could also modify the specification above and incorporate the probability of foreclosure into the continuation value as explained in section 2; this probability can be estimated non-parametrically from observed behavior and would most probably depend on the elapsed time since initial default, which would be incorporated as a state variable.

There are computational issues that have to be addressed. Specifically, the size of the state space imposes limitations on the ability to compute the model fast, due to the curse of dimensionality. These limitations would have to be addressed on a case by case basis.

In (3.7), the identification of the individual components of δ is not clear. The "search" value W_t is assumed to be a time-changing parameter, constant to all debtors; if there is no other fixed-time effect, then we can presumably identify it from our data panel. We have already pointed out that this is an endogenous value that may in general depend on the characteristics of debtors, but of course identifying an individual- and time- specific effect is virtually impossible. We may posit a parametric form for W that would depend on the debtor's characteristics. Even though such function would not be separately identified from the part in $u(.)$ that also depends on debtors' characteristics, obtained estimates should be structurally consistent and stable.

3.3. **The dynamic model of default behavior with aggregate data.** When only aggregate data is observed, things are more complicated. After normalizing the utility of The predicted default probability $\wp$ yielded by the model is:

$$\int \frac{exp(\delta_{i,t} + \beta EV_i(S_{i,t+1}))}{1 + exp(\delta_{i,t} + \beta EV_i(S_{i,t+1}))} dF_{\theta,\alpha,Y} \equiv \wp_t \tag{3.8}$$

Again, integration is made with respect to the assumed distribution of preference parameters. In this case, estimation should proceed by equating the predicted default probability to its sample analog. Again, as explained before, given the identity between predicted and observed default rates, it should possible to solve for the implied unobserved randomness that is common to all debtors who choose to default, given any vector of parameters. This error can then be interacted with instruments to construct a set of moment conditions.

The difficulty is computing the integral in (3.8). Again simulation techniques are available, but then each simulated draw of the joint distribution of the parameters requires the solution of a fixed point for computing $EV(.)$. The computational burden is significative, because the efficient computation of the integral requires the use of as many simulations as possible for *each* evaluation of the parameters. Additionally the computational burden of computing each point increases exponentially in the number of state variables. The applicability of the technique should then be evaluated on a case by case basis.

4. Final remark

We have addressed the problem of estimating structural models of behavior of mortgage debtors. We are interested in uncovering the primitives of a structural model which is dynamic and heterogeneous across debtors. The discussion regarding the estimation of the model progressed from a very simple static case with homogeneous debtorss to models with heterogeneous debtors to dynamic models. Two types of data availability have been considered: micro data with individual information, and aggregate data with additional external information regarding the distribution of variables of interest. The overview has been general enough to allow for specific modelling variations. Definite application of the discussed techniques will, nevertheless, depend on the specifics of the available data and the focus of the study.

A final word of caution. Throughout the paper we have assumed that the sample of debtors is random. If there are reasons to believe that the characteristics of debtors change systematically over time, due for example to cyclical changes in financing restrictions, then this condition doesn't hold. If that is the case and such effect is deemed to be empirically significant, the more general model that generates observed debtors should be the starting point of the analysis. It may still be possible to generate simple estimation techniques, but as usual the specifics will depend on the case.

University of Wisconsin-Madison
E-mail address: juanes@ssc.wisc.edu