Université de Montréal
Faculté des arts et des sciences
Département de sciences économiques

CAHIER 9528

ON THE REPAYMENT OF PERSONAL LOANS UNDER ASYMMETRICAL INFORMATION: A COUNT DATA MODEL APPROACH

G. Dionne[1], M. Artís[2] and M. Guillén[2]

1 Département de sciences économiques and CRT, Université de Montréal.

2 University of Barcelona.

May 1995

M. Guillén and M. Artís acknowledge support received from the DGICYT under grant PB92-0545. M. Guillén thanks the University of Texas at Austin where she was a visiting researcher in 1994 and G. Dionne thanks the FFSA for financial support and the Paris-X-Nanterre-CEPREMAP Chair of Insurance for their hospitality during the academic year 1993-94. Charles Vanasse, the editor and the referee are acknowledged for their comments.

C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7

Télécopieur (FAX): (514) 343-5831
Courrier électronique (E-Mail): econo@tornade.ERE.Umontreal.CA

Dépôt légal - 1995
Bibliothèque nationale du Québec
Bibliothèque nationale du Canada

ISSN 0709-9231

RÉSUMÉ

Dans la littérature récente sur la classification des risques bancaires, l'emphase a été mise sur l'estimation des probabilités de défaillance sans vraiment se préoccuper des bénéfices et des coûts associés aux prêts. Il est bien connu qu'un mauvais prêt peut devenir coûteux à la banque en introduisant des frais de gestion supplémentaires. Dans cet article, nous proposons une extension des modèles connus pour estimer conjointement les probabilités de défaillance et les espérances de coûts des bons et des mauvais risques. Nous montrons que les variables significatives des différents modèles ne sont pas les mêmes et nous obtenons que les distributions conditionnelles des accidents de paiements des bons et des mauvais risques diffèrent. Ces résultats impliquent que limiter l'analyse à l'estimation des probabilités de défaillance pour évaluer les risques de crédit n'est pas suffisant.

Mots clés : prêts personnels, modèles de comptage, Poisson tronqué, probabilité de défaillance, information asymétrique.

Journal of Economic Literature Classification Numbers : G2, C35.

ABSTRACT

In the recent literature on credit scoring, the emphasis was made on the estimation of default probabilities without any real effort of considering the different costs and benefits of the loans. However, an accepted loan may become a bad loan after some costly reminders and may even introduce collection and other bad debt costs. In this paper, we extend the hurdle model defined by Mullahy (1986) to estimate jointly the default probability and the two conditional truncated distributions of non-payments of good and bad loans respectively. We first show that the significant variables that affect the three distributions are not the same. Moreover, the two truncated non-payments distributions (before and after the identification of a bad loan) do not follow the same distribution. These results imply that limiting the analysis to the estimation of the default probabilities to evaluate credit scoring is not sufficient to obtain an appropriate evaluation of the ex-ante bank funding.

Keywords : Personal loans, hurdle count data model, truncated Poisson, default probability, asymmetrical information.

Journal of Economic Literature Classification Numbers : G2, C35.

1 Introduction

In the recent literature on credit scoring, the emphasis was made on the estimation of default probabilities without any real effort of considering the different costs and benefits of the loans[1]. However, an accepted loan may become a bad loan after a predetermined number of costly reminders; it may also introduce collection and other bad debt costs after some additional non-payments. In the two corresponding time periods (before and after being identified as a bad loan) there are administrative costs that have to be estimated by the bank when she takes the decision to give a loan. *Ex-ante*, these individual costs are random since the bank does not know how many non-payments a bad loan will generate.

Non-payments are non-negative integer or count data. Many methodologies can be used to estimate the individual's expected number of non-payments. If we assume that the non-payments follow a Poisson distribution, for example, it is not clear that the count data follow the same process before and after the loan is identified to be a bad loan or once a given threshold is reached. In this paper we extend the hurdle model methodology defined by Mullahy (1986) to estimate jointly the default probability[2] and the two

[1] See, however, Boyes, Hoffman and Low (1989) and some other few exceptions discussed in Altman, Avery, Eisenbeis and Sinkey (1981).

[2] As we will see later, we will not estimate the default probability traditionally discussed in the literature, since we do not have information on those who have not received a loan. However, the methodology proposed here can be extended to cope with the sample selection problem. Therefore, in this paper, default probability means the probability to become a bad loan once the loan is accepted.

conditional truncated distributions of non-payments of good and bad loans respectively. Our results indicate that the two conditional distributions do not follow the same distribution. We also show that the significant variables that affect the three distributions are not the same which implies that limiting the analysis to the estimation of default probabilities to evaluate the credit scoring is not sufficient to obtain an appropriate evaluation of the expected net returns of bank funding. Precise estimates of default probabilities <u>and</u> conditional expected administrative costs of loans are necessary to evaluate both the probability to have a good loan and the conditional behavior of borrowers concerning their scheduling of non-payments. We will show how our model can be used to study the consequences of risks misclassification in terms of expected profits for the bank.

The paper is organized as follows. In the next section, we propose a theoretical model that trades off the benefits and costs of loans with emphasis on non-recovered debt and expected administrative costs of good and bad loans. We also discuss the links between credit rationing and credit scoring in a world of asymmetrical information in credit markets. Section 3 is devoted to the presentation of the extended econometric model. Specifically we present a Negative Binomial hurdle specification that is convenient for the estimation of the truncated distributions of non-payments. Section 4 describes the variables and the data obtained from a Spanish bank and Section 5 presents and discusses the main results of the study. A conclusion summarizes the contributions of the article and suggests some avenues of future research.

2 Theoretical Model

We consider a model where a bank has to allow credit to borrowers of different types. To set the model we first assume that all the borrowers' characteristics are observable. Then we introduce asymmetrical information on default probabilities and expected collection costs. In particular, we consider two types of expected collection costs that correspond to the behavior of the bank from which the data come from: 1) those from "good loans" with no more than three monthly non-payments and 2) those from "bad loans" with four or more monthly non-payments.

When considering a client, the bank faces the following risky situation (Fourgeaud et al., 1990):

Quality of loan \ Decision	Accepted	Refused
good loan (with probability p)	$D[r-\mu] - \mathrm{E}_0(c_0)$	$-c$
bad loan (with probability $1-p$)	$D[r-\mu] - \mathrm{E}_1(c_1)$ $-D(1+r)\pi$	$-c$

where

D	is the amount of loan;
r	is the interest rate;
μ	is the opportunity cost of the bank;
$\mathrm{E}_0(c_0)$	is the conditional expected administrative cost of good loans; ex-ante the bank does not know the number of monthly non-payments that will be accumulated (0,1,2,3). However, under full information the bank knows the parameters of each individual distribution;

$E_1(c_1)$ is the conditional expected administrative cost of bad loans including all collection costs; again the bank does not know the number of non-payments ex-ante;

π is the fraction of the loan that will not be recovered and will not be available for the next period.

Intuitively $c < c_0 < c_1$ and $E_1(c_1) > E_0(c_0)$. Moreover, under full information $c \equiv 0$.

We can then write the expected profit of the bank for a given client with default probability $(1-p)$ and expected costs $E_0(c_0)$ and $E_1(c_1)$ as:

$$D(r-\mu) - pE_0(c_0) - (1-p)E_1(c_1) - (1-p)D(1+r)\pi \tag{1}$$

Under competition, this expected profit for a given type of risk has to be null which implies that the equilibrium interest rate of a given type solves:

$$r^* = \frac{D\mu + D\pi(1-p) + pE_0(c_0) + (1-p)E_1(c_1)}{D(1-\pi(1-p))} \tag{2}$$

where r^* is increasing in $(1-p)$ and where $D\mu$ is the opportunity cost of capital. $D\pi(1-p)$ is the expected loss of future business, $pE_0(c_0)$ is the expected administrative cost of good loans and $(1-p)E_1(c_1)$ is the expected cost of bad loans.

It can be shown that all potential clients that know their true characteristics will accept to pay the competitive interest rate corresponding to their individual risk and will maximize their welfare by choosing an optimal amount of loan at that interest rate. Without asymmetrical information the optimal form of the contract does not matter.

Under asymmetrical information matters are less simple. Let us first discuss the case where only p is not observable by the bank. Many mechanisms have been developed to reduce the social cost of asymmetrical information. One is the use of credit rationing and debt in an environment of adverse selection and costly state verification (Boyd and Smith, 1993; Stiglitz and Weiss, 1981; Jaffee and Russell, 1976). Such mechanisms permit the self-selection of borrowers. However, they limit the access to credit for some good borrowers and some equilibrium may not be optimal in the sense of Harris and Townsend (1981). Multi-period contracting can improve resource allocation under adverse selection (Dionne and Doherty, 1994) as well as risk classification (Crocker and Snow, 1986) or in our context, credit scoring. Credit scoring will improve resource allocation from a second best allocation, particularly, when differences between the potential borrowers are large or when the cost of information is not too high which is usually the case for many banking industries, including the one we study.

From equation (2) we observe that r^* is directly related to p, $E_0(c_0)$ and $E_1(c_1)$. When the bank cannot observe p or the parameters of both $E_0(c_0)$ and $E_1(c_1)$, she can complement self-selection mechanisms by estimating these numbers and by using credit scoring. Appropriate estimations of the parameters are important for the bank in order to reduce misclassification of both good and bad risks that would imply less profits: bad risks that are incorrectly classified mean *ex-post* losses and good risks classified as bad risks will leave.

In the standard literature on credit scoring applications, the emphasis was put on the evaluation of default probability $(1-p)$ (Altman et. al., 1981; Steenackers et al., 1989). However, to our knowledge, no study has investigated the estimation of both $E_0(c_0)$ and $E_1(c_1)$ or of the parameters of their underlying distributions.

In the next section we introduce the specification of the econometric models that permit the estimation of these distributions.

3 Specification of the Econometric Models

Poisson models are frequently used to estimate non-negative integer outcomes that suffer from restrictions involving the equality of mean and variance (conditional on a set of explanatory variables) of the process to be estimated.

The Poisson model establishes that Y_i, the number of unpaid payments in our application, is Poisson distributed

$$\text{Prob}(Y_i = j) = \frac{e^{-\lambda_i}\lambda_i^j}{j!} \qquad \begin{array}{l} j = 0, 1, 2, \ldots \\ \lambda_i > 0 \end{array}$$

where λ_i is the parameter that may differ for every individual, since it may be a function of individual characteristics. As we said, for this model, the restriction establishes that $\text{E}(Y_i|X_i) = \text{Var}(Y_i|X_i) = \lambda_i$, where X_i is the vector of observable individual characteristics.

A widely accepted way to overcome this restriction that has been used by many authors is the extension to the case of a Negative Binomial distribution that allows for over-dispersion[3]. This generalization can be motivated as the result of a mixture of a Poisson distribution with a Gamma distribution for the parameter (λ). Other generalizations have been suggested by Mullahy (1986) and Lambert (1992). The basic idea underlying these extensions is the fact that, in certain situations, it is natural to assume that small values of the integer outcome are generated by one distribution, while larger values of this

[3]Alternatively, some authors have extended the Poisson model by embedding it in a wider family, such as the Katz family of distributions. On count data models, see Gourieroux and Monfort, 1984a and 1984b; Hausman, Hall and Grilliches, 1984; Cameron and Trivedi, 1986; King, 1989; Winkelmann and Zimmermann, 1991; Gurmu and Trivedi, 1992; Dionne and Vanasse, 1992; and Winkelmann, 1994.

integer outcome are generated by a different distribution (which may actually have a different functional form). These authors have specifically focussed on the case of zero being different from the other possible outcomes, but no applications have been found using a strictly positive value at the point where the change of distribution takes place.

In this paper, we will assume that the two disjoint subsets for which we have different generating distributions are fixed and known in the sense that the bound defining the frontier between them is known (in fact, for our application this bound is set at four payments by the bank).

Given the nature of our data set, we are concerned with clients that were accepted by the bank. We want to isolate the significant characteristics influencing the probability for such borrowers to become defaulters before the end of the contract. At the same time, we are interested in taking into account the behavior of the clients by looking at the expected number of non-payments and, consequently, at their conditional expected costs.

The models presented below introduce the fact that there could be two different processes for those clients returning a loan: one for the clients that will never be defaulters, and another one for the clients that have missed enough payments to become defaulters. Usually, the literature in this field has considered the two processes as being the same and therefore has not used truncated models. Altman et alt. (1981) and many others pool all of them together, mainly because their focus is on the decision *prior* to the contract. When discussing the problems in the definition of the groups arising in the classification process for through-the-door applicants, Altman et alt. (1981, pp. 192) suggest that this *posterior* possibility of default is worth being considered. The same authors point out the need to account, *ex-ante*, for the change in the expected costs once the client changes his behavior.

3.1 Extended hurdle specification

Following Mullahy (1986), we define the studied specifications as extended hurdle models. The name *hurdle* was first suggested by Cragg (1971). Its basic feature is that the relative probabilities of two disjoint subsets differ from those that would be implied if only one parent distribution was defined over both of them.

Let us use **I** to denote the set of all non-negative integers. K is a fixed integer, $K \in \mathbf{I}$, that defines $\mathbf{I}_1$ and $\mathbf{I}_2$: [4]

$$\mathbf{I}_1 = \{0, 1, 2, \ldots, (K-1)\}$$

and

$$\mathbf{I}_2 = \{K, (K+1), (K+2), \ldots\}.$$

Let $\phi_1(y;\theta_1)$ and $\phi_2(y;\theta_2)$ be two discrete probability density functions defined on **I**, depending respectively on the parameter vectors θ_1 and θ_2. The observed non-negative outcomes are denoted by y_i, $i = 1, \ldots, n$, where n is the total sample size. We define $\Omega_1 = \{i | y_i \in \mathbf{I}_1\}$ (non-defaulters) and $\Omega_2 = \{i | y_i \in \mathbf{I}_2\}$ (defaulters). n_i is the number of individuals with an observation in either $\mathbf{I}_1$ or $\mathbf{I}_2$ (non-defaulters and defaulters respectively). Moreover,

$$\sum_{i \in \Omega_1} \phi_1(y_i;\theta_1) = \Phi_1(\theta_1), \quad \sum_{i \in \Omega_2} \phi_2(y_i;\theta_2) = \Phi_2(\theta_2),$$

where θ_1 and θ_2 may depend on some fixed covariates. By allowing either

[4] Note that the fact that we only have two different sets is at present a simplification. If there were more sets the same construction would be possible, although it might be necessary to use iterative estimation methods that might be computationally cumbersome. This more general situation will not be discussed here because our dataset contains only two different subsets

$\phi_1(y_i;\theta_1)$ or $\phi_2(y_i;\theta_2)$ (or both) to differ from a discrete probability density function by a proportionality constant, we have that $\Phi_1(\theta_1)+\Phi_2(\theta_2)=1$.

The data generating process in the situation that we will deal with is as follows:

$$\text{Prob}(y_i|y_i\in \mathbf{I}_1)=\begin{cases}\phi_1(y_i;\theta_1)\cdot\Phi_1(\theta_1)^{-1} & \text{when } y_i\in \mathbf{I}_1\\ 0 & \text{otherwise}\end{cases}$$
$$\text{Prob}(y_i\in \mathbf{I}_1)=\Phi_1(\theta_1)$$

and

$$\text{Prob}(y_i|y_i\in \mathbf{I}_2)=\begin{cases}\phi_2(y_i;\theta_2)\cdot\Phi_2(\theta_2)^{-1} & \text{when } y_i\in \mathbf{I}_2\\ 0 & \text{otherwise}\end{cases}$$
$$\text{Prob}(y_i\in \mathbf{I}_2)=\Phi_2(\theta_2).$$

Note that as usually the vector parameter varies from individual to individual, $\theta_{1i}=\exp(X_i\beta_1)$ and $\theta_{2i}=\exp(X_i\beta_2)$, where X_i are observable characteristics. Then, the above distribution may be regarded as a distribution conditional on the set of explanatory variables. We omit the i subscripts of θ_{1i} and θ_{2i} in order to make notation easier.

The likelihood function for this process is:

$$\Lambda^H=\prod_{i\in\Omega_1}\phi_1(y_i;\theta_1)\cdot\Phi_1(\theta_1)^{-1}\cdot\prod_{i\in\Omega_2}\phi_2(y_i;\theta_2)\cdot\Phi_2(\theta_2)^{-1}[1-\Phi_1(\theta_1)]$$

When ϕ_1 and ϕ_2 are not of a particular form, the maximization of the likelihood function may become complicated. Mullahy (1986) chooses ϕ_1 such that the likelihood can be easily maximized separately for each vector parameter, and is therefore very straightforward. Our presentation is similar because our process is also a binary choice between two alternatives (default or not)[5], where K is the first default payment or the threshold. This means

[5]As already pointed out, Mullahy (1986) discusses the case of a process governing the zero outcomes being different from the process for the other possible outcomes.

that the borrower will be considered as defaulter if he or she does not pay K payments or more.

3.2 Negative Binomial hurdle specification

Since we are going to study a process which may contain some unobserved heterogeneity (because some individual characteristics cannot be observed), for convenience we assume that the non-observed component (or the model error) is Gamma distributed. Therefore, we will use Negative Binomial models, which will be tested against Poisson models in the application.

A first model is to consider only one process and estimate a Negative Binomial model for all the data. In this situation, we would use the Type II Negative Binomial model (as referred to in Cameron and Trivedi (1986)). To keep it short, we will omit the words 'Type II' (with ratio of conditional variance to conditional expectation as a linear function in the conditional expectation) because this is the only type of model used throughout this paper.

The Negative Binomial model is as follows:

$$P(Y_i = j) = \frac{\Gamma(j+\nu)}{\Gamma(\nu)\Gamma(y_i+1)} g^{\nu}(1-g)^{y_i}, \quad j = 0, 1, 2, ...$$

with $\nu > 0$, $g = \frac{\nu/\lambda}{1+\nu/\lambda}$. Furthermore, $E(Y_i|X_i) = \lambda_i$ and $Var(Y_i|X_i) = \lambda_i + \alpha\lambda_i^2$, where $\alpha = \frac{1}{\nu}$.

Since it is not evident that the count process for borrowers having a good repayment behavior is similar to the count process of those clients that have already been considered defaulters, we will specify two hurdle models.

Let us first consider the case where we estimate the parameters of the count process in the credit default case ($\mathbf{I_2}$).

We assume that $\text{Prob}(y_i|y_i \in \mathbf{I}_2)$ follows a truncated (from bellow at K) Negative Binomial distribution with mean $\lambda_i = \exp(X_i\beta)$ and variance equal to $\exp(X_i\beta)(1+\alpha)$, where α is the dispersion parameter. When $\alpha = 0$ this reduces to the truncated Poisson distribution, as well known.

The choice of the binary outcome distribution as a Bernouilli with parameter $p = \exp(-\theta_1)$ leads naturally to a logit specification. This is equivalent to taking:

$$\text{Prob}(y_i \in \mathbf{I}_1) = \exp(-\theta_1).$$

Moreover, the probability distribution on $\mathbf{I}_1$ is taken to be proportional to:

$$\text{Prob}(y_i|y_i \in \mathbf{I}_1) \propto \begin{cases} \frac{\exp(-\theta_1)}{K} & \text{when } y_i \in \mathbf{I}_1 \\ 0 & \text{otherwise} \end{cases} .$$

Therefore, the likelihood function is proportional to:

$$\Lambda^H = \prod_{i\in\Omega_1} \exp(-\theta_1) \cdot \prod_{i\in\Omega_2} (1-\exp(-\theta_1)) \cdot \prod_{i\in\Omega_2} \text{Prob}(y_i|y_i \in \mathbf{I}_2). \qquad (3)$$

Λ^H can be regarded as a likelihood function for the binary outcome determining whether an observation belongs to $\mathbf{I}_1$ or $\mathbf{I}_2$ times a likelihood function for a truncated Negative Binomial model on $\mathbf{I}_2$. Furthermore, using the same argument given by Mullahy (1986), the maximum likelihood estimates can be obtained by **separate** maximization of the two parts; i.e. a logit model governing the binary outcome and a truncated from below Negative Binomial model. Standard econometric computer packages, and particularly LIMDEP (Greene, 1992) can be used for the estimation of the parameters and other statistics as the variance-covariance matrix.

Let us now specify a model for the non-default case where the count process is defined on $\mathbf{I}_1$ (the number of unpaid payments for the individuals that are considered non defaulters by the bank). We assume that $\text{Prob}(y_i|y_i \in \mathbf{I}_1)$ follows a truncated (from above at K) Negative Binomial distribution, and that the default probability is $\text{Prob}(y_i \in \mathbf{I}_2) = \exp(-\theta_2)$.

In order to define the probability distribution on $\mathbf{I}_2$, we establish that it is proportional to:

$$\text{Prob}(y_i|y_i \in \mathbf{I}_2) \propto \begin{cases} \frac{\exp(-\theta_2)}{K^*-K+1} & \text{when } y_i \in \mathbf{I}_2^* \subset \mathbf{I}_2 \\ 0 & \text{otherwise} \end{cases},$$

where $\mathbf{I}_2^* = \{K, (K+1), \ldots, K^*\}$, K^* is a large enough positive integer such that $y_i \in (\mathbf{I}_1 \cup \mathbf{I}_2^*) \subset \mathbf{I}$, $i = 1, \ldots, n$. In fact in our application K^* is the maximum observable outcome [6].

Since the two vector parameters can be estimated independently, the likelihood function is like in (3) with the appropriate notational modifications.

Due to the fact that by definition and construction $\Phi_1(\theta_1) + \Phi_2(\theta_2) = 1$, the same estimates for the binary outcome (say the logit model) will be obtained in the two cases. However, since the two processes are different, it is not clear that the same characteristics will be significant in the two count processes. In the case of our application, $\text{Prob}(y_i|y_i \in \mathbf{I}_2)$ is used in the first model as the distribution of the number of non payments for individuals that are considered defaulters. Conversely, $\text{Prob}(y_i|y_i \in \mathbf{I}_1)$ is used in the second model as the distribution of the number of non-payments in the non-default group. Finally, the characteristics that explain the probability of default may differ from those that explain the number of non-payments in each group[7].

[6]But note that there is no need to explicitly define K^* in the model since it turns out to be irrelevant for likelihood maximization.

[7]Note that the presentation offered by Mullahy (1986) is a special case because he defines $\mathbf{I}_1 = \{0\}$. Therefore the probability of obtaining zero equals the probability of obtaining one observation from that subset. Our case is slightly different. When we look at the set $\{K, (K+1), (K+2), \ldots, K^*\}$ we must perform the construction on a finite set in order to be able to define a discrete uniform distribution. In fact, we are only interested in the probability that an observation belongs to $\mathbf{I}_2$.

4 Data and Variables

A Spanish bank provided the data set, having information from clients who had been granted credit. The final sample contains 2,446 observations after the elimination of some individuals with incomplete records and outlier values[8]. We are limited to the study of the probabilities of default for those clients who had already a credit. We think, however, that the methodology introduced here is worth being considered for other applications in this field, including the granting decision when data on refused clients are available. To preserve confidentiality we keep the source anonymous, however our final dataset is available upon request. Finally, individuals who were delinquent for at least 12 months were rejected since they could also be considered as outliers by standard criteria.

The data is a simple random sample for all the bank clients at a given date. It contains information about clients that had obtained a loan for consumption. Mortgages were deliberately excluded because they imply the existence of a substantive collateral that corresponds to a creditor risk exposure that differs from the one considered here.

The type of credit we study is known as personal loan. Personal loans are characterized by small amounts of money being lend, with not necessarily a collateral. Usually, the loan is paid back after a short period of time and is often due monthly with constant payments along the loan period. This kind of credit generally corresponds to consumer loans for the purchase of a car, some durable goods or other similar purposes.

Variables included in the dataset have been collected by the bank from

[8]In order to eliminate values that may distort our results, we have eliminated all the observations having at least one quantitative variable with a value outside the interval $[\bar{x} - 3s, \bar{x} + 3s]$, where $\bar{x}$ stands for the sample mean and s is the standard error.

different sources. Some items were provided in the application form by the individuals applying for credit. The bank always checked the truthfulness of this information. Other characteristics are already in the file of the financial institution. Since we are interested in the behavior of those who are still paying back, we had access to the number of monthly payments that were not paid and should have been paid at data collection time.

There are three groups of variables in the dataset:

a) Personal variables (date of birth, marital status, scholarship,...)

b) Socio-economic variables (net monthly income, housing ownership, geographical location,...)

c) Financial variables (monthly instalment, availability of credit card, amount requested, interest rate,...)

The data contain information about the contractual total length of the credit return period, the number of defaulted payments and the number of months from the beginning of the loan in order to control for different exposures to the risk of default. All records contain individuals who were granted credit and started paying back at least six months before the sample was taken. Fenn (1981) gives a good account of the problem of just considering individuals with the total contractual length is completed, although this was in another context. For modelization purposes, some interacting variables were also considered to be relevant.

We will use the number of non-payments as the dependent variable, and also the binary variable characterizing an individual as being defaulter or not. The bank establishes that an individual is a defaulter whenever this person owes 4 or more monthly payments. The percentage of defaulters in the sample is 13.74 %, according to this criterium. The sample mean and variance of the number of non-payments are 1.109 and 4.860, respectively, suggesting the presence of some overdispersion in the data. Table I presents the frequency table of the dependent variable.

This table means that, for example, 73 individuals owe 3 monthly payments to the bank and represent 3% of the data set. It is interesting to observe, however, that there are 106 individuals owing 4 monthly payments, which suggests that there is a hurdle or, at least, that the process is not a simple Poisson process.

The explanatory variables used in the models are presented in Table A1 in the Appendix. Table A2 presents descriptive statistics for these variables.

(Insert Table 1 here)

Since we are using hurdle models, we will specify two different models: the first one will estimate the individual probability to be in the first group of non-defaulters and their corresponding distribution of non-payments (0, 1, 2 or 3). On the other hand, the second model will estimate the default probability and the count process for the defaulters, i.e. the number of defaults in this latter group (4, 5, 6, ...).

5 Results

The results[9] come from hurdle regression models. The following tables present the maximum likelihood estimates of the parameters for the previous

[9]All the results shown in this paper have been obtained using the econometric software LIMDEP. W. H. Greene has recently implemented the possibility to estimate hurdle models. The variance of the parameters in the Truncated Negative Binomial models were estimated using the BHHH approximation. Both the database and the LIMDEP sources may be provided upon request.

econometric specifications.

Table 2 gives the results of the hurdle models for the non-defaulters. Table 3 gives the Chi squared test statistic. Table 4 presents the results for the hurdle models for the defaulters (4-5-6-...) and Table 5 gives the Chi squared test statistic.

In Table 2, the first model (hurdle 1) presents the logit and the Negative Binomial estimates without covariates, while the second model (hurdle 2) introduces the explanatory variables. As suspected, many significant variables that define the probability of default differ from those that explain the distribution of non-payments. For example, the coefficient of M1 is positive and highly significant to explain the default probability over the complete sample (2,446 observations). This variable is not significant in the estimation of the expected number of non-payment for the non-defaulters (2,110 individuals). This result means that married non-owners with a salary under 3,000$ per month have a higher probability of default than married owners with a salary over 3,000$. On the other hand, the two groups do not differ in explaining the expected number of non-payments and, consequently, the expected administrative costs of good loans.

We obtain a different result for the variable AGE2. Here, being in the group of 25-39 years of age affects the distribution of non-payments, while it does not affect the default probability. Note that the dispersion parameter α is highly significant, which means that the Negative Binomial model does not control all the heterogeneity.

(Insert Table 2 about here)

Finally, the significant variables that explain the default probabilities are common to those obtained in classical studies: age, salary, level of education,...

(Insert Table 3 about here)

Table 3 gives the χ^2 goodness-of-fit statistic and the cumulative frequencies of the two models. It shows that both models fit the data at a confidence level higher than 95%.

(Insert Table 4 about here)

Table 4 presents the model for the non-payments of the defaulters. It shows that the significant parameters of the Truncated Negative Binomial model also differs from the logit parameters. Moreover, it shows that some significant parameters in the Truncated Negative Binomial model (hurdle 4) differ from those in the previous Truncated Negative Binomial model (hurdle 2). For example, AGE1 is now significant to explain the distribution of non-payments. Conversely, ETU2 is not significant to explain the non-payments for defaulters, while it was significant when we estimated the non-defaulters distribution.

The same differences apply to the Z variables NM1, RECSAL and CENTRE. This confirms that the bank has to take into account these results in order to evaluate the individual risks and costs when she makes the decision to allow or not a loan and when she writes the loan contracts.

Another interesting fact is that α is not significant in the hurdle 4 model, which means that we cannot reject the Poisson distribution assumption in this case. It also means that our model controls quite well for the heterogeneity of this smaller group of observations (336 vs. 2,110). These results may also indicate that there is more homogeneity in the default group than in the non-default group.

Table 5 presents the χ^2 goodness-of-fit statistic for the latter models and indicates that they fit the data at reasonable confidence levels.

(Insert Table 5 about here)

Table 6 presents the reclassification results using the logit model with different threshold levels. The percentages of good classification rates are 65.6%, 57.9% and 61.3%, which are percentages similar to those generally reported in the related literature (see, for example, Steenackers and Goovaerts, 1989).

(Insert Table 6 about here)

Finally, a good comprehension of the results is achieved by computing the values of $E_0(c_0)$ and $E_1(c_1)$ for different individuals. For our purpose we define $E_0(c_0) = c_0 E_0(Y)$, $E_1(c_1) = c_1 E_1(Y)$, and we assume that $c_0 = 50\$$ and that $c_1 = 100\$$.

We will compare the expected costs, assuming first that the borrowers are in the non-defaulters group, of two individuals with mean characteristics but belonging to a different age group. Afterwards, we will compare the differences in expected costs for two individuals having characteristics equal to the mean, except for the variable RECSAL (which indicates that they receive their salary in the bank that gave the loan).

Then we are going to see the differences in expected costs when assuming that they are defaulters. We will compare two mean individuals being in different age groups. Two average individuals will be compared by choosing married individuals and house proprietors with different salary levels (i.e. M3=1 or M4=1)

We can now calculate the expected cost using the parameter estimates of the Negative Binomial model for the non-defaulters and the Poisson model for the defaulters[10].

For example, the expected cost of a good-loan individual being in the age group AGE2 (25-39 years) is 39$, while a good-loan individual in the age group AGE3 (40+ years) has an expected cost of 22$. If we assume that we face two average individuals with the only difference being in the reception of the salary variable (RECSAL), then the expected administrative costs for good loans are 38$ if the salary is not received at the bank versus 21$, otherwise.

For the defaulters (or bad loans), if we take two individuals in the previous age groups, the estimated expected costs are much higher but also very different: 489$ for middle-aged and 396$ for older clients. If we consider two married owners, with different levels of salary and all the other characteristics equal to the mean, we obtain the following expected costs: 585$ for lower (M3) and 427$ for higher salaries (M4).

Finally, we give examples illustrating the values of r^* in equation (2). We assume that an individual has average characteristics but receives the salary in the same bank that accepts the loan. If we fix $D = 10,000\$$, $\mu = 9\%$ (5%)[11], $\pi = 40\%$ and we estimate p using the logit results, then $r^* = 12.55\%$ (8.44%). The value is $r^* = 17.44\%$ (13.17%) for the individual that does not receive his salary at the bank.

The corresponding values for two individuals with average characteristics

[10]The truncated Poisson parameter estimates were calculated but are not reported here for brevity.

[11]For matters of comparison we use two values of μ: 9% which represents the lowest interest rate available at the bank during the studied period and 5% a theoretical value to show the effects of different values for μ. The values in parenthesis give the values of r^* corresponding to $\mu = 5\%$. For the bank 9% is optimal when $(1-p) = 0$ and $E_0(c_0) = 0$.

but different ages would be $r^* = 16.03\%$ (11.81%) when the individual is under 25 years (AGE1) and $r^* = 14.61\%$ (10.42%) if he is over 40 years of age (AGE3).

6 Conclusion

In this article we have extended the hurdle models methodology in order to analyse the credit scoring for personal loans in detail and to improve banks profit maximization under asymmetrical information. We have estimated jointly the default probability and two conditional truncated distributions of non-payments of good and bad loans. We have verified that the two conditional distributions do not follow the same distribution and we have shown that the significant variables that affect the three distributions differ. Our application confirms that the methodology increases the degree of flexibility in comparison with standard count models. For example, we did not reject that the (conditional) non-payments distribution of defaulters is a Poisson distribution. This means that the regression component contains enough information to evaluate with precision the effect of different variables on their non-payments while, for the non-defaulters, many non-observable factors still do matter.

Our data limited the analysis. Two extensions would be to introduce information on those who have not received the loan and to consider the past experience of the individuals on the loan market. This last source of information would permit to reduce the heterogeneity problem and to cope with repetitors.

References

Altman, E.I., R.B. Avery, R.A. Eisenbeis and J.F. Sinkey, 1981, Application of Classification Techniques in Business, Banking and Finance, (JAI Press. Greenwich, CT).

Boyd, J.H. and B. D. Smith, 1993, The Equilibrium Allocation of Investment Capital in the Presence of Adverse Selection and Costly State Verification, Economic Theory 3, 427-451.

Boyes, W.J., Hoffman, D.L. and S.A. Low, 1989, An Econometric Analysis of the Bank Credit Scoring Problem, Journal of Econometrics 40, 3-14.

Cameron, A.C. and P.K. Trivedi, 1986, Econometric Models Based on Count Data: Comparison and Application of Some Estimators and Tests, Journal of Applied Econometrics 1, 29-53.

Cragg, J.G., 1971, Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods, Econometrica 39, 829-844.

Crocker, K.J. and A. Snow, 1986, The Efficiency Effects of Categorical Discrimination in the Insurance Industry, Journal of Political Economy 94, 321-344.

Dionne, G. and N. A. Doherty, 1994, Adverse Selection, Commitment and Renegotiation: Extension to and Evidence from Insurance Markets, Journal of Political Economy 102, 209-236.

Dionne, G. and C. Vanasse, 1993, Automobile Insurance Ratemaking in the Presence of Asymmetrical Information, Journal of Applied Econometrics 7, 149-165.

Fenn, P., 1981, Sickness duration, residual disability and income replacement: an empirical analysis, The Economic Journal 91, 158-173.

Fourgeaud, C, Gourieroux, C and J. Pradel, 1990, Selection de clientèle et tarification de prêt bancaire, Working Paper 9004, CEPREMAP.

Gourieroux, C., Monfort, A. and A. Trognon, 1984a, Pseudo-Maximum Likelihood Methods: Theory, Econometrica 52, 681-700.

Gourieroux, C., Monfort, A. and A. Trognon, ,1984b, Pseudo-Maximum Likelihood Methods: Applications to Poisson Models, Econometrica 52, 701-720.

Greene, W. H., 1992, LIMDEP User's Guide. (Econometric Software, New York).

Guillén, M., 1992, Análisis econométrico del credit scoring. Modelos Count Data (Ph. D. Dissertation. University of Barcelona. -in Spanish-).

Gurmu, S. and P.K. Trivedi, 1992, Recent developments in models of event counts: a survey (University of Virginia).

Harris and R. Townsend, 1981, Resource Allocation Under Asymmetrical Information, Econometrica 52, 909-938.

Hausman, J., Hall, B.H. and Z. Grilliches, 1984, Econometric Models for Count Data with an Application to the Patents-R&D Relationship, Econometrica 52, 909-938.

Jaffee, D. M. and T. Russell, 1976, Imperfect Information, Uncertainty and Credit Rationing, Quarterly Journal of Economics, 651-666.

King, G., 1989, Variance Specification in Event Count Models: from Restrictive Assumptions to a Generalized Estimator, American Journal of Political Science 33, 762-784.

Lambert, D., 1992, Zero-inflated Poisson Regression with an Application to Defects in Manufacturing, Technometrics 34, 1, 1-14.

Mullahy, J., 1986 Specification and Testing of Some Modified Count Data Models, Journal of Econometrics 33, 3, 341-366.

Steenackers, A. and M.J. Goovaerts, 1989, A Credit Scoring Model for Personal Loans, Insurance: Mathematics and Economics 8, 31-34.

Stiglitz, J.E. and A. Weiss, 1981, Credit rationing in markets with imperfect information, American Economic Review 71, 393-410.

Winkelmann, R and K.F. Zimmermann, 1991, A new Approach for Modelling Economic Count Data, Economics Letters 37, 139-143.

Winkelmann, R., 1994, Count Data Models (Springer-Verlag, Berlin).

Appendix: Variable Definitions

Table A1. List of variables used in the econometric analysis.

Variable	Description
Y	= Number of non-payments.
YDUM	= { 1 if the number of nonpayments is equal to or greater than 4 0 otherwise.
DT6	= { 1 if total contract duration of return period is more than 4 years 0 otherwise (reference group).
DUREEA	= Number of months from the beginning of the contract
AGE1	= { 1 if the age group is 18-24 years. 0 otherwise.
AGE2	= { 1 if the age group is 25-39 years. 0 otherwise.
AGE3	= { 1 if the age group is 40 years or more. 0 otherwise.
DESTIN	= { 1 if the credit is used to purchase a good with a collateral. 0 otherwise.
ETU1	= { 1 if the client has not completed primary education. 0 otherwise.
ETU2	= { 1 if the client has completed primary education. 0 otherwise.
ETU3	= { 1 if the client has completed higher education. 0 otherwise.
ETU4	= { 1 if the client has a university degree. 0 otherwise.
RECSAL	= { 1 if the client receives the salary through the bank. 0 otherwise.

(...)

Table A1. (Continued)

M1	=	1 if married, non-owner, salary under 3,000$. 0 otherwise.
M2	=	1 if married, non-owner, salary higher than (equal to) 3,000$. 0 otherwise.
M3	=	1 if married, owner, salary under 3,000$. 0 otherwise.
M4	=	1 if married, owner, salary higher than (equal to) 3,000$. 0 otherwise.
NM1	=	1 if not married, non-owner. 0 otherwise.
NM2	=	1 if not married, owner. 0 otherwise.
CENTRE	=	1 if the credit is granted by a store. 0 otherwise.
RESID	=	1 if resident in the city for at least 4 years. 0 otherwise.
Z1	=	1 if south (Andalucía, Canarias, Castilla-La Mancha, Extremadura, Murcia). 0 otherwise.
Z2	=	1 if north (Aragon, Asturias, Cantabria, Castilla-León, Galicia, Navarra, País Vasco). 0 otherwise.
Z3	=	1 if east (Baleares, Catalunya, Valencia). 0 otherwise.
Z4	=	1 if center (Madrid). 0 otherwise.

Table A2. Sample statistics.

Variable	Mean	Variance
DT6	0.335	0.223
DUREEA	18.812	117.69
AGE1	0.087	0.080
AGE2	0.483	0.250
AGE3	0.439	0.245
DESTIN	0.453	0.248
ETU1	0.043	0.041
ETU2	0.486	0.250
ETU3	0.320	0.218
ETU4	0.152	0.129
RECSAL	0.403	0.241
M1	0.206	0.164
M2	0.020	0.019
M3	0.079	0.073
M4	0.695	0.212
NM1	0.219	0.171
NM2	0.781	0.171
CENTRE	0.158	0.133
RESID	0.738	0.193
Z1	0.282	0.202
Z2	0.292	0.207
Z3	0.297	0.209
Z4	0.129	0.113

Table I. Frequency table of the number of non-payments.

j	Frequency	Percent
0	1665	68.1
1	271	11.1
2	101	4.1
3	73	3.0
4	106	4.3
5	72	2.9
6	43	1.8
7	31	1.3
8	31	1.3
9	25	1.0
10	19	0.8
11	9	0.4

Table 2. Negative Binomial hurdle model for non-defaulters.

Variables	hurdle 1 (Logit)	hurdle 1 (BNTron)	hurdle 2 (Logit)	hurdle 2 (BNTron)
Constant	-1.8373	-0.48086	-2.305	-0.794
	(-31.280)	(-2.425)	(-7.791)	(-1.550)
DT6	-	-	0.516	0.244
			(4.045)	(0.950)
DUREEA	-	-	0.028	0.009
			(4.879)	(0.740)
AGE1	-	-	0.223	0.200
			(0.893)	(0.452)
AGE2	-	-	0.061	0.561
			(0.443)	(2.100)
AGE3	-	-	-	-
DESTIN	-	-	-0.573	-1.156
			(-4.364)	(-4.055)
ETU1	-	-	0.761	-0.016
			(2.500)	(-0.030)
ETU2	-	-	0.374	0.823
			(1.866)	(2.347)
ETU3	-	-	0.231	0.179
			(1.075)	(0.563)
ETU4	-	-	-	-
RECSAL	-	-	-0.932	-0.611
			(-6.540)	(-2.419)
M1	-	-	0.619	0.404
			(3.948)	(1.325)
M2	-	-	0.486	-0.167
			(1.119)	(-0.214)
M3	-	-	0.509	1.133
			(2.205)	(2.282)
M4	-	-	-	-
NM1	-	-	0.348	0.504
			(2.004)	(1.658)
NM2	-	-	-	-
CENTRE	-	-	0.038	-0.845
			(0.207)	(-2.684)

(...)

Table 2. (Continued)

Variables	hurdle 1 (Logit)	hurdle 1 (BNTron)	hurdle 2 (Logit)	hurdle 2 (BNTron)
RESID	-	-	-0.250	-0.198
			(-1.811)	(-0.672)
Z1	-	-	-	-
Z2	-	-	-0.086	0.187
			(-0.568)	(0.609)
Z3	-	-	-0.425	0.245
			(-2.584)	(0.800)
Z4	-	-	-0.646	-0.633
			(-2.569)	(-1.809)
Alpha	-	4.7566	-	4.089
		(7.796)		(8.017)
No. parameters	1	2	19	20
No. observations	2,446	2,110	2,446	2,110
Log-likelihood	-978.8	-1505.1	-896.4	-1466.7
Total Log-likelihood	-2483.9		-2363.1	

Table 3. Negative Binomial hurdle model for non-defaulters. χ^2 statistic [12].

Value	Observed freq.	hurdle 1	hurdle 2
0	1665	1665.80	1665.08
1	271	261.34	264.52
2	101	118.01	115.93
3	73	64.88	64.47
4+	336	335.99	336.00
Chi-2		3.83	3.21
P-value (4 d.f.)		0.430	0.523

Cumulative frequencies

Value	Observed freq.	hurdle 1	hurdle 2
0	1665	1665.80	1665.08
1	1936	1927.13	1929.60
2	2037	2045.14	2045.53
3	2110	2110.02	2110.00
4+	2446	2446.01	2446.00

[12] The Chi squared statistic is computed using the following expression:

$$\widehat{\chi^2} = \sum_k \frac{\left(N_k - \sum_i \hat{p_i(k)}\right)^2}{\sum_i \hat{p_i(k)}}$$

where $\hat{p_i(k)}$ is the predicted probability for individual i to have k unpaid installments, N_k is the observed absolute frequency for cell k. Under H_0: predicted frequencies = observed frequencies. This statistic will be compared with a χ^2 distribution with as many degrees of freedom as the number of cells considered minus one.

Table 4. Negative Binomial hurdle model for defaulters.

Variables	hurdle 3 (Logit)	hurdle 3 (BNTron)	hurdle 4 (Logit)	hurdle 4 (BNTron)
Constant	-1.8373	1.4208	-2.305	1.487
	(-31.280)	(11.418)	(-7.791)	(6.982)
DT6	-	-	0.516	-0.102
			(4.045)	(-1.165)
DUREEA	-	-	0.028	0.007
			(4.879)	(1.515)
AGE1	-	-	0.223	0.273
			(0.893)	(1.961)
AGE2	-	-	0.061	0.231
			(0.443)	(2.489)
AGE3	-	-	-	-
DESTIN	-	-	-0.573	-0.234
			(-4.364)	(-2.481)
ETU1	-	-	0.761	0.065
			(2.500)	(0.339)
ETU2	-	-	0.374	0.054
			(1.866)	(0.424)
ETU3	-	-	0.231	-0.170
			(1.075)	(-1.168)
ETU4	-	-	-	-
RECSAL	-	-	-0.932	-0.139
			(-6.540)	(-1.530)
M1	-	-	0.619	0.103
			(3.948)	(1.023)
M2	-	-	0.486	0.070
			(1.119)	(0.122)
M3	-	-	0.509	0.341
			(2.205)	(2.330)
M4	-	-	-	-
NM1	-	-	0.348	0.040
			(2.004)	(0.392)
NM2	-	-	-	-
CENTRE	-	-	0.038	0.017
			(0.207)	(0.147)

(...)

Table 4. (Continued)

Variables	hurdle 3 (Logit)	hurdle 3 (BNTron)	hurdle 4 (Logit)	hurdle 4 (BNTron)
RESID	-	-	-0.250	-0.045
			(-1.811)	(-0.532)
Z1	-	-	-	-
Z2	-	-	-0.086	-0.180
			(-0.568)	(-1.980)
Z3	-	-	-0.425	-0.278
			(-2.584)	(-2.292)
Z4	-	-	-0.646	0.049
			(-2.569)	(0.313)
Alpha	-	0.16248	-	0.049
		(1.392)		(0.907)
No. parameters	1	2	19	20
No. observations	2,446	336	2,446	336
Log-likelihood	-978.8	-638.9	-896.4	-617.8
Total Log-likelihood	-1617.7		-1514.2	

Table 5. Negative Binomial hurdle model for defaulters. Chi-2 statistic.

Value	Observed freq.	hurdle 1	hurdle 2
0-1-2-3	2110	2110.03	2110.00
4	106	94.14	92.20
5	72	76.89	76.99
6	43	57.49	58.08
7	31	40.14	40.62
8	31	26.55	26.81
9	25	16.79	16.92
10+	28	24.00	24.38
Chi-2		12.97	13.64
P-value (7 d.f.)*		0.073	0.058

* H_0 is not rejected at the 95% confidence level

Cumulative frequencies

Value	Observed freq.	hurdle 1	hurdle 2
0-1-2-3	2110	2110.03	2110.00
4	2216	2204.17	2202.20
5	2288	2281.06	2279.19
6	2331	2338.54	2337.27
7	2362	2378.69	2377.89
8	2393	2405.24	2404.70
9	2418	2422.03	2421.62
10+	2446	2446.03	2446.00

Table 6. Reclassification results.

Threshold equal to the frequency of defaulters			
		Predicted	
		Good Non-default	Bad Default
Observed	Good	1376	734
	Bad	107	229

Threshold equal to the estimated probability of default for the mean individual			
		Predicted	
		Good Non-default	Bad Default
Observed	Good	1152	958
	Bad	72	264

Threshold equal to 0.125, as suggested by the bank managers			
		Predicted	
		Good Non-default	Bad Default
Observed	Good	1258	852
	Bad	88	248

Université de Montréal
Département de sciences économiques
Centre de documentation
C.P. 6128, succursale Centre-ville
Montréal (Québec)
H3C 3J7

Cahiers de recherche (Discussion papers)
1994 à aujourd'hui (1994 to date)

Si vous désirez obtenir un exemplaire, vous n'avez qu'à faire parvenir votre demande et votre paiement (5 $ l'unité) à l'adresse ci-haut mentionnée. / To obtain a copy ($ 5 each), please send your request and prepayment to the above-mentioned address.

9401 : Mercenier, Jean et Bernardin Akitoby, "On Intertemporal General-Equilibrium Reallocation Effects of Europe's Move to a Single Market", janvier 1994, 41 pages.

9402 : Gauthier, Céline et Michel Poitevin, "Using Ex Ante Payments in Self-Enforcing Risk-Sharing Contracts", février 1994, 38 pages.

9403 : Ghysels, Eric et Joanna Jasiak, "Stochastic Volatility and Time Deformation : an Application of Trading Volume and Leverage Effects", février 1994, 37 pages.

9404 : Dagenais, Marcel G. et Denyse L. Dagenais, "GMM Estimators for Linear Regression Models with Errors in the Variables", avril 1994, 33 pages.

9405 : Bronsard, C., Fabienne Rosenwald et Lise Salvas-Bronsard, "Evidence on Corporate Private Debt Finance and the Term Structure of Interest Rates", avril 1994, 42 pages.

9406 : Dinardo, John, Nicole M. Fortin et Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992 : A Semiparametric Approach", avril 1994, 73 pages.

9407 : Campbell, Bryan et Jean-Marie Dufour, "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter", avril 1994, 32 pages.

9408 : Bollerslev, Tim et Eric Ghysels, "Periodic Autoregressive Conditional Heteroskedasticity", mai 1994, 29 pages.

9409 : Cardia, Emanuela, "The Effects of Government Financial Policies : Can We Assume Ricardian Equivalence?", mai 1994, 42 pages.

9410 : Kollmann, Robert, "Hidden Unemployment : A Search Theoretic Interpretation", mai 1994, 9 pages.

9411 : Kollmann, Robert, "The Correlation of Productivity Growth Across Regions and Industries in the US", juin 1994, 14 pages.

9412 : Gaudry, Marc, Benedikt Mandel et Werner Rothengatter, "Introducing Spatial Competition through an Autoregressive Contiguous Distributed (AR-C-D) Process in Intercity Generation-Distribution Models within a Quasi-Direct Format (QDF)", juin 1994, 64 pages.

9413 : Gaudry, Marc et Alexandre Le Leyzour, "Improving a Fragile Linear Logit Model Specified for High Speed Rail Demand Analysis in the Quebec-Windsor Corridor of Canada", août 1994, 39 pages.

9414 : Lewis, Tracy et Michel Poitevin, "Disclosure of Information in Regulatory Proceedings", juillet 1994, 38 pages.

9415 : Ambler, Steve, Emanuela Cardia et Jeannine Farazli, "Export Promotion and Growth", août 1994, 41 pages.

9416 : Ghysels, Eric et Haldun Sarlan, "On the Analysis of Business Cycles Through the Spectrum of Chronologies", août 1994, 37 pages.

9417 : Martel, Jocelyn et Timothy C.G. Fisher, "The Creditors' Financial Reorganization Decision : New Evidence from Canadian Data", août 1994, 21 pages.

9418 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Entrance Quotas and Admission to Medical Schools : A Sequential Probit Model", septembre 1994, 26 pages.

9419 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Major Choices : Undergraduate Concentrations and the Probability of Graduation", septembre 1994, 26 pages.

9420 : Nabeya, Seiji et Pierre Perron, "Approximations to Some Exact Distributions in the First Order Autoregressive Model with Dependent Errors", septembre 1994, 40 pages.

9421 : Perron, Pierre, "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", octobre 1994, 50 pages.

9422 : Vogelsang, Timothy J. et Pierre Perron, "Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time", novembre 1994, 57 pages.

9423 : Ng, Serena et Pierre Perron, "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag", décembre 1994, 41 pages.

9424 : Perron, Pierre, "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors", décembre 1994, 37 pages.

9425 : Ghysels, Eric et Pierre Perron, "The Effect of Linear Filters on Dynamic Time Series with Structural Change", décembre 1994, 35 pages.

9426 : Boyer, Marcel, Jean-Jacques Laffont, Philippe Mahenc et Michel Moreaux, "Sequential Location Equilibria Under Incomplete Information", décembre 1994, 38 pages.

9427 : Perron, Pierre et Serena NG, "Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties", décembre 1994, 41 pages.

9428 : Garcia, René et Pierre Perron, "An Analysis of the Real Interest Rate Under Regime Shifts", décembre 1994, 42 pages.

9501 : Boyer, Marcel et Jean-Jacques Laffont, "Environmental Risks and Bank Liability", janvier 1995, 46 pages.

9502 : Margolis, David. N., "Firm Heterogeneity and Worker Self-Selection Bias Estimated Returns to Seniority", décembre 1994, 29 pages.
9503 : Abowd, John M., Francis Kramarz et David N. Margolis, "High-Wage Workers and High-Wage Firms", janvier 1995, 73 pages
9504 : Cardia, Emanuela et Steve Ambler, "Indexation Lags and Heterodox Stabilization Programs", janvier 1995, 29 pages.
9605 : Garcia, René et Huntley Schaller, "Are the Effects of Monetary Policy Asymmetric?", février, 42 pages.
9506 : Parent, Daniel, "Survol des contributions théoriques et empiriques liées au capital humain", février 1995, 70 pages.
9507 : Parent, Daniel, "Wages and Mobility : The Impact of Employer-Provided Training", février 1995, 34 pages.
9508 : Parent, Daniel, "Industry-Specific Capital and the Wage Profile : Evidence from the NLSY and the PSID", février 1995, 21 pages.
9509 : Parent, Daniel, "Matching, Human Capital, and the Covariance Structure of Earnings", février 1995, 54 pages.
9510 : Garcia, René, "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", mars 1995, 50 pages.
9511 : Garcia, René, Annamaria Lusardi and Serena Ng, "Excess Sensivity and Asymmetries in Consumption : An Empirical Investigation", mars 1995, 26 pages.
9512 : Sprumont, Yves, "An Axiomatization of the Pazner-Schmeidler Rules in Large Fair Division Problems", mars 1995, 26 pages.
9513 : Ghysels, Eric, Lynda Khalaf and Cosmé Vodounou, "Simulation Based Inference in Moving Average Models", mars 1995, 10 pages.
9514 : Ng, Serena, "Looking for Evidence of Speculative Stockholding in Commodity Markets", mars 1995, 25 pages.
9515 : Ng, Serena and Huntley Schaller, "The Risky Spread, Investment, and Monetary Policy Transmission: Evidence on the Role of Asymmetric Information", mars 1995, 26 pages.
9516 : Ng, Serena, "Testing for Homogeneity in Demand Systems when the Regressors are Non-Stationary", mars 1995, 26 pages.
9517 : Ghysels, Eric, Clive W.J. Granger and Pierre L. Siklos, "Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?", mars 1995, 34 pages.
9518 : Ghysels, Eric, Alastair Hall and Hahn S. Lee, "On Periodic Structures and Testing for Seasonal Unit Roots", mars 1995, 45 pages.
9519 : Sprumont, Yves, "On the Game-Theoretic Structure of Public-Good Economies", mars 1995, 21 pages.
9520 : Charles, Sandra, François Vaillancourt and Nicolas Marceau, "The Impact of Decentralization on Growth and Democracy: A Note", 13 pages.
9521 : Sprumont, Yves, "Balanced Egalitarian Redistribution of Income", 17 pages.
9522 : Bronsard, Camille, Lise Salvas-Bronsard and Alain Trognon, "On the Residual Dynamics Implied by the Rational Expectations Hypothesis", mars 1995, 18 pages.

9523 : Campbell, Bryan and Eric Ghysels, "An Empirical Analysis of the Canadian Budget Process", mars 1995, 30 pages.

9524 : Ghysels, Eric, Alain Guay and Alastair Hall, "Predictive Tests for Structural Change with Unknown Breakpoint", avril 1995, 29 pages.

9525 : Ghysels, Eric, "On Stable Factor Structures in the Pricing of Risk", avril 1995, 37 pages.

9526 : Kollmann, Robert, "Mark Up Fluctuations in U.S. Manufacturing and Trade: Empirical Evidence Based on A Model of Optimal Storage", avril 1995, 18 pages.

9527 : à venir

9528 : Dionne, G., M. Artis and M. Guillen, "On the Repayment of Personal Loans Under Asymmetrical Information: A Count Data Model Approach", mai 1995, 32 pages.