UNIVERSITÉ DE STRASBOURG

Numéro :

# DOCTORAT DE L'UNIVERSITÉ

# DE STRASBOURG

## RAPPORT de PRESENTATION de la THÈSE

FORMULAIRE A RENVOYER OBLIGATOIREMENT AVEC LE RAPPORT
15 jours avant la soutenance sous peine de report de celle-ci

Nom du candidat : **M. Gilles STUPFLER**

**Titre de la thèse** : Un modèle de Markov caché en assurance et Estimation de frontière et de point terminal

Composition du jury de soutenance :

Directeur(s) de thèse :
Mme A. GUILLOU
M. S. GIRARD
Rapporteur(s) :

**Mme I. GIJBELS**
M. H. ALBRECHER
Autre(s) membre(s) du jury :
M. S. LOISEL
M. J-N. BACRO
M. L. GARDES

Rapport de **Mme I. GIJBELS**

**1° ÉVALUATION GÉNÉRALE**

Par comparaison avec des thèses de Doctorat récentes soutenues dans votre Université, ou dont vous avez eu connaissance personnellement, cette thèse est-elle à votre avis digne d'être soutenue en vue du Doctorat ?

**OUI** [X] **NON** [ ] (*)

(*)le refus d'attribution du grade de docteur devra toujours faire l'objet d'un rapport circonstancié.

Dans l'affirmative et avant soutenance, cette thèse est-elle d'un niveau scientifique :

**SATISFAISANT [ ]** **BON [ ]** **TRÈS BON [ ]** **EXCEPTIONNEL [X]**

*Date :* le 26 octobre
*Signature du Rapporteur :*

Gijbels

*Vu et autorisation de soutenance.*
*Le Président de l'Université*

## 2° CONTENU SCIENTIFIQUE

Veuillez commenter les résultats qui, dans le mémoire, vous paraissent :

- d'un intérêt scientifique particulier, justifiant la soutenance,
- mériter une discussion lors de la soutenance,
- exiger une révision ou des compléments avant la soutenance,
- trop éloignés de votre compétence pour que vous puissiez donner un avis valable.

see the enclosed report please.

# Evaluation report concerning the thesis "Un modèle de Markov caché en assurance et estimation de frontière et de point terminal" of Mr Gilles Stupfler

**Evaluation on the scientific contribution**

The thesis consists of two parts: in part I estimation of the parameters of a Markov-modulated loss process in insurance is discussed; and in part II the focus is on estimating endpoints and frontier functions.

The thesis contributes thus in two quite distinct areas. The thesis is very well written, with very clear explanations and an excellent presentation of the material. It is an impressive piece of theoretical work, accompanied with nice motivations, attention for estimation algorithms, implementation issues as well as finite-sample performances.

The excellent quality of writing is already clear from the start, from the introductory chapter. This chapter explains the reader step by step why a Markov-modulated loss process is of interest, in particular in the insurance context. In subsequent steps it discusses homogeneous Poisson processes, (possible inhomogeneous) Poisson processes, (homogeneous) Markov chains, (homogeneous) Markov processes, Markov-modulated Poisson processes, and hidden Markov models. For each of these processes it is very clearly explained what are the properties and what is possibly lacking to get to an appropriate modeling of losses in insurance.

This excellent quality of presentation is continued when explaining the context of extreme value statistics, the issue of endpoint estimation and the related frontier estimation problem. Such an excellent exposition is already an important indication of an excellent candidate.

Chapter 1 deals with the estimation of the parameters in a Markov-modulated loss process in insurance. All elements of the considered process are clearly described, and the available data are discussed, as well as where are the extra difficulties coming in. For this part the thesis contributions consist of: (i) providing the likelihood expression and finding the estimators; (ii) providing an EM-algorithm for the computation of the maximum likelihood estimators; (iii) establishing strong consistency of the maximum likelihood estimators; (iv) numerical illustrations for the method involving a non-life insurance example and a life insurance example. The asymptotic normality result for the maximum likelihood estimators is also established by the candidate (and co-authors) but this result is not included in the thesis (due to its high-technical level). All together this is a very complete and thorough treatment of the statistical inference for an important process. A remarkable fact is that all aspects of the problem are very well studied: motivation of the process, estimation and quality of estimation, algorithm and implementation and numerical illustrations. It is rather seldom to see each of these aspects treated in such a nice way. Another remarkable fact is that the applicant objectively comments on possible limitations. This part I constitutes thus a very original and nice contribution to the area of modeling of loss processes.

Chapters 2 and 3 are related and concern the estimation of endpoints. In frontier estimation the endpoint is the upper boundary of the support of a bivariate density (in the univariate covariate case). Here the contribution consists of introducing a new endpoint estimator that remedies the drawback of the high order moments estimator of Girard and Jacob (2003). The latter estimator has the drawback that an asymptotic Gaussian distribution of the estimator only holds under the assumption that the conditional distribution of $Y$ given $X$ is uniform. The new high order moment estimator is introduced first in the univariate setting and then in the multivariate setting. The latter extension is rather straightforward (as admitted).

Chapter 2 deals with the new endpoint estimator in general (in the non-regression context), and in a first stage contributions (consistency in probability and asymptotic normality) are provided for the case of positive random variables. In particular the proof of the asymptotic normality result is quite involved. The thesis also nicely discusses the special case of the Hall model, and the focus in the discussion is on the differences between the general setup and the special set up in the Hall model. The contribution then goes on with the treatment of the case of a general r.v. $X$ (i.e. dropping the positiveness of $X$). This is done in a simple way by just, for general $X$, considering the positive random variable $Y = e^X$. The higher order moment estimator is discussed and its consistency in probability and asymptotic normality are proven, with also for this general case a comparison with the situation encountered in the Hall model. Another contribution consists of a discussion on optimality of the convergence rate, showing that the convergence rate of the proposed high order moment estimator is essentially the same as that of the extreme-value moment estimator of Aarssen and de Haan (1994). The chapter concludes by a simulation study that involves also a comparison with two available estimators in the literature: the (naive) maximum endpoint estimator and the extreme-value estimator of Aarssen and de Haan (1994). Simulations are done for a sample size $n = 500$ which seems rather large. But even for this already large sample size the differences between the three estimators are still visible. For the high order moment estimator a choice of the parameter $a$ and the sequence $p_n$ should be made. In the simulation study $p_n = n^{1/a}/\ln\ln n$ is made and a set of plausible $a$ values is considered. Among the presented results are these for the 'best' choices of $a$. The sensitivity for the parameter $a$ is illustrated in figures. There is less information on the sensitivity to the choice of the sequence $p_n$. It would have been interesting to see simulation results for smaller sample sizes. Remarkable on this chapter is also that all aspects of the studied problem are treated: the theoretical performance, a discussion on the assumptions made, and the finite-sample performance of the proposed method.

Chapter 3 then looks into the estimation of the frontier function, and introduces in this regression context the high order moments estimator, as such dealing with the drawback of the frontier estimator of Girard and Jacob (2003). Since this is a regression context, the estimator now involves the choices of a bandwidth parameter $h_n$ as well as the parameters $a$ and $p_n$. The chapter starts by showing that the estimator of Girard and Jacob (2003) does not lead to an asymptotic Gaussian distribution when the conditional distribution of $Y$ given $X$ is not uniform. The estimator in this context of covariates is based on nonparametric kernel estimation of mean regression functions and as such also suffers from a curse of dimensionality. Consistency in probability and the asymptotic normality of the proposed high order moment frontier estimator is established under some assumptions. In frontier estimation it is also of interest to consider situations in which the joint density shows a different behaviour close to the frontier (staying away from zero, tending to zero or tending to infinity). The simulation study involves illustrations of the finite-sample performance of the method for the univariate and bivariate case. Sample sizes are respectively $n = 500$ and $n = 1000$ in the case $d = 1$ and $d = 2$ respectively. Again these are rather large sample sizes. Practical choices of the various parameters are briefly discussed, but limited to choices that satisfy the theoretical assumptions complemented with ad hoc choices of constants involved. It would have been interesting to see further discussion on important issues such as: the assumptions made on the joint density in this frontier estimation context, the curse of dimensionality problem, sensitivity to the choices of $a$, $p_n$ and $h_n$ in this context, and the use of the method in real data analysis. Also this chapter contains an original and very nice contribution to the area of frontier estimation.

**Conclusive remarks**

This is a thesis of a very good to excellent scientific level. It provides nice and important contributions to two distinct areas: modeling loss processes in insurance and estimation of endpoints and frontier functions. Remarkable facts about the thesis are: (i) its excellent level of presentation which makes its reading a real scientific enjoyment, (ii) its thorough study of the problems considered including important theoretical contributions requiring a solid knowledge of several advanced mathematical techniques, as well as attention to practical implementation and finite-sample performance. Last but not least the studied problems as well as the approaches taken are very well motivated, and hence the candidate shows a remarkable level of scientific maturity (also illustrated in the discussions).

In summary, this is an excellent piece of doctoral research, and therefore a positive advice is given for the defense of this work.

Gijbels

Irène Gijbels
Full Professor in Statistics
Katholieke Universiteit Leuven
Department of Mathematics and
Leuven Statistics Research Centre
Belgium

Leuven, October 26, 2011.