Concentration risk in mortgages portfolios: the role of geographic factors and loans characteristics in a multi-factor framework

Michel DIETSCH[1,2]
Joël PETEY[3]
Adeline VANDAELE[4]

PRELIMINARY – PLEASE DO NOT QUOTE

July 2010

Abstract: This paper evaluates the potential of concentration risk in mortgages portfolios. In large portfolios of small exposures, concentration risk comes from correlated defaults among groups of borrowers. Consequently, measurement of concentration risk needs to take into account borrowers' heterogeneity. In this paper, we extend the standard asymptotic single factor framework to introduce additional factors of systematic risk varying across borrowers' population. A generalized linear mixed model is used to estimate the dependence structure among exposures subject to the influence of multiple risk factors. Risk concentration is measured using a database of more than 340 000 French mortgages, including the quarterly ratings history of their holders over 2006-2009. The paper produces estimates of credit risk parameters and economic capital levels taking into account location factors and mortgages characteristics (fixed or adjustable interest rate, subsidized loans or loans contracted on the free mortgage market). Results show that the standard single factor model fails in capturing potential risk concentration. Concentration risk occurs in adjustable rate mortgages portfolio's segment, but local environmental factors help to reduce potential concentration risk in fixed rate mortgages portfolios segments as well as in subsidized loans segments.

[1] Corresponding author: michel.dietsch@unistra.fr.
[2] Université de Strasbourg and BPCE
[3] Université de Strasbourg
[4] BPCE

1. Introduction

Mortgages defaults concentration is often considered as a major starting point of the 2007-2008 financial crisis. In retail banking, where most portfolios are granular, credit risk concentration refers to a potential situation of correlated defaults in given portfolio's segments. Such correlations come from exposures' dependency to common macroeconomic factors but also from dependency to other systematic factors which characterize specific segments of the portfolio or particular clienteles. For example, mortgages borrowers located in a given geographic area can suffer from local economic recession or bad state of housing market. Thus, it is necessary to identify sources of borrowers' heterogeneity. Geographic differences could be one of these sources. But, in mortgages portfolios, as demonstrated by the subprime crisis, other sources of heterogeneity could be linked to the characteristics of the mortgages loans themselves. One is the choice between adjustable and fixed interest rate, and another one is the subsidized versus not subsidized – i.e. free - nature of the mortgage loans. In most European countries, less wealthy borrowers have access to public supported home ownership program and most European governments are strongly pushing such programs. Now, recent research tend to demonstrate that, in the U.S., the increase in the supply of finance to low-income borrowers drove lending and subsequent mortgage defaults in the recent subprime crisis (Mian and Sufi, 2008). But originators selling mortgages were also a main cause of the U.S. mortgage default crisis. So, it is interesting to try to measure the relative level of credit risk in populations of low-income borrowers benefiting from public program and in populations of borrowers using the free mortgage market.

One way to take into account this heterogeneity is to expand the asymptotic single risk factor (ASRF) model to consider additional risk factors (Lucas et al., 2001, Tasche, 2006). Then, the main problem is to identify the risk factors leading to borrowers' default in retail banking portfolios. In fact, a general theory of household's failure is missing. A natural way in searching for explicit risk factors would be to make explicit the latent factor in the Merton-Vasicek framework by introducing a set of macroeconomic variables. However, retail banking markets are local by nature and data availability at the local level may limit the implementation of such an approach. In addition, lack of time series data on potential risk factors at the local level (for example, regional economic activity or regional prices level data), which might be the relevant data for retail clienteles, impedes using the same methodologies which are commonly used to compute dependency structure parameters in large corporate exposures portfolios. Thus, finding a good methodology to compute risk parameters in retail loans portfolios is a real challenge.

This paper proposes an alternative way to expand the asymptotic single risk factor framework. The approach consists in expanding the one-factor model by adding new latent factors that can be linked to observable characteristics of the borrowers, such as location, wealth or profession, or to observable characteristics of their loans, such as the type of interest rate (fixed or adjustable) or the nature of mortgage loans (subsidized or not). These factors are expected to reinforce or attenuate the effect of general economic conditions on portfolio losses. Thus, the extension of the single-factor model to a multi-factor model could improve substantially the computation of the dependency structure across exposures in a typical mortgages' portfolio.

This paper uses extensive information related to a large French mortgages portfolio, including borrowers' ratings history. We apply the methodology of generalized linear mixed models (GLMM) to produce estimates of portfolio's credit risk parameters in a multi-factor context. This model implements in a coherent way the latent factor default model. It produces estimates of default thresholds considered as fixed effects and covariance matrixes of a set of latent random effects corresponding to the set of systematic factors. The estimation of such parameters allows computing economic capital as buffer of losses in portfolios exposed to different systematic risk factors.

Measurement of concentration risk calls for the computation of marginal contributions of distinguished segments to the total portfolio's losses, in order to allocate economic capital to each segment. As mentioned before, portfolio's segmentation is the result of a process that consists to identify separate groups of borrowers with the same observable characteristics which expose these borrowers to the same risk factors. In a multi-factor context, capital allocation can be implemented at the segment level such that it is possible to investigate the heterogeneity in capital allocation induced by the various systematic risk factors. We will show later that the differences between capital allocations coming from the choice of different risk factors can be significant. Thus, a single factor homogeneous framework could induce a misrepresentation of the concentration risk even in large portfolios of retail exposures. This could also lead to biased risk adjusted performance measures in the capital allocation process to business lines or decisions units. Moreover, while they are calibrated using a single factor model, Basel 2 regulatory formulas of capital requirements in the IRB approach could be of limited interest in allocating capital.

In this paper, we use the risk parameters estimates given by the multi-factor econometric model to compute the level of loans losses associated to distinct segments of the mortgages portfolio. Computation of portfolio's value-at-risk (VaR) and marginal contributions of segments exposed to additional risk factors are made by using a methodology recently proposed by Tasche (2009). This methodology has the advantage to take into account the impact of borrowers' heterogeneity on economic capital charges and capital allocation.

Section 2 presents the asymptotic multifactor credit risk model, its economic specification as a general linear mixed model (GLMM) and the chosen dependence structure of risk factors. Section 3 presents portfolio credit risk parameters results and show the consequences of introducing additional risk factors on the economic capital requirements. Section 4 presents the methodology we use for the computation of marginal contributions and analyses the concentration risk through a characterization of the heterogeneity in capital charges. Section 5 concludes.

2. Credit Risk Model specification

2.1 The asymptotic multi-factor credit risk model

Measures of portfolio credit risk require estimates of dependency across assets. The credit risk literature has emphasized the importance of obligors' exposure to common factors in determining the shape of the distribution of losses. Losses at the portfolio level can be defined as the sum of individual losses on defaulting expositions, given the individual severity of losses, i.e. loss given default. Hence, defining u_i as obligor's i loss given default (LGD) and $\mathbf{1}_{Di}$ the default indicator variable of obligor i:

$$L = \sum_{i=1}^{n} u_i \mathbf{1}_{D_i}$$

In structural credit risk models, default occurs when the value of one obligor's assets become smaller than the value of due debt. As asset values may be difficult to observe, this framework has been extended by generalizing the modelling of default as the crossing of an unobservable threshold. Thus, the financial health (or final asset value) of obligor *i* is represented by a latent (unobservable) variable U_i, which level is determined by the realizations of a set of risk factors such that:

$$U_i = \mathbf{w'_i}\, s + \sqrt{1 - \mathbf{w'_i}\, R\, \mathbf{w_i}}\ \varepsilon_i \quad (1)$$

where S is a vector of "systematic" risk factors with realization s, $\mathbf{w_i}$ is the vector of sensitivities (or factor loadings) of the *i*-th borrower to the set of systematic factors and ε_i is a specific risk factor for borrower *i*. R is the correlation matrix of the risk factors. Assuming risk factors are multivariate Gaussian, the sensitivity to specific risk in equation (1) ensures that U_i is standard normal. Specific risk factors are assumed to be uncorrelated among obligors and also independent from the systematic factors. In this framework, default occurs when the latent variable U_i falls below a default threshold calibrated according to the stationary (long term) default probability $\overline{p}_i$ of obligor *i*. Denoting Φ the standard normal cdf, default occurs when:

$$1_{D_i} = 1 \Leftrightarrow U_i = \mathbf{w_i}' s + \sqrt{1 - \mathbf{w_i}' R\, \mathbf{w_i}}\ \varepsilon_i < \Phi^{-1}\left(\overline{p}_i\right)$$

Moreover, assuming specific risk can be entirely diversified away, the default indicator variable of a given obligor can be approximated by its expected value, that is by its default probability conditional to the realization of systematic risk factors. Conditional portfolio losses are then defined as:

$$L(s) \approx \sum_{i=1}^{n} u_i \Phi\left[\frac{\Phi^{-1}\left(\overline{p}_i\right) - \mathbf{w_i}' s}{\sqrt{1 - \mathbf{w_i}' R\, \mathbf{w_i}}}\right] \quad (2)$$

This framework is called the asymptotic multi-factor framework of credit risk (e.g. Lucas et al., 2001). Equation (2) assumes that each obligor can be characterised by individual default threshold and factor sensitivities. However, in retail loans portfolios, default rates are computed by rating classes and sensitivities to risk factors cannot be computed on an individual basis. Thus, assumptions are required in order to reduce the number of parameters of the loss variable. Following previous literature on large portfolios of loans, we assume that obligors belonging to the same rating notch r will share the same default threshold.[5] We further assume that the vector of risk factor sensitivities is the same for obligors belonging to the same segment of a portfolio. Hence, assuming a portfolio composed of K segments, losses can be rewritten as:

$$L(s) \approx \sum_{i=1}^{n} u_i \Phi\left[\frac{\Phi^{-1}(\overline{p}_r) - \mathbf{w_{ki}}' s}{\sqrt{1 - \mathbf{w_{ki}}' R\, \mathbf{w_{ki}}}}\right]$$

Now, considering a multi-factor framework calls for the specification of the dependence structure of risk factors as well as for the estimation of the default thresholds and factor sensitivities.

2.2 Estimation of credit risk parameters using Generalized Linear Mixed Models.

The methodological option followed here is to extend the single factor model by adding new latent factors that can be linked to easily observable characteristics of businesses. Many

[5] Recent results have highlighted the importance of firm level heterogeneity in the measurement of credit risk. Pesaran, Schuermann, Treutler (2005) illustrate the importance of the dispersion of default probabilities within ratings on the level of credit risk. Duellmann, Scheicher and Schmieder (2007) show the prominence of name concentration over sector concentration. These results are obtained considering larger (mostly) quoted companies for which equity price changes can be related to explicit risk factors. We assume here that the size of our portfolio is sufficiently large to mitigate these problems. Nevertheless, it could remain in our portfolio a share of firm level heterogeneity which is not accounted for.

researches have underlined the importance of industry effects when considering credit risk (Carling et al., 2004, Pesaran et al., 2005, Heitfield et al., 2006). Here, we will also concentrate on industry effects in determining the level of losses.. In order to estimate the impact of additional risk factors on portfolio credit risk, we use an econometric model belonging to the class of generalized linear mixed models. GLMM models combine fixed and random effects for observable and (latent) unobservable factors. Detailed presentations of GLMM models in credit risk modelling can be found in Frey and McNeil (2003) and McNeil and Wendin (2007).

Defining Y the $(N \times 1)$ vector of observed default data and γ the $(K \times 1)$ vector of random effects, the conditional expected default probability of obligor i is given by:

$$E\left[Y_i = 1 \middle| \gamma\right] = g\left(X\beta + Z\gamma\right)$$

where $g(\cdot)$ is a differentiable monotonic link function and Y_i the default indicator variable for obligor i. In the following applications, we will focus on the probit link function because the normal distribution is also the underlying link function assumed by the Basel 2 framework of credit risk, thus $g(x) = \Phi(x)$. X is a $(N{\times}P)$ matrix containing the (observed) fixed effects, and Z is the $(N{\times}K)$ design matrix for the random effects. The random effects are assumed to be normally distributed with mean 0 and covariance matrix G. β is the vector of parameters associated to the fixed effects.

Considering a portfolio of N obligors dispatched in $r = 1, \ldots, R$ (non-default) rating classes and given a vector γ_t of random effects, the default probability of borrower i at time t is given by:

$$P\left(Y_{ti} = 1 \middle| \gamma_t\right) = \Phi\left(x'_{ti}\,\mu_r + z'\gamma_t\right)$$

Where:

- μ_r denotes the vector of parameters of the fixed effect of rating. If the rating scale is properly built, we expect these thresholds to be ordered and increasing as credit quality decreases.
- $x'_{ti} = [0,...,1,...0]$ is a $(1\times R)$ vector defining the rating of borrower *i* in time period *t*. As we assume exchangeability of borrowers within segments, the estimation is not done on individual borrowers, but on annual default rates within segments. This leads assuming homogeneity with respect to credit rating.
- *z* is the design matrix of random effects. Its form depends on the assumed dependence structure of random effects, as developed in the following section.

Here, we consider one fixed effect: the borrower's credit quality – measured by its rating - defining default thresholds - and two different types of random effects: industry effects (assuming a firm can be assigned to a unique sector) in addition to a general latent risk factor. Thus, the additional random effects are introduced through an industry segmentation of rating and default histories. Given conditional independence, defaults follow a binomial distribution. Hence, the joint conditional distribution of $\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tn})'$ is:

$$P(Y_t = \mathbf{y}_t | \gamma_t) = \prod_{i=1}^{n_t} P(Y_{ti} = 1 | \gamma_t)^{y_i} (1 - P(Y_{ti} = 1 | \gamma_t))^{1-y_i} \text{ for all } y_i \in \{1,0\}^{n_t}$$

Further assuming that random effects are serially independent (but possibly cross-sectionally correlated in the case of multiple random effects), the unconditional density of $\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tn})'$ is, defining θ as the hyper-parameter of the random effects distribution and *Q* the number of random effects:

$$f(\mathbf{y}_t | \beta, \theta) = \int \cdots \int_{Q^r} \left(\prod_{i=1}^{n_t} P(Y_t = \mathbf{y}_t | \gamma_t, \beta) \right) g(\gamma_t | \theta) d\gamma_t$$

The likelihood function with serially independent random effects is then:

$$L(\beta,\theta|\, data) = \prod_{t=1}^{T} f(\mathbf{y}_t|\beta,\theta)$$

However, the introduction of a large number of, possibly correlated, random effects makes the estimation of the model's parameters trickier. Methods based on the approximation of the objective function (like Laplace or Gaussian quadrature) can be implemented as long as the number of random effects remains low and their dependence structure simple. Consequently, the model is estimated by restricted pseudo-likelihood, which is based on the linearization of the likelihood function (Wolfinger and O'Connell, 1993) where the optimization is based on some linear expansion of the objective function. The drawback of this approach is the absence of a true objective function for the overall optimization process, which can lead to biased estimates of covariance parameters.[6] Indeed, simulation studies show that restricted pseudo-likelihood estimation of GLMM models of binary outcome may lead to large biases in variance components (*G* matrix) parameters as well as in their standard errors (see e.g. Breslow and Lin, 1995, Lin and Breslow, 1996). However, these studies also show that grouping observations in homogeneous clusters substantially reduce these biases (Lin, 1997). Thus, grouping binary (default) observations into binomial responses with high numbers of individuals as done here at the industry/rating level limits the impact of these biases.

2.3 Dependence structure of random effects.

Extending the single factor model calls also for a specification of the risk factors' dependence structure. By assuming that the general risk factor (the risk factor of the single factor model) represents the impact on default rates of variations in general economic conditions, it seems straightforward to consider that additional risk factors can reinforce or weaken the sensitivity of a given subset of borrowers in the portfolio to general economic conditions. This corresponds to the idea that a geographic location factor, for example, can be either

[6] We estimate the models using the SAS 9.1 Glimmix procedure.

procyclical, cycle neutral or countercyclical. In order to capture these effects, we will estimate the correlation between the general risk factor and additional factors associated to a given segmentation of the portfolio. In order to keep the model tractable, we further assume that the additional factors, i.e. shocks that affect segments of the portfolio, are independent. This specification implies in particular that the inter-segment correlation is not directly attributable to the risk factors of segments but rather to the impact of general economic conditions and the dependence between this latter factor and the other risk factors. The covariance structure we will focus on is of the form:

$$G = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \sigma_{1,q+1} \\ 0 & \ddots & 0 & \sigma_{2,q+1} \\ 0 & 0 & \sigma_q^2 & \vdots \\ \sigma_{1,q+1} & \sigma_{2,q+1} & \dots & \sigma_{q+1}^2 \end{bmatrix}$$

considering q latent segment factors and one systematic factor (denoted q+1). This structure implies that a given borrower is affected by two (or more) factors: the factor representative of general economic conditions and additional factors attached to the region or to another risk class he belongs to. In our empirical analysis, we will consider separately or all together additional random factors such as the borrower's location, the nature of interest rate attached to the loan, or the type of loans. The last column of G defines a random effect common to all obligors and reflects the heterogeneity in default rates related to time. This random effect corresponds to the heterogeneity in default rates attributable to time-heterogeneity, which is assumed to be related to general economic conditions. In this specification, the linear predictor in the logistic regression contains an intercept term that randomly varies at the year level, the highest level in the modelling, where all other effects are nested in. In other words, a random intercept is drawn separately and independently for each year.

3. Risk parameters estimates and computation of economic capital.

3.1 Data

We use a large database of French mortgages and historical default data, which contains 343.423 credit files including 527.041 mortgages (the same borrower could get two different loans). This database comprises a large proportion of loans took out in the framework of public social program. The database covers four years (2006 to 2009) including the recent economic crisis years. The unique objet of the loans in this study is to finance principal home ownership. This database provides monthly information about the type of loan interest rate, the kid of loan (took out in the free market or in the framework of a public home ownership financing support program), and the borrower's geographic location. The database provides also the complete history of borrowers' credit quality as measured by an internal ratings system, in which borrowers are ranked from rating grade 1 (low risk) to 5 (high risk) and rating 6 corresponds to bank default. Some details about of portfolio's characteristics are presented in table 1.

The portfolio contains mortgages with two different interest rates: fixed rate mortgage (FRM) and adjustable rate mortgage (ARM). Notice that, in France, borrowers can benefit from public subsidized mortgages, and also from other public supported programs, which allow reducing interest charges, as reward for previous savings effort.

Table 1 Main characteristics of the French mortgages portfolio used in the present study (averages values over 2006-2009)

	Number of borrowers (in %)	Average EAD in euros (end of period)
Type of principal loan's interest rate		
FRMs	53,4	43 718
ARMs	46,6	104 847
Type of principal loan		
Borrowers using only public supported loans	42,70	63 825
Borrowers using only market loans	39,7	64 155
Borrowers using both public supported and non public supported loans	17,6	122 564

The loan's size is sometimes considered as a main element of household choice and a main factor of risk management. Table 1 shows that average loans' amount varies deeply with loans characteristics. Due to regulatory limits upon the total amount each borrower could raise, the average amount of zero interest rate mortgages shows the lowest average observed value in the database. Borrowers using only subsidised loans are less indebted than borrowers which contracted in the free market. Finally, borrowers using ARMs or borrowers mixing subsidised and non subsidised loans are the most indebted in this study's portfolio.

3.2 The choice of additional risk factors

We consider two types of additional factors which create potential heterogeneity in borrowers' populations: those linked to the mortgages' characteristics and the geographical location factors.

The mortgages' characteristics

Two source of borrowers' heterogeneity can be linked to mortgages characteristics. First, one can expect that risk factors affecting mortgages benefiting from a public financial support might be different from those that do not benefit from such help and borrow without support. In fact, the main effect of the public support is to lower the interest charges over the borrowers' life cycle, what render financial constraints less binding and may contribute to lower default rates. Second, heterogeneity could come from the choice between fixed rate mortgage (FRM) and adjustable rate mortgage (ARM). There are two strands of literature considering this issue. The first strand emphasized the relevance of mortgage characteristics in solving the adverse selection problem between lender and borrower. The second one considers the choice of a mortgage contract as a problem in household risk management.

Considering the first line of arguments, mortgage characteristics (maturity, loan-to value, payment schedules, interest rate) could be used by lenders as mechanisms to screen the unobservable default risk of borrowers (Ben-Shahar, 2005). Here, we focus on the type of interest rate in order to highlight potential differences in the credit risk parameters. Posey and Yavas (2001) focus in a theoretical model on the self-selection of borrowers between fixed rate mortgages and adjustable rate mortgages. Although they obtain mixed results depending on the magnitude of (borrower) costs of default, their main result is that there exist separating equilibriums in which lenders offer both ARM and FRM, high risk borrowers choosing ARMs. The intuition of the result in Posey and Yavas (2001) relies on the combination of the (random) interest payments and the random income stream of borrowers. Since high-risk borrowers are more likely to have low income, they valuate the possibility to support low interest payments, while low risk borrowers are less likely to default because of high interest charges. Thus, high-risk borrowers have a preference for ARMs. Therefore, one expects adjustable rate borrowers to exhibit ex post higher risk characteristics which, in our framework, are captured by the variance of latent factors. Latent factor variance indicates the scope of potential concentration of defaults when economic conditions worsen. A limitation of this intuition is that the data do not contain any information concerning the causes of default; in particular we cannot observe if defaults are attributable to increases in interest rates or to other risk factors directly attached to borrowers, like unemployment, or divorce.

We consider now the second line of arguments in which mortgage's choice is determined by household's risk management (Campbell, 2006, Campbell and Cocco, 2003, Koijen et ali., 2009). When deciding on the type of mortgage, an important consideration is labor income and the risk of its volatility. FRMs expose household to a wealth risk, because the real capital is affected by inflation. On the contrary, an ARM is a safe contract in the sense that its real value is stable. But ARMs expose households to income risk, i.e. to short-term volatility in real payments each month. This risk is problematic if the borrower faces binding financial constraints and cannot borrow against future income. So, low income borrowers are vulnerable to this risk, because buffer savings might be exhausted in states of the world with low income. Now, borrowers are also exposed to the risk of variation in real interest rates. FRMs protect against this risk while ARMs do not. Finally, if the borrower is borrowing constrained, the most appropriate mortgage contract is the one with lowest interest rate payments, which is generally the ARM contract. However, borrowers are also sensitive to the yield spread between ARMs and FRMs which is driven by the long-term – short-term yield spread. If the yield spread is low, borrowers will prefer with certainty FRM, due to the low level of the FRM premium.

However, ARMs' borrowers, while selecting such type of contract implicitly accept the additional default risk stemming from potentially increasing interest rates. Thus, the higher ex post risk could partly be attributed to this higher risk-taking behavior. Indeed, our sample shows that the average quarterly default rate is higher in the population of borrowers choosing ARMs. In the sample, the average default rate in this population is twice the default rate of borrowers choosing FRMs.

A peculiarity of the French mortgages market is to propose various kinds of subsidized loans. Borrowers could access these public supported programs if they fulfil income conditions. This peculiarity is exemplified in our database. Likely because subsidies tend to reduce budget constraints, borrowers benefiting from such public supported loans are characterised by significantly low default rates, on average, as compared to default rate of borrowers which go to the non subsidised mortgage market. However, borrowers using a majority of subsidised ARMs' loans show on average the same default rate than other borrowers choosing ARMs in the free market.

The geographic location

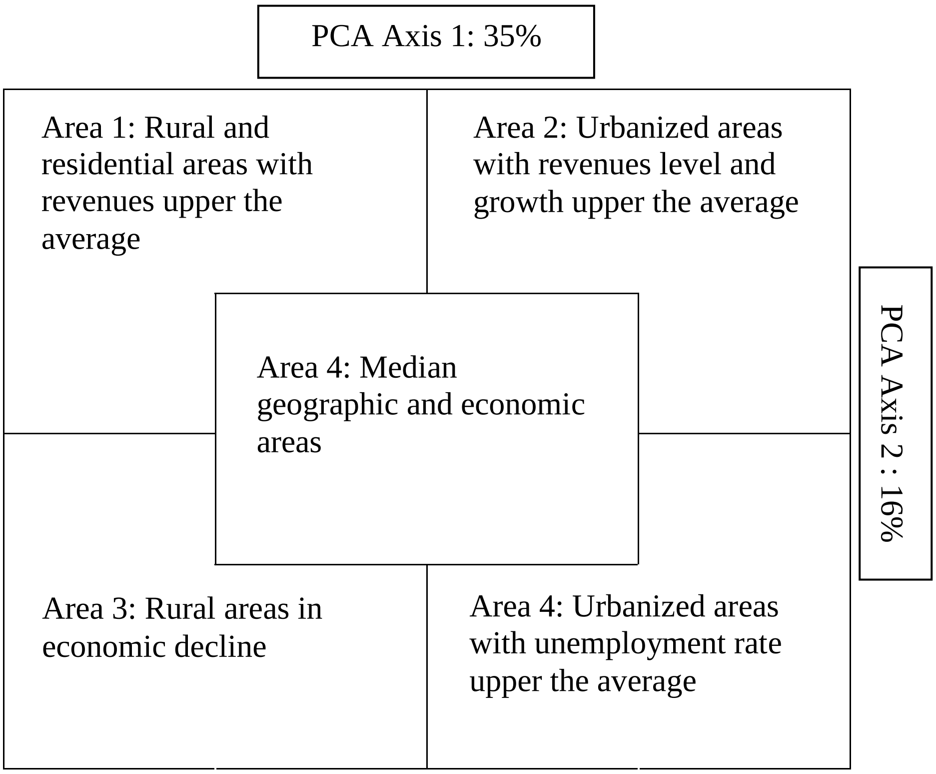
Another important risk characteristic may be geography, i.e. characteristics of the environments – in terms of demographic and economic growth - where borrowers are living. In fact, local socio-economic features (wage levels, types of jobs, living costs), lead to a high degree of heterogeneity across housing markets and consequently across credit risk on these markets. Indeed, higher price levels in highly urbanized areas might only partially being compensated by higher wages associated to more qualified jobs or sector specializations. For instance, Calem & LaCour-Little (2004) have highlighted, in the American context, the importance of local diversification in the credit risk of mortgage portfolios and the calculation of corresponding capital requirements. Others have focused on Europe considered as a single market (Diaz-Serrano, 2005). Here, we will go one step further and focus on the sensitivity of borrowers' credit quality to general economic conditions and the existence of local risk factors. This gives an insight in the geographic repartition of credit risk on these markets as well as a characterization of potential diversification at the country level. More precisely, we are interested here in highlighting credit risk heterogeneity between various geographic environments.

To measure geographic factors synthetically, we have grouped obligors in their employment areas (*bassins d'emploi* in French) which are defined as areas in which most workers work *and* live and most firms find labour force to hold their supplied jobs. We characterise these areas – which amount to 366 in France - in a multidimensional way by computing average values of several socio-economic indicators within each area. We used twelve socio-economic indicators[7]: unemployment rate, population growth, average taxable income, population density, share of senior and retired people in total population, share of owners of their principal housing in the population, share of senior executive managers and high educated people in the active population, share of social housing in total housing, share of vacant housing in total housing, share of second residence in total housing, proportion of households owning two cars or more, and finally a business sector concentration Herfindhal-Hirchmann index in the area.

Then, we have applied a Principal Component Analysis (PCA) on this database of French employment areas. Results of this analysis (not presented here) show that the two first factors explain around 52% of the total variance. The first factor (35% of the variance) is mainly determined by geographic variables and by the ageing structure of the population. In brief, it

[7] Information about all these variables at the town level is provided by the French National Institute of Statistics (INSEE).

distinguishes rural and urban areas. The second factor (16% of the variance) is determined by revenue and wealth conditions as well as by unemployment rate and the share of senior executives. In the final step of the process, we used the quartiles of the coordinates of the employment areas on the two first component axes to distinguish five classes of geographic areas and group borrowers living in the same town or sub-area in these classes. The definition of geographic areas follows the logic behind the following figure.



The first environment is characterised by a low urbanization rate and good economic conditions, in particular in terms of revenues and life conditions. The second environment is also characterized by good economic conditions but differs from the previous one because of its high urbanization rate and high demographic growth. The third environment corresponds to declining rural zones with a low demographic growth and a high proportion of retired people and seniors. The fourth environment is characterised by high level of urbanisation and also an unemployment rate upper to the national average. Finally, the fifth environment gathers mean-sized towns and median employment areas, that is employment areas which appear in the core of the principal component analysis results.

Knowing the borrower's mailing address, we ranked each borrower in the environment to which its employment area belongs. Thus, each borrower is characterised by its own geographic factor – its environment – and we added five distinct geographical factors in the modelling of mortgages' portfolios credit risk. Environments are likely to exert an impact on borrowers' mortgages choices and financial health, as shown by Table 2 below.

Table 2 Main characteristics of the French mortgages portfolio used in the present study by geographic area (averages values over the 2006-2009 period)

Geographic areas	Number of borrowers %	Average quarterly default rate	Average EAD in euros (end of period)
Area 1 (rural/residential – growing)	8,7	0,71%	59 535
Area 2 (urban – wealthy and growing)	35,3	0,76%	82 586
Area 3 (rural – ageing)	6,0	0,77%	55 951
Area 4 (urban – high unemployment)	32,3	0,85%	79 855
Area 5 (median areas)	17,7	0,77%	62 696

3.3 Estimation results

We proceeded in fourth steps to progressively introduce additional risk factors in different specifications of the GLMM binomial model presented in the preceding section. In a first step, we estimated a single factor "benchmark" model where all borrowers are assumed homogeneous in the segment to which they belong in terms of default thresholds and sensitivity to general systematic risk. In this specification, all borrowers within the same segment share the same sensitivity (asset correlation) to a general systematic risk factor. Then, in a second step, we introduced the loans characteristics as distinct risk factors. Here, the objective is to compare risk sensitivities of different categories of borrowers. Firstly, we estimated a type of interest model that extends the benchmark model by introducing main types of interest rate (fixed or adjustable rates) as additional random effects. Secondly, we estimated another model that considers categories of borrowers choosing different types of loans: public supported loans only, free market loans only, and a mix of the two types of loans. However, estimation results showed that this latter model only slightly discriminated borrowers. In fact, the role of the loans' type risk factor appears more distinctly when the model combines interest rate and loan type risk factors. In that case, we consider six distinct combinations of the loan's type and the rate's nature as additional random effects. This model aims at capturing the risk sensitivities of different populations of borrowers who actually solve budget constraints by choosing specific characteristics of their loans. In a third step, we estimated a "geographic" model that considers borrowers' geographic areas as random factors by using the geographic segmentation of borrowers presented above. Thus, the matrix of random effects is extended to five geographic risk factors in addition to the general random factor. And finally, in order to evaluate the relative impact of the loans' characteristics factors and the environmental factors, we estimated a model combining geographic areas and type of rates in two distinct sub-portfolios: a portfolio containing only subsidized mortgages and another one only free market mortgages.

Estimation details - default thresholds and covariance matrixes of random effects associated to the various models - are shown in the appendixes. Goodness-of-fit statistics of the GLMM models (in particular the generalized chi-square to degrees of freedom ratio) indicate that the single factor model has a poor fit as the residual heterogeneity remains high. Results show the model which combines interest rate and type of loans has a better fit than models that consider each loans' characteristic separately. The geographic model and the models combining geographic factors and loans' characteristics show also good statistical performances, what means that the addition of these factors is valuable[8].

Table 3: Covariance matrix of random effects

	Types of rate		
	Fixed rate	Adjustable rate	General
Fixed rate	0.000283		-0.00097
Adjustable rate		0.04132	0.01657
General	-.000970	0.01657	0.01481

	Fixed rate * subsidized loans	Fixed rate * market loans	Fixed rate * both subsidized and market loans	Adjustable rate * subsidized loans	Adjustable rate * market loans	Adjustable rate * both subsidized and market loans	General
Fixed rate * subsidized loans	0.008906						-0.008880
Fixed rate * market loans		0.05763					-0.044370
Fixed rate * both subsidized and market loans			0.017490				0.003955
Adjustable rate * subsidized loans				0.018480			0.009253
Adjustable rate * market loans					0		0.000000
Adjustable rate * both subsidized and market loans						0.01678	-0.005750
General	-.008880	-0.04437	0.003955	0.009253	0	-0.00575	0.050500

	Area 1	Area 2	Area 3	Area 4	Area 5	General
Area 1	0.000577					0.002314
Area 2		0.002081				-0.002010
Area 3			0.001823			-0.001430
Area 4				0		0.000000
Area 5					0	0.000000
General	0.002314	-0.002010	-0.001430	0	0	0.012340

[8] In addition, we tested for additional risk factors effects. A first test rejected the null hypothesis that each of the additional factors separately is zero, for most factors considered in the various models. A second test rejected the hypothesis that the off-diagonal elements in the matrix are equal to zero. Thus, the integration of the elements of interaction between the additional factors linked to the characteristics or to the geographic factors, on the one side, and the general factor, on the other, brings generally additional relevant information. See results of the tests in appendix.

Observation of matrixes of random effects (table 3) highlights the divergence between borrowers choosing adjustable rate mortgages (ARMs) and those choosing fixed rate mortgages (FRMs). Thus, the loan's interest rate exerts a significant effect on default rate. Indeed, covariance matrixes show that assets correlation within a segment is higher for borrowers taking out ARMs, taking account for values of the covariance between the general macroeconomic factor and the segment specific factor. Interestingly, the model mixing the type of loan and the type of interest rate variables highlights diversification effects within mortgages' portfolios. Intra-segment covariances tend to increase when the type of loan factor is added to the fixed rate factor, and to decrease when it is added to the adjustable rate factor. But most covariance between the general factor and the segment specific factor are negative in the FRMs segment, while they are low or only slightly negative in the ARMs segment. That means that the combination of the general and specific risk factors contribute to reduce credit risk in the portfolio's segments composed of FRMs, and to slightly increase it in the ARMs segment. Finally, even if the value of the intra-segment covariance is zero in the segment of ARMs contracted in the free market, the credit risk is higher in this segment than in the segment of FRMs contracted in the free market, due to the risk counteracting effect (negative covariance) of the local factor in the latter segment. Following that logic, results show that credit risk is at its highest level in the segments of ARMs subsidized mortgages. By comparison, portfolios composed of subsidized FRMs appear less risky.

Finally, the geographic factor model shows only weak differences between environments. The intra-segment correlation appears to be higher in area 2, which corresponds to urban employment areas in good economic state, but this effect on credit risk is completely mitigated by the negative covariance between general and geographic factors. The same is true in area 3.

Finally, credit risk seems to be higher in the three other areas, even if the covariance values associated to the geographic factors are zero or lower than in the two others. But, overall, the impact of general systematic risk factor is rather small in this geographic model.

However, the models mixing environmental factors and interest rate factors verify the existence of a geographic effect (see results in appendix 2.D and 2.E). Indeed, in the two populations of borrowers –using only subsidized loans or using only market loans – credit risk is distinctively higher in ARMs segments, because, in this segment, the effect of the local geographic factor is equal to zero or small and the interaction of the local factor with the general – i.e. national – factor is very weak. Thus, for borrowers choosing ARMs, the general risk factor clearly dominates the local risk factor and the covariance of the general factor is higher in the segment of market mortgages than in the segment of subsidized ones (0.07 versus 0.01 : see tables D in appendix 2.D and 2.E). For borrowers choosing fixed rates, the reverse is true, whatever the type of loans, subsidized or not. Indeed, in the two portfolios, and in the five segments crossing fixed rates and environments, covariances within segments diverge across environments and, in most cases, covariances between the general (national) factor and the geographic (local) factor are negative. This latter result means that local factors frequently counteract the effect of the national factor and contribute to reduce the sensitivities of fixed rates exposures to risk factors[9]. To resume, geographic factors appear to play a significant role to reduce borrowers' risk sensibilities in most segments of the population of borrowers choosing FRMs, while in the population of borrowers choosing ARMs, local factors do not play a counteracting role, and the impact of general (national) risk factors distinctively prevails. The final consequence is that global sensitivity to risk factors is higher in the ARMs segments.

[9] However, in the population of borrowers choosing market loans, in areas 1 and 5, the local and the national factors interact positively. We will see below that this interaction provokes a credit risk concentration effect that increases potential losses in these two segments.

As mentioned before, because covariances between additional risk factors and general macroeconomic factor could be negative, the intra-segment covariances results do not presume final effects on capital ratios. Now, it is useful to consider the impact of the risk parameters on the economic capital consumption. To this aim, economic capital is computed by simulation of the risk factors given default thresholds and risk factor sensitivities, which are the outputs of the GLMM models[10]. Here, exposures at default are computed as the average value of mortgages in each bucket built by taking rating and interest rate type, rating and type of mortgages class, and so on.

Table 4 shows economic capital (% of total exposures) for various quantiles of the distribution of losses. Quite significant differences can be observed between models when considering the level of economic capital. Taking into account some additional risk factors could induce a significant increase of economic capital requirements, as compared to the requirements we found when taking account for a single factor. In addition, table 4 shows that values of capital ratios are very low. Many factors explain this result: high proportion of very safe borrowers in the population, quite short maturity of mortgages (around 15 years), low individual loan amount and very low LGD (high rate of recovery due to efficient guarantees). Yet, most quarters of the period (2006 to 2009) under review were characterized by economic recession which may have led to higher default rates.

[10] At this stage, a technical difficulty arises as estimated covariance matrices are rarely definite positive. Yet, definite positivity being a necessary condition for correctly specified correlation matrices, we correct the estimated matrices by applying the correction suggested by Higham (2002).

Table 4: Economic capital quarterly ratio (% of total exposure)

Quantile:	90%	95%	99%	99.5%	99.9%	99.97%
One factor model	0.277	0.294	0.329	0.342	0.371	0.390
Type of interest rate as random effect	0.452	0.514	0.647	0.703	0.840	0.929
Combination of interest rate and loan types model as random effects	0.434	0.483	0.589	0.635	0.726	0.798
Geographic area environment as random effect	0.276	0.293	0.326	0.339	0.367	0.385

Note: Table shows the value-at-risk expressed as percentage of total portfolio exposure. All computations assume a 20% LGD.

4. Capital allocation, borrowers' heterogeneity and credit risk concentration.

In this section, we assess the ability of each model to detect potential concentration. If groups of obligors are homogenous in terms of credit risk, capital ratios should not differ across these groups. On the contrary, if there is obligors' heterogeneity, capital ratios should differ and a substantially higher capital ratio would indicate potential risk concentration in a given sub-portfolio. A straightforward source of heterogeneity comes from the differences of credit ratings which are accounted for in the GLMM estimation of default thresholds. Moreover, in mortgages portfolios, geographic location and loan contracts characteristics could determine additional sources of risk heterogeneity. Differences in capital ratios along these sources of risk should enlighten a potential for concentration risk. Thus, a portfolio's segment could need a relatively higher economic capital requirement than another one, what would indicate a relative concentration of economic capital in that segment. The mortgages portfolio under consideration is highly granular due to its size. Thus, concentration risk will mainly come from additional risk factors which determine the portfolio's segmentation.

4.1 Computing marginal contributions by Monte Carlo simulation

The detection of potential concentration requires allocating portfolio-wide economic capital to sub-portfolios or individual assets. It is the (relative) importance of economic capital "consumed" by given sub-portfolios that highlights the potential concentration of losses on these sub-portfolios and should lead to restructuration and investment decisions. Grounding on the works of Tasche (1999) and Gouriéroux and al. (2000), marginal contributions to portfolio VaR can be expressed as the expected loss on a given exposure conditional on losses reaching VaR:

$$RCVAR_i = E\left[L_i \middle| L = VaR_\alpha(L)\right] = \frac{E\left[L_i \mathbf{1}_{VaR_\alpha(L)=L}\right]}{P\left[L = VaR_\alpha(L)\right]}$$

This last definition shows that, given a positive probability for losses reaching VaR, the computation of marginal contributions relies heavily on the ability to estimate individual losses when aggregate losses approach VaR. Thus, when considering Monte Carlo simulation, the conditional mean might be based on a limited number of simulations, leading to unreliable estimates. In order to improve estimation, some authors (Tasche, 2009, Glasserman and Li, 2005, Egloff et al., 2005) have used importance sampling. Importance sampling consists in shifting the distribution's parameters in order to increase the likelihood of observing some desired realizations of the risk factors. Here, the main difficulty in importance sampling lies in the choice of the alternative distribution, i.e. in the determination of the parameters of F^*. We follow Tasche (2009) in shifting only the risk factors (S) means such as:

$$S_i^* = S_i - E_F\left[S_i\right] + \mu_i,\ \mu_i = E\left[S_i \middle| L = VaR_\alpha(L)\right]$$

The next step is the computation of conditional expectation (equation 3). As the computation of VaR is done by Monte Carlo simulation, one has as well the realizations of each risk factor as the resulting credit losses. This allows using the non parametric Naradaya-Watson estimator for conditional expectations. Using the standard normal density as kernel and denoting h the bandwidth of the kernel, the estimator of the conditional expectation for risk factor k is defined as:

$$\hat{E}\left[S_k \middle| L = VaR_\alpha(L)\right] = \frac{\sum_{t=1}^{T} S_k \phi\left(\frac{VaR - L_t}{h}\right)}{\sum_{t=1}^{T} \phi\left(\frac{VaR - L_t}{h}\right)}$$

$$h = 1.06\sigma_L T^{-1/5}$$

In order to limit the computational burden of the simulation of marginal contributions, we rely on the size of the portfolio under consideration to assume complete diversification of idiosyncratic risk. This allows to simulate losses using conditional probabilities rather than to simulate defaults and associated losses. Associated to the assumed homogeneity of exposures within sub-portfolios resulting from the crossing of a given segmentation variable and a ratings class, it is possible to compute a single marginal contribution by sub-portfolio rather than to proceed at the asset level. Losses are then approximated by:

$$L \approx \sum_{j=1}^{N} u_{rk} \Phi\left(\frac{\Phi^{-1}(\bar{p}_r) - \mathbf{w}' s_k}{\sqrt{1 - \mathbf{w}'\Sigma\mathbf{w}}}\right)$$

Once the shifts in the means are computed for all the risk factors, the next step is to draw realizations of the risk factors under the new distribution in order to compute again the aggregate losses as well as individual losses within each sub-segment and each rating grade. Tasche (2009, proposition 4.2) using Klebaner (2007) establishes that the expected losses conditional to VaR under the natural distribution can be defined as:

$$E_F\left[L_i \middle| L = VaR_\alpha(L)\right] = \frac{E_{F^*}\left[L_i R \middle| L = VaR_\alpha(L)\right]}{E_{F^*}\left[R \middle| L = VaR_\alpha(L)\right]}$$

As previously, these conditional expectations can be computed using the Naradaya-Watson estimator and the simulations of risk factors and losses can be realized under the shifted distribution.

4.2 Heterogeneity and risk concentration in mortgages portfolios

Here, to characterize the potential concentration risk, we consider the distinct mortgages portfolio's segments using mortgages characteristics and borrowers' location areas as criterions of segmentation. We computed economic capital ratios for the different segments (in % of total exposures). We also computed the Basel 2 IRBA regulatory capital charges, no longer considering estimated stationary default probabilities, but instead the empirical average default rates (BCBS, 2006, §328). Table 5 and graphs 1a to 1c gather the capital allocation results and presents them in terms of capital ratio by segment.

Due to heterogeneity of portfolio risk parameters and exposures, marginal contributions in term of capital charges vary significantly across segments. In fact, table 5 shows a decrease of economic capital charges in some portfolio segments, relative to the capital charge computed by using a single factor model, as well as increases in other segments. Thus, the single factor model fails in capturing potential risk concentration.

Moreover, results show that loan interest rate as a random factor has a strong impact on the capital consumption: the capital ratio is higher for ARMs than for FRMs. A reason could be that the borrowers choosing ARMs exhibit on average a lower credit quality – a higher risk level - than borrowers choosing FRMs. In fact, riskier borrowers could be oriented towards ARMs, this bank's policy serving - as mentioned above - as a device to control adverse selection problems, or they simply choose ARMs to reduce current interest burden. The higher consumption in ARMs segment is also explained by the higher concentration between borrowers in this segment. Another potential cause could come from the size of the exposure. Indeed, the average EAD of ARMs is very large – around twice the one of the FRMs.

Crossing interest rate type and loan type underlines the difference across sub-portfolios in term of capital consumption. The resulting capital ratio is the combination of the default rate's level, the average EAD and the degree of concentration (measured by covariance parameters) in each segment. We observe the highest economic capital consumption in the segment crossing adjustable rate and subsidized loans. Again, for low-income borrowers benefiting from public supported loans, choosing an adjustable rate mortgage loan helps to satisfy financial constraints which are likely more severe than in the other borrowers population.

Finally, the effect of the environment on concentration risk varies with the interest rate borrower's choice, whatever the type of mortgages, subsidized or not. When considering the role of geographic factors separately, we observe the lowest economic capital consumption in areas 2 and 3. But differences in economic capital consumption between geographic areas (rural or urban) are very weak (as illustrated in graph 1b). In fact, the role of the environment factor appears more clearly when we consider the impact of geographic factors in interaction with loans characteristics factors. First, in the segment of borrowers choosing FRMs, taking geographic risk factors into account leads generally to a decrease of economic capital consumption (graph 1c), as compared to the capital consumption computed by using risk parameters (covariances) given by the model combining type of loans and type of interest rate factors (graph 1a). This result comes from the counteracting effect (negative covariance) of the local factor, as mentioned before. Second, in the segments of borrowers choosing ARMs, the capital ratio is higher than in the segments of FRMs, whatever the mortgage is subsidized or not. The corresponding portfolio's segments do not benefit from the counteracting effect of local factors, so that the general factor and the loans characteristics factors reinforce each other, accruing capital ratio. In other words, ARMs segments are strongly exposed to the effect of the general macroeconomic factor and do not really benefit from diversification effects the local factors could bring. Third, the capital ratio is higher in the free market mortgages segment (ARMs and FRMs) than in the subsidized mortgages segment. The reason comes from the higher value of asset correlation in the first segment (shown by higher covariance of the general factor, as mentioned before).

Table 5a: Capital ratios (in percent) by loans' characteristics

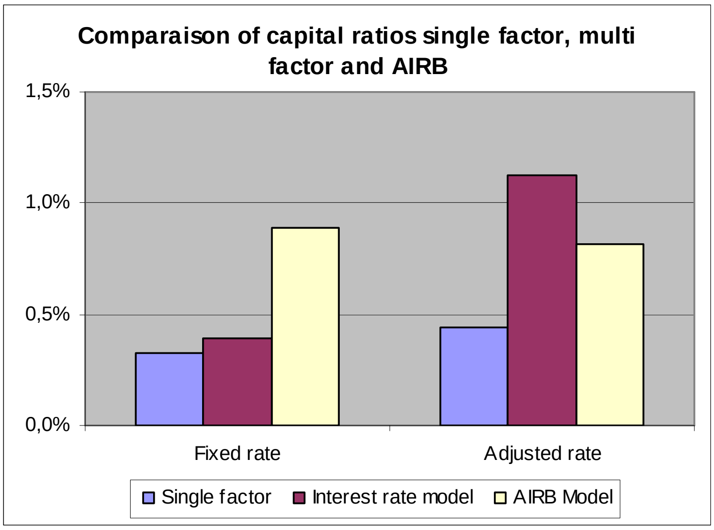
		Single factor model	Interest rate model	Combination of interest rate and loans type model	IRBA model
Type of loans' rate	Fixed rate	0.329	0.391		0.890
	Adjusted rate	0.444	1.127		0.815
Type of loans* Type of rates	Fixed rate * Borrowers using only subsidized loans	0.185		0.462	0.642
	Fixed rate * Borrowers using only market loans	0.463		0.290	1.018
	Fixed rate * Borrowers using both subsidized and market loans	0.573		0.899	0.959
	Adjustable rate * Borrowers using only subsidized loans	0.421		1.079	0.799
	Adjustable rate * Borrowers using only market loans	0.464		0.829	0.906
	Adjustable rate * Borrowers using both subsidized and market loans	0.479		0.840	0.721

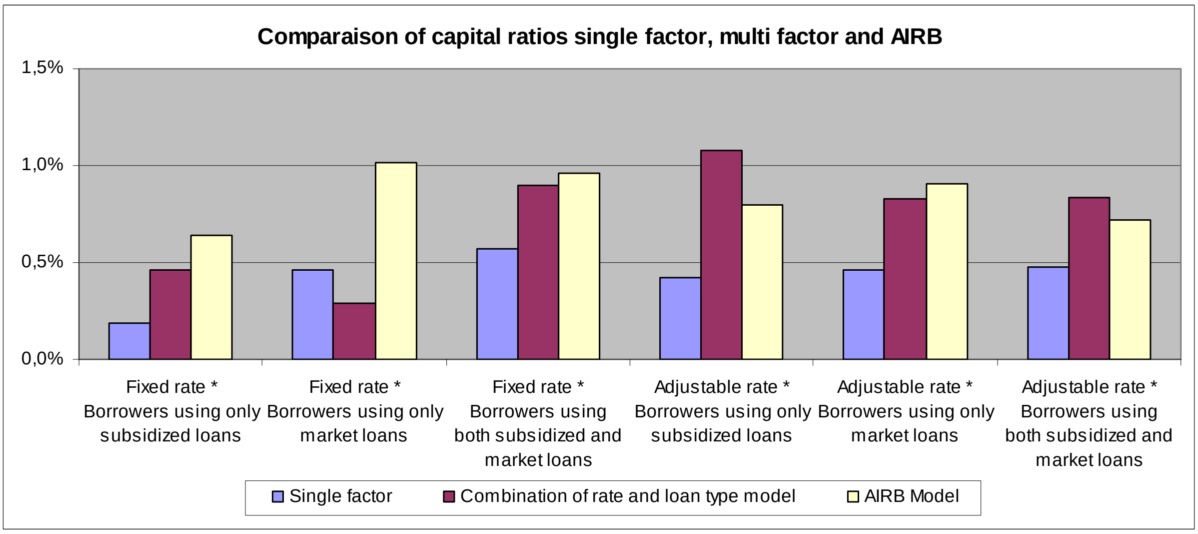
Table 5b: Capital ratios (in percent) by geographic area

		Single factor model	Geographic model	IRBA model
Type of loan* Type of interest rate	Area 1 (rural – stable)	0.392	0.413	0.821
	Area 2 (urban – developing)	0.348	0.360	0.804
	Area 3 (rural – ageing)	0.324	0.358	0.851
	Area 4 (urban – in decline)	0.404	0.404	0.849
	Area 5 (median areas)	0.379	0.400	0.882

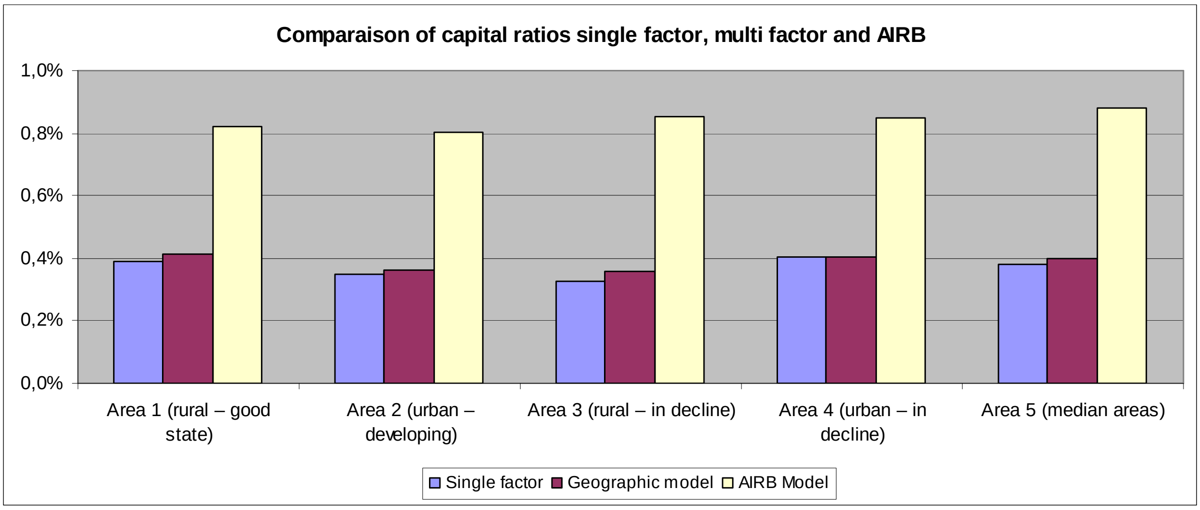
Note: Column IRBA presents the IRBA regulatory capital charges. All computations are based on a LGD of 20%, close but higher than the effective computed rate for the bank which provided data.

Graph 1a: Comparison of capital ratios (in percent) by loans' characteristics (interest rate type only and combination of interest rate type and type of loans)

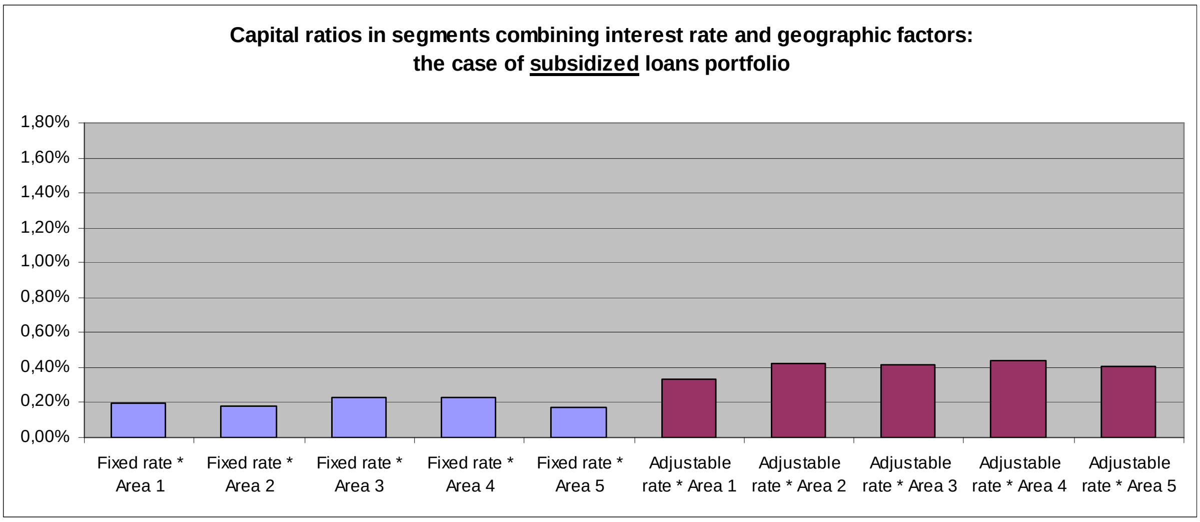


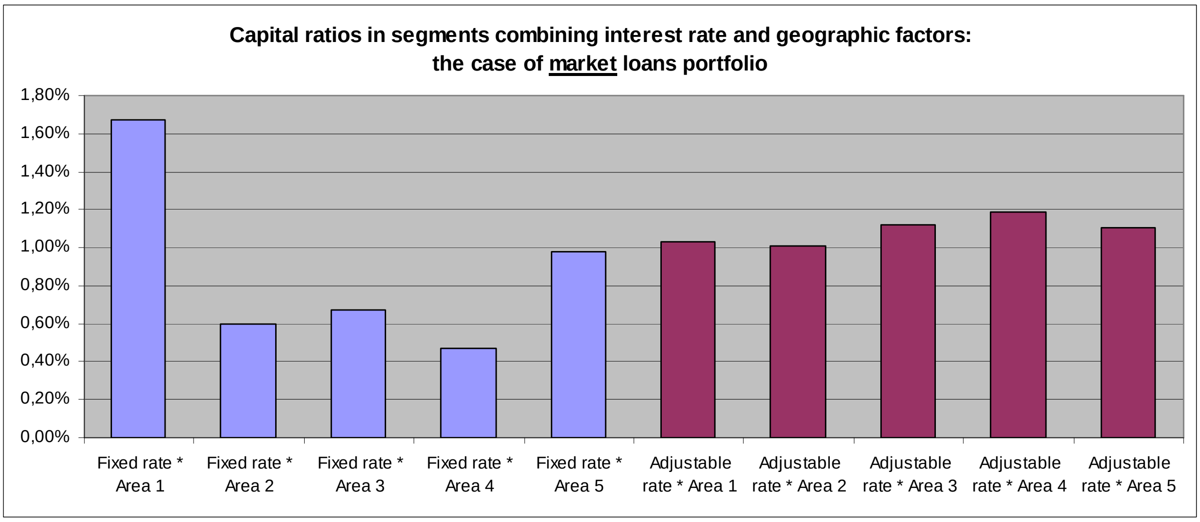


Graph 1b: Comparison of capital ratios (in percent) by environment



Graph 1c: Comparison of multifactor capital ratios (in percent) by environment and type of interest rate





These results suggest that the introduction of additional risk factors has an important effect on the representation of the risk's distribution within a mortgage loans portfolio. Moreover, the dimensions explored in the paper show also that the suggested overlap of interest rate type, loan type and location factors might be limited, as they lead to quite different capital allocations, given the chosen aggregation variable for economic capital.

The last main result is the limited ability of the IRBA weighting scheme to capture heterogeneity in the portfolio. This can be explained because, in mortgages portfolio, the correlation is fixed (equal to 15%) in Basel II formula. Consequently, only differences in rating levels can lead to differences in capital ratios. Differences coming from additional risk factors are not accounted for. The usefulness of the regulatory model may be rather limited in the context of the Pillar 2 of Basel II. In addition, we notice that IRBA ratios are higher than the economic capital ratio, except for the sub-portfolio composed of adjustable rate mortgages.

5. Conclusion

In the context of Pillar 2 of Basel 2, the detection and measurement of credit risk concentration in large portfolios calls for the extension of the regulatory framework in order to introduce additional sources of systematic risk in the modelling of credit risk. In this context, our results show that it is necessary to deeply investigate the various origins of borrower's failures in mortgages portfolios. Such investigation should allow a better understanding of the reasons why the choice between ARMs and FRMs as well as the borrower's geographic location influence the losses distribution function at the portfolio's level.

This paper gives some qualitative insights on the potential concentration within large portfolios of mortgages by extending the standard asymptotic single factor model to a multi-factor framework. Two additional types of factors were introduced in the multifactor model: location factors and mortgage characteristics factors. Taking into account mortgage characteristics factors – ARM or FRM, subsidized mortgage or free market mortgage – really improves the discrimination of portfolios segments in terms of credit risk. Thus, on average, subsidized mortgages are less risky than free market mortgages. Consequently, the single factor model fails in capturing potential risk concentration. Thus, the heterogeneity captured by credit ratings, the only source of heterogeneity in the asymptotic one factor framework, fails to describe the effective heterogeneity in default rates within large retail portfolios and other factors might be at play. In addition, our findings show that the combination of environment factors (captured by the characteristics of the borrowers' employment area) and mortgages characteristics factors improves the accurate measurement of portfolio credit risk. Indeed, the local factors could compensate the effect of the general macroeconomic factor, contributing to reduce the economic capital consumption. This is the cas for FRMs, whatever the type of mortgage, subsidized or not. In particular, in the mortgages portfolio we studied, even if the choice between fixed and adjustable interest rate determines the most important source of heterogeneity in loans losses. Our findings show that home ownership supporting public programs induced additional heterogeneity. In fact, by releasing financial constraints, these programs succeed in lowering credit concentration, in particular if borrowers use FRMs public supported loans.

To summarize, beside interest rate type choices, access to home ownership public support programs and local economic conditions play a role in the characterization of risk concentration in mortgages portfolios. Indeed, the combination of geographic location and loans characteristics effect seems to improve the modelling of mortgages credit risk.

References

BCBS, 2006, Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version, Basel Committee on Banking Supervision.

BCBS, 2008, Range of practices and issues in economic capital modelling, Consultative document, Basel Committee on Banking Supervision.

Ben-Shahar D., 2004, Screening mortgage default risk: A unified theoretical framework, *Journal of Real Estate Research* 28, 215-239.

Blume M.E., F. Lim, A.C. Macklinlay, 1998, The declining credit quality of US corporate debt: myth or reality, *Journal of Finance* 53 (4), 1389-1413.

Calem P. S., M. LaCour-Little, 2004, Risk-based capital requirements for mortgage loans, *Journal of Banking and Finance* 28, 647-672.

Campbell, J.Y, 2006, Household finance, Journal of Finance, 61, 1553-1604

Campbell, J.Y, and J. Cocco, 2003, Household risk management and optimal mortgage choice, Quarterly Journal of Economics, 118, 1449-1494.

Diaz-Serrano L., 2005, Income volatility and residential mortgage delinquency across the EU, *Journal of Housing Economics* 14, 153-177.

Dietsch M., J. Petey, 2004, Should SME exposures be treated as retail or corporate exposures ? A comparative analysis of default probabilities and asset correlations in French and German SMEs, *Journal of Banking and Finance* 28, 773-788.

Egloff D., M. Leippold, S. Jöhri, C. Dalbert, 2005, Optimal importance sampling for credit portfolios with stochastic approximation, working paper.

Frey, R., A. McNeil, 2003, Dependent defaults in models of portfolio credit risk. *Journal of Risk* 6 (1), 59–92.

Glasserman P., J. Li, 2005, Importance sampling for portfolio credit risk. *Management Science*, 51(11), 1643–1656.

Gouriéroux C., J.P. Laurent, O. Scaillet, 2000, Sensitivity analysis of Values at Risk, *Journal of Empirical Finance* 7, 225-245.

Heitfield N., S. Burton, S. Chomsisengphet, 2006, Systematic and idiosyncratic risk in syndicated loan portfolios, *Journal of Credit Risk* 2 (3), 3-31.

Higham N., 2002, Computing the nearest correlation matrix: a problem from finance, *IMA Journal of Numerical Analysis* 22, 329-343.

Klebaner F. C., *Introduction to Stochastic Calculus with Applications*. Imperial College Press, second edition, 2005.

Koijen,R., O. Van Hemert, and S. Van Nieuwerburgh, 2009, Mortgage Timing, Journal of Financial Economics, 93 (2, August), 292-324.
Lopez J.A., M.R. Saidenberg, 2000, Evaluating credit risk models, *Journal of Banking and Finance* 24, 151-165.

Lucas A., P. Klaassen, P. Spreij, S. Straetmans, 2001, An analytic approach to credit risk of large corporate bond and loan portfolios, *Journal of Banking and Finance* 25, 1635-1664.

McNeil A., J. Wendin J, 2007, Bayesian inference for generalized linear mixed models of portfolio credit risk, *Journal of Empirical Finance* 14, 131-149.

Mian A. and A.Sufi, (2008) The consequences of mortgage credit expansion: evidence from the 2007 mortgage default crisis, NBER Working Paper 13936, http://www.nber.org/papers/w13936

Pesaran M.H., T. Schuermann, B.J. Treutler, 2005, The role of industry, geography and firm heterogeneity in credit risk diversification, IEPR working paper 05.25.

Posey L. L., A. Yavas, 2001, Adjustable and fixed rate mortgages as a screening mechanism for default risk, *Journal of Urban Economics* 49, 54-79.

Tasche, D., 1999, Risk contributions and performance measurement, Working paper, Technische Universität München.

Tasche D., 2006, Measuring sectoral diversification in an asymptotic multi-factor framework, *Journal of Credit Risk* 2 (3), 33-55.

Tasche D., 2009, Capital allocation for credit portfolios with kernel estimators, *Quantitative Finance* 9(5), 581-595.

Wolfinger R., M. O'Connell, 1993, Generalized linear mixed models: A pseudo-likelihood approach, *Journal of Statistical Computation and Simulation* 4, 233-243.

Appendix 1.

Estimation results: Homogeneous one factor model

A. Goodness-of-fit measures

-2 Res Log Pseudo-Likelihood	304.51
Pseudo AIC	306.51
Pseudo BIC	306.99
Generalized Chi-Square	512.90
Gener. Chi-Square / DF	9.33

B. Covariance parameters

Parameter	Mean	Std error
Intercept	0.01243	0.005382

C. Default thresholds and probabilities

Rating	Estimate	Std error	P-value	Default probability (%)
1	-3.3501	0.03644	<.0001	0.040
2	-3.3094	0.03385	<.0001	0.047
3	-3.1994	0.03473	<.0001	0.069
4	-1.9679	0.03296	<.0001	2.454
5	-0.6710	0.03260	<.0001	25.111

Appendix 2.A

Estimation results: Interest rate model

This specification considers two additional risk factors associated to two types of interest rate classes in the population: fixed rate mortgages or adjustable rate mortgages.

A. Goodness-of-fit measures

-2 Res Log Pseudo-Likelihood	939.56
Pseudo AIC	949.56
Pseudo BIC	951.98
Generalized Chi-Square	1294.29
Gener. Chi-Square / DF	11.25

B. Covariance parameters

Parameter	Mean	Std error
CHOL(1,1)	0.01681	.
CHOL(2,2)	0.2033	0.04258
CHOL(3,1)	-0.05761	.
CHOL(3,2)	0.08150	0.06512
CHOL(3,3)	0.06964	0.02127

Note: CHOL(,) denotes the position (row, column) of the estimated parameter in the Cholesky root of the covariance matrix of random effects. Non-shown elements of the matrix are zero. For all models, the covariance matrix of random effect is parameterized through its Cholesky root *A*. This parameterization ensures that the resulting variance-covariance matrix should at least be positive semi-definite. If all diagonal values are nonzero, it is positive definite. The variance-covariance matrix of random effects *C* is then simply defined as *C* = *A'A*.

C. Default thresholds and probabilities

Rating	Estimate	Std error	P-value	Default probability (%)
1	-3.25	0.02843	<.0001	0.058
2	-3.15	0.02546	<.0001	0.081
3	-3.00	0.02712	<.0001	0.133
4	-1.78	0.02474	<.0001	3.726
5	-0.46	0.02457	<.0001	32.240

D. Covariance matrix of random effects

	Types of rate		
	Fixed rate	Adjustable rate	Syst
Fixed rate	0.000283		-0.00097
Adjustable rate		0.04132	0.01657
General	-.000970	0.01657	0.01481

Note: Syst denotes the general risk factor.

E. Tests of Covariance Parameters

Test Null hypothesis	-2 Res Log P-Like	ChiSq	Pr > ChiSq
$\sigma_1^2 = 0$	939.56	0.00	1.0000
$\sigma_2^2 = 0$	939.56	0.00	1.0000
$\sigma_{13} = \sigma_{23} = 0$	941.23	1.68	0.4325

Appendix 2.B

Estimation results: Combination of rate and loan type model

This specification considers six combinations of the loan's type and the rate's nature as additional random effects.

A. Goodness-of-fit measures

-2 Res Log Pseudo-Likelihood	1257.05
Pseudo AIC	1279.05
Pseudo BIC	1284.39
Generalized Chi-Square	1974.23
Gener. Chi-Square / DF	5.56

B. Covariance parameters

Parameter	Mean	Std error
CHOL(1,1)	0.09437	0.02335
CHOL(2,2)	0.2401	0.05049
CHOL(3,3)	0.1322	0.03614
CHOL(4,4)	0.1359	0.02994
CHOL(5,5)	0	.
CHOL(6,6)	0.1295	0.02910
CHOL(7,1)	-0.09408	0.02465
CHOL(7,2)	-0.1848	0.07301
CHOL(7,3)	0.02991	0.01827
CHOL(7,4)	0.06807	0.03545
CHOL(7,5)	-3,21E-24	0.02080
CHOL(7,6)	-0.04436	0.02087
CHOL(7,7)	0	.

Note: CHOL(,) denotes the position (row, column) of the estimated parameter in the Cholesky root of the covariance matrix of random effects. Non-shown elements of the matrix are zero. For all models, the covariance matrix of random effect is parameterized through its Cholesky root *A*. This parameterization ensures that the resulting variance-covariance matrix should at least be positive semi-definite. If all diagonal values are nonzero, it is positive definite. The variance-covariance matrix of random effects *C* is then simply defined as $C = A'A$.

C. Default thresholds and probabilities

Rating	Estimate	Std error	P-value	Default probability (%)
1	-3.30	0.01983	<.0001	0.049
2	-3.15	0.01548	<.0001	0.083
3	-3.02	0.01800	<.0001	0.128
4	-1.78	0.01425	<.0001	3.714
5	-0.45	0.01399	<.0001	32.685

D. Covariance matrix of random effects

	Fixed rate * subsidized loans	Fixed rate * market loans	Fixed rate * booth subsidized and market loans	Adjustable rate * subsidized loans	Adjustable rate * market loans	Adjustable rate * subsidized and market loans	Syst
Fixed rate * subsidized loans	0.008906						-0.008880
Fixed rate * market loans		0.05763					-0.044370
Fixed rate * both subsidized and market loans			0.017490				0.003955
Adjustable rate * subsidized loans				0.018480			0.009253
Adjustable rate * market loans					0		0
Adjustable rate * subsidized and market loans						0.01678	-0.005750
General	-.008880	-0.04437	0.003955	0.009253	0	-0.00575	0.050500

Note: Syst denotes the general risk factor.

E. Tests of Covariance Parameters

Test Null hypothesis	-2 Res Log P-Like	ChiSq	Pr > ChiSq
$\sigma_1^2 = 0$	1301.89	44.84	<.0001
$\sigma_2^2 = 0$	1628.66	371.60	<.0001
$\sigma_3^2 = 0$	1281.08	24.03	<.0001
$\sigma_4^2 = 0$	1384.97	127.91	<.0001
$\sigma_5^2 = 0$	1257.05	0.00	1.0000
$\sigma_6^2 = 0$	1357.49	100.43	<.0001
$\sigma_{17} = \sigma_{27} = \sigma_{37} = 0,$ $\sigma_{47} = \sigma_{57} = \sigma_{67} = 0$	1450.15	193.10	<.0001

Appendix 2.C

Estimation results: Geographic model

This specification considers five geographic areas as random factors by using the geographic segmentation of borrowers.

A. Goodness-of-fit measures

-2 Res Log Pseudo-Likelihood	-27.97
Pseudo AIC	-11.97
Pseudo BIC	-8.09
Generalized Chi-Square	787.46
Gener. Chi-Square / DF	2.67

B. Covariance parameters

Parameter	Mean	Std error
CHOL(1,1)	0.02401	0.01668
CHOL(2,2)	0.04562	0.01266
CHOL(3,3)	0.04269	0.02223
CHOL(6,1)	0.09634	0.02706
CHOL(6,2)	-0.04409	0.03348
CHOL(6,3)	-0.03344	0.04377
CHOL(6,4)	5.38E-24	0.08576
CHOL(6,5)	-1,01E-35	0.08576
CHOL(6,6)	0	.

Note: CHOL(,) denotes the position (row, column) of the estimated parameter in the Cholesky root of the covariance matrix of random effects. Non-shown elements of the matrix are zero. For all models, the covariance matrix of random effect is parameterized through its Cholesky root *A*. This parameterization ensures that the resulting variance-covariance matrix should at least be positive semi-definite. If all diagonal values are nonzero, it is positive definite. The variance-covariance matrix of random effects *C* is then simply defined as *C* = *A'A*.

C. Default thresholds and probabilities

Rating	Estimate	Std error	P-value	Default probability (%)
1	-3.34	0.03127	<.0001	0.041
2	-3.31	0.02801	<.0001	0.047
3	-3.19	0.02909	<.0001	0.070
4	-1.96	0.02691	<.0001	2.476
5	-0.67	0.02646	<.0001	25.251

D. Covariance matrix of random effects

	Area 1	Area 2	Area 3	Area 4	Area 5	Syst
Area 1	0.000577					0.002314
Area 2		0.002081				-0.00201
Area 3			0.001823			-0.00143
Area 4				0		0
Area 5					0	0
Syst	0.002314	-0.00201	-0.00143	0	0	0.01234

Note: Syst denotes the general risk factor.

E. Tests of Covariance Parameters

Test Null hypothesis	-2 Res Log P-Like	ChiSq	Pr > ChiSq
$\sigma_1^2 = 0$	-25.8631	2.10	0.1470
$\sigma_2^2 = 0$	-9.8113	18.15	<.0001
$\sigma_3^2 = 0$	-25.9268	2.04	0.1533
$\sigma_4^2 = 0$	-27.9660	0.00	1.0000
$\sigma_5^2 = 0$	-27.9660	0.00	1.0000
$\sigma_{16} = \sigma_{26} = \sigma_{36} = \sigma_{46} = \sigma_{56} = 0$	401.70	429.67	<.0001

Appendix 2.D

Estimation results: Combination of geographic areas and types of interest rate model for subsidized loans

This specification considers ten classes issued of the combination of the five geographic areas and the two types of interest rate. First five columns correspond to fixed rate loans and the following five to adjustable rate loans.

A. Goodness-of-fit measures

-2 Res Log Pseudo-Likelihood	300.72
Pseudo AIC	334.72
Pseudo BIC	342.96
Generalized Chi-Square	954.22
Gener. Chi-Square / DF	1.60

B. Covariance parameters

Parameter	Mean	Std error
CHOL(1,1)	0.08395	0.04613
CHOL(2,2)	0.09124	0.02887
CHOL(3,3)	0.1328	0.05332
CHOL(4,4)	0.08237	0.02915
CHOL(5,5)	0.1465	0.03904
CHOL(7,7)	0.08056	0.02380
CHOL(9,9)	0.03723	0.02450
CHOL(11,1)	-0.03399	0.02653
CHOL(11,2)	-0.05439	0.03751
CHOL(11,3)	0.004519	0.02806
CHOL(11,4)	-0.06207	0.03630
CHOL(11,5)	-0.05129	0.04116
CHOL(11,6)	-9,82E-25	0.03037
CHOL(11,7)	-0.01268	0.02392
CHOL(11,8)	-2,62E-24	0.03037
CHOL(11,9)	0.004743	0.02713
CHOL(11,10)	-2,63E-53	0.03037
CHOL(11,11)	0	.

Note: CHOL(,) denotes the position (row, column) of the estimated parameter in the Cholesky root of the covariance matrix of random effects. Non-shown elements of the matrix are zero. For all models, the covariance matrix of random effect is parameterized through its Cholesky root *A*. This parameterization ensures that the resulting variance-covariance matrix should at least be positive semi-definite. If all diagonal values are nonzero, it is positive definite. The variance-covariance matrix of random effects *C* is then simply defined as $C = A'A$.

C. Default thresholds and probabilities

Rating	Estimate	Std error	P-value	Default probability (%)
1	-3.35	0.03305	<.0001	0.041
2	-3.32	0.02192	<.0001	0.045
3	-3.09	0.02670	<.0001	0.100
4	-1.96	0.01975	<.0001	2.527
5	-0.67	0.01843	<.0001	25.044

D. Covariance matrix of random effects

	Area 1 * fixed rate	Area 2 * fixed rate	Area 3 * fixed rate	Area 4 * fixed rate	Area 5 * fixed rate	Area 1 * adjust. rate	Area 2 * adjust. rate	Area 3 * adjust. rate	Area 4 * adjust. rate	Area 5 * adjust. rate	general
Area 1 * fixed rate	0.007048										-0.00285
Area 2 * fixed rate		0.008325									-0.00496
Area 3 * fixed rate			0.01764								0.000600
Area 4 * fixed rate				0.006785							-0.00511
Area 5 * fixed rate					0.02145						-0.00751
Area 1 * adjust. rate						0					0
Area 2 * adjust. rate							0.006490				-0.00102
Area 3 * adjust. rate								0			0
Area 4 * adjust. rate									0.001386		0.000177
Area 5 * adjust. rate										0	0
general	-0.00285	-0.00496	0.000600	-0.00511	-0.00751	0	-0.00102	0	0.000177	0	0.01080

Note: general denotes the general risk factor.

E. Tests of Covariance Parameters

Test Null hypothesis	-2 Res Log P-Like	ChiSq	Pr > ChiSq
$\sigma_1^2 = 0$	302.65	1.93	0.1645
$\sigma_2^2 = 0$	310.45	9.73	0.0018
$\sigma_3^2 = 0$	304.64	3.93	0.0475
$\sigma_4^2 = 0$	307.22	6.51	0.0108
$\sigma_5^2 = 0$	322.48	21.77	<.0001
$\sigma_6^2 = 0$	300.72	0.00	1.0000
$\sigma_7^2 = 0$	314.46	13.74	0.0002
$\sigma_8^2 = 0$	300.72	0.00	1.0000
$\sigma_9^2 = 0$	301.68	0.97	0.3252
$\sigma_{10}^2 = 0$	300.72	0.00	1.0000
$\sigma_{111} = \sigma_{211} = \sigma_{311} = \sigma_{411} = \sigma_{511} = \sigma_{611} = \sigma_{711} = \sigma_{811} = \sigma_{911} = \sigma_{1011} 0$	430.49	129.78	<.0001

Appendix 2.E

Estimation results: Combination of geographic areas and types of interest rate model for market loans

This specification considers ten classes issued of the combination of the five geographic areas and the two types of interest rate. First five columns correspond to fixed rate loans and the following to adjustable rate loans.

A. Goodness-of-fit measures

-2 Res Log Pseudo-Likelihood	756.13
Pseudo AIC	790.13
Pseudo BIC	798.37
Generalized Chi-Square	1314.65
Gener. Chi-Square / DF	2.21

B. Covariance parameters

Parameter	Mean	Std error
CHOL(1,1)	0.2420	0.05713
CHOL(2,2)	0.1851	0.04289
CHOL(3,3)	0.2845	0.06705
CHOL(4,4)	0.2064	0.04671
CHOL(5,5)	0.2364	0.05312
CHOL(7,7)	0.01535	0.05968
CHOL(9,9)	0.05674	0.02540
CHOL(11,1)	0.1239	0.05404
CHOL(11,2)	-0.1335	0.09107
CHOL(11,3)	-0.01526	0.06365
CHOL(11,4)	-0.2015	0.08673
CHOL(11,5)	0.03339	0.07941
CHOL(11,6)	9.22E-24	0.04308
CHOL(11,7)	-0.00088	0.04262
CHOL(11,8)	1.31E-24	0.04308
CHOL(11,9)	0.01344	0.04590
CHOL(11,10)	-3,14E-56	0.04308
CHOL(11,11)	0	.

Note: CHOL(,) denotes the position (row, column) of the estimated parameter in the Cholesky root of the covariance matrix of random effects. Non-shown elements of the matrix are zero. For all models, the covariance matrix of random effect is parameterized through its Cholesky root *A*. This parameterization ensures that the resulting variance-covariance matrix should at least be positive semi-definite. If all diagonal values are nonzero, it is positive definite. The variance-covariance matrix of random effects *C* is then simply defined as *C* = *A'A*.

C. Default thresholds and probabilities

Rating	Estimate	Std error	P-value	Default probability (%)
1	-3.32	0.03020	<.0001	0.046
2	-3.14	0.02552	<.0001	0.085
3	-3.03	0.02925	<.0001	0.121
4	-1.68	0.02333	<.0001	4.669
5	-0.47	0.02320	<.0001	32.077

D. Covariance matrix of random effects

	Area 1 * fixed rate	Area 2 * fixed rate	Area 3 * fixed rate	Area 4 * fixed rate	Area 5 * fixed rate	Area 1 * adjust. rate	Area 2 * adjust. rate	Area 3 * adjust. rate	Area 4 * adjust. rate	Area 5 * adjust. rate	general
Area 1 * fixed rate	0.05856										0.02997
Area 2 * fixed rate		0.03425									-0.02470
Area 3 * fixed rate			0.08097								-0.00434
Area 4 * fixed rate				0.04262							-0.04159
Area 5 * fixed rate					0.05588						0.007894
Area 1 * adjust. rate						0					0
Area 2 * adjust. rate							0.000236				-0.00001
Area 3 * adjust. rate								0			0
Area 4 * adjust. rate									0.003219		0.000763
Area 5 * adjust. rate										0	0
general	0.02997	-0.02470	-0.00434	-0.04159	0.007894	0	-0.00001	0	0.000763	0	0.07528

Note: general denotes the systematic general risk factor.

E. Tests of Covariance Parameters

Test Null hypothesis	-2 Res Log P-Like	ChiSq	Pr > ChiSq
$\sigma_1^2 = 0$	808.68	52.55	<.0001
$\sigma_2^2 = 0$	799.60	43.47	<.0001
$\sigma_3^2 = 0$	806.59	50.46	<.0001
$\sigma_4^2 = 0$	810.04	53.91	<.0001
$\sigma_5^2 = 0$	825.78	69.65	<.0001
$\sigma_6^2 = 0$	756.13	0.00	1.0000
$\sigma_7^2 = 0$	756.15	0.02	0.8940
$\sigma_8^2 = 0$	756.13	0.00	1.0000
$\sigma_9^2 = 0$	758.79	2.67	0.1025
$\sigma_{10}^2 = 0$	756.13	0.00	1.0000
$\sigma_{111} = \sigma_{211} = \sigma_{311} = \sigma_{411} = \sigma_{511} = \sigma_{611} = \sigma_{711} = \sigma_{811} = \sigma_{911} = \sigma_{1011} 0$	825.84	69.71	<.0001