Endogeneity & Instrumental Variables in Dynamic Processes
Inverse Problems in Finance

Guillaume Simon
Université de Toulouse

May 2011

Contents

Remerciements

Le processus doctoral[1] est un cheminement tortueux et éreintant sur les sentiers scientifiques. Ce long périple n'apprend pas toujours à marcher et en pratique, l'attrait pour la recherche met rarement fin à ces tribulations. Mais ce chemin permet avant tout de mieux se connaître et de multiplier les rencontres enrichissantes. Au moment de divulguer mon carnet de voyage, je voudrais ici témoigner de ces rencontres et de ce qu'elles m'ont apporté pour exprimer ma gratitude à ceux m'ayant permis d'en arriver jusque là.

La première personne que je souhaiterais ici remercier est Jean-Pierre Florens, qui aura su me guider jusqu'à des contrées intellectuelles qu'il me semblait impensable de pouvoir visiter il y a encore quatre ans. Son enthousiasme pour l'étude de problèmes complexes m'aura permis de m'approcher au plus près de ceux-ci ; sa curiosité perpétuelle et son redoutable instinct scientifique nous aurons permis d'en apprivoiser de nouveaux. Les semaines où je "descendais à Toulouse" resteront pour moi des moments privilégiés, d'apprentissage et de réflexion, toujours dans un climat d'échange et de pleine collaboration. Outre la satisfaction scientifique du travail bien fait, je retiendrai avant tout la satisfaction personnelle des conditions humaines exceptionnelles dans lesquelles ce travail a été mené, et pour lesquelles je souhaiterais très chaleureusement remercier Jean-Pierre. En espérant que ces quelques lignes traduiront fidèlement la reconnaissance que je cherche à exprimer.

"Toute marche de 10000km commence par un pas." Beaucoup de pas (et surtout le premier) auront été possibles grâce à la ténacité de Serge Darolles. Sans lui ce doctorat n'aurait certainement jamais eu de début. Son action et ses idées auront permis d'initier cette thèse et de lui donner corps. Je veux le remercier pour sa perpétuelle confiance et l'intérêt qu'il a su me transmettre pour nos sujets de recherche. Durant sept ans (!) Serge m'aura toujours permis de développer de nouvelles idées, tout en mettant en valeur mes initiatives. De la tour Total aux couloirs de l'Ensae, du palais Brongniart au boulevard Haussmann, notre binôme se sera tacitement mué en VRP de l'enseignement de l'économétrie de la finance. L'entente amicale qui en a résulté est le meilleur témoignage de cette complicité.

Je souhaiterais aussi remercier (dans l'ordre alphabétique) les rapporteurs et membres du jury René Garcia, Marc Hoffmann, Nour Meddahi, et Eric Renault, qui ont accepté de lire ce manuscrit tout en me donnant conseils et avis. La satisfaction qui est la mienne à l'heure de conclure cette aventure tient en grande partie à leur présence, leurs commentaires et leur attention pour mon travail.

[1] May the English readers forgive me: this part is far better in French !

J'adresse mes gages de profonde gratitude à Anna Simoni et Florian Ielpo, actifs co-auteurs lors de ces années de doctorat. Travailler avec eux aura été une joie et une vraie satisfaction, grâce à leurs qualités humaines et leur intuition scientifique. Avec deux styles différents, je sais ce que je leur dois. Je leur souhaite pleine réussite dans leurs vies académiques et professionnelles respectives, en espérant que je n'aurai pas été un acolyte trop encombrant et que nous aurons l'opportunité de poursuivre nos collaborations.

Ma sincère reconnaissance va également à Marie-Hélène Dufour et Aude Schloesing, et plus largement toute l'équipe administrative de l'IDEI et de l'école doctorale. À de nombreuses reprises leur action et leur efficacité ont évité à mon parcours de rencontrer ornières et obstacles. Toujours avec disponibilité, gentillesse et affabilité, elles ont su faciliter toutes mes démarches et venues. Qu'elles en soient ici chaleureusement remerciées.

Ce doctorat s'est déroulé dans le cadre d'une thèse CIFRE en partenariat avec Société Générale Asset Management, devenue Lyxor. Cette période aura été extrêmement riche en évènements, expériences et sujets. Mais ce sont probablement les rencontres humaines qui m'auront le plus marqué. Je souhaiterais ainsi remercier Emmanuelle Jay, Isabelle Sérot, Thierry Béchu, Mabrouk Chetouane, Patrick Ducharme, Kirill Evlampiev, Abdul Koudiraty, Stéphane Martinetti, Christian Masciulli, Patrick Merheb, Angelo Pessaris, Vincent Pouderoux, et Thierry Roncalli pour leur bonne humeur, leur soutien, et leur sympathie. L'aide patiente et attentive d'Emmanuelle (qui était là au tout début !) et la compréhension de Thierry sur mes derniers mois de thèse auront été des éléments importants. La particularité de la période financière récente aura au moins eu le mérite de me montrer combien le recrutement de cette équipe était pertinent, homogène et fait de personnalités originales. Je salue ici amicalement Mabrouk, témoin privilégié de cette époque, convive recherché de déjeuners hauts en couleur et compagnon d'infortune dans le monde impitoyable des thésards CIFRE. Je ne peux enfin que remercier de manière appuyée Angelo dont l'aide décisive a permis quelques petits miracles. Travailler avec lui aura été un réel plaisir qui je l'espère sera renouvelé un jour.

Dès mars 2010 et pendant quelques mois, j'ai aussi eu la chance et la joie de faire un peu partie de la joyeuse équipe du CREST. Ceci aura notamment été possible grâce à Christian Gouriéroux, Gaëlle LeFol et Jean-Michel Zakoian. Je souhaiterais les remercier vivement pour leur accueil, mais aussi pour leurs conseils et leur intérêt pour mes travaux. Cette période n'aurait pas été si productive et agréable sans l'amitié et l'aide de Gulten Mero et Guillaume Lepage. Le quotidien partagé avec eux pendant quelques mois ne laissera chez moi que des souvenirs simples, heureux et authentiques. La bonne humeur de Céline, Giorgia, Fanda, Nadine, Sophie, Julien, Guy et Linas restera de même une évocation des plus douces.

C'est aussi pour moi l'occasion de remercier l'ENSAE, institution, professeurs et assistants. Les enseignements suivis m'auront donné le goût pour parfaire mon apprentissage, et permis d'atteindre un niveau académique suffisant pour envisager un doctorat. Le souvenir et l'attachement à mon école restera fort, et je souhaite légitimement en témoigner ici. Plus personnellement, je salue amicalement, scientifiquement (et sportivement) Mathieu Rosenbaum, Romuald Elie, Xavier d'Haultfœuille, Johan Hombert et Christian-Yann Robert. Leur aide, leurs conseils et les discussions (pas toutes sérieuses certes) menées avec eux auront été bien souvent des plus éclairantes. Je remercie de plus Yves Duval pour m'avoir su me donner le goût des mathématiques et de de la topologie. Son enseigenement, sa passion et son intégrité auront été des plus motivants dans mon parcours académique. Je souhaiterais

enfin remercier ici Anne-Marie Lebufnoir, qui aura été là au tout début, pour avoir su avec bienveillance me donner goût du travail bien fait et confiance dans mes capacités.

Alors même que ce manuscrit n'était pas fini, Jean-Philippe Bouchaud, Laurent Laloux et Marc Potters m'ont donné ma chance au sein de Capital Fund Management. Je tiendrai donc à les en remercier, ainsi que toute l'équipe Ventus, Alexandre, Bruno, Emmanuel, Gilles et Peter, pour les conditions de travail exceptionnelles qui sont les miennes désormais. Travailler dans un environnement aussi porteur, sympathique et stimulant est une récompense rare que je sais chaque jour apprécier.

Je tiens enfin à exprimer avec toute la chaleur possible, mes remerciements les plus affectueux à tous mes proches. Mes parents Catherine et Lionel (jamais deux thésards sans trois) auront donné beaucoup pour que leurs enfants fassent ce qu'ils avaient envie de réaliser. Les associer à la conclusion de mon doctorat est une de mes plus grandes fiertés. J'ai tenté, du mieux que je pouvais, de faire cas de leurs conseils, de leurs opinions, et de leur affection. Leur soutien indéfectible et toujours serein aura été la plus fidèle des boussoles dans les moments de doute. Clotilde et Wandrille montreront à leur tour dans quelques années qu'ils sont bien plus brillants que moi. En attendant, je leur attribue sans hésiter leur doctorat de meilleurs frères et sœurs, qu'ils ont passé avec mention ! Je ne remercierai jamais assez Julie d'avoir été une compagne si drôle, attentionnée et compréhensive tout au long de cette aventure. Du sac à dos autrefois rempli d'articles jusqu'à nos années futures bientôt remplies de biberons, l'avoir auprès de moi est ma seule réussite incontestable. Je conclurai enfin ces déjà trop longues lignes en dédiant modestement ce travail à la mémoire de Camille dont le souvenir et l'affection ont été un guide perpétuel ces dernières années, et sont encore aujourd'hui des plus vivants.

Chapter 1

Introduction

Not the least charm of a theory to be refuted.
Friedrich NIETZSCHE

The primary goal of econometrics is to model an outcome variable Y with the help of explanatory variables Z. This relationship between Y and Z is however possibly perturbed by other unobservable variables. Their influence is generally carried through a disturbance term U (also called noise or perturbation) incorporated in the model. Regardless of the sophistication of the model linking Y with Z and U, no estimation is possible as long as no assumption is made on U. Generally, estimation begins with an assumption on the (in)dependence between U and Z. This (in)dependence condition also ensures (or not) the identification of the object of interest, depending on the form of the model. Exogeneity is the situation where this independence condition, under its various forms (non-correlation between U and Z, conditional independence or full independence) holds. Its treatment is standard in econometrics and well mastered for many years now. Exogeneity may also be considered as the econometric counterpart of the concept of "experimentation" in other scientific fields. We understand through the term of *experimentation*, situations where the explanatory variable may be fixed, controlled or assigned by the experimenter.

Endogeneity is generally defined as the non-exogeneity situation and is, beyond mathematical definitions, a difficult concept. In a wide range of economic and social studies, the independence between the explanatory variables and the unobserved ones makes no sense. It is of course always mathematically possible to make such an assumption between Z and the perturbation U. But the result of the related estimation procedure is strongly different from the object of interest. In the static framework, solutions to deal with endogeneity are known for long: Instrumental Variables and Control Functions are such examples. In a dynamic framework, contributions and illustrations are however harder to find, both from a theoretical and a practical point of view. The aim of this thesis is to see how the notion of endogeneity and the concept of Instrumental Variables could be extended from static variables to processes such as diffusions, counting processes and duration models.

The motivations and the origins of such a question may be found in *observational studies* (as in medicine, economics, or social studies) where experimentation is not possible. When experimentation is possible, identical conditions for an experience may be recreated and repeated. The causal effect of an explanatory variable is then estimated as the average response of the outcome of interest to different manipulations of the explanatory variable. In presence of endogeneity, experimentation is generally not possible since the explanatory variable cannot be manipulated. Any relationship estimated between the outcome and the explanatory variable must take into account this potential endogeneity. If not, any "what-if" scenario would be totally pointless. In order to estimate a causal effect in this case, the average response of the outcome of interest must be conceived as the difference between counterfactual situations. The manipulation of the explanatory variable is only hypothetical, as two counterfactual situations cannot be observed in the same time. The notion of counterfactual is one of the salient features of the work of James Heckman, Nobel Prize in 2004. We detail this notion in Section 1.1.2.1.

This is even more complex when the explanatory variable takes the form of a process that affects through time the outcome of interest, namely an other process. In medical studies, this outcome process may be a level of contamination, a degree of illness, or a duration process underlying a time-to-recovery. The aim of a causal analysis would be to estimate the causal effect of the so-called explanatory process that could dynamically influence this outcome. Examples for this explanatory process may be the level of drug or treatment given to the patient. When experimentation is possible, the treatment is totally controlled in the sense that the path of the dynamic treatment may be preliminary and arbitrarily fixed according to a given scheme. The causal effect is not difficult to estimate, simply because as expressed before, the difference between two different paths of treatment would be directly observable. Conversely, in situations of endogeneity, the explanatory process influences the outcome but its path is not arbitrarily assigned. Most of the time, the treatment is chosen according to the patient status but also according to its probability to recover, then anticipating on the potential response of the patient to the given treatment. This explains how endogeneity arises.

Some attempts to circumvent such biases have been made, for instance *G-computation*, that is detailed in Section 1.2.3.3. Those attempts are however very specific and focus on a particular use. We aim to be more general and to give a whole framework for the use of Instrumental Variables in this context. This is the main objective of this work as developed in Chapter 2. Chapter 2 concentrates the main theoretical novelties of this work. In particular, we will see that in the particular case of duration models, the translation of the concept of endogeneity implies unusual assumptions and sophisticated resolution involving Inverse Problems.

We will also be interested in practical examples where endogeneity lies in a dynamic context. The study of financial, speculative funds called Hedge Funds will provide us with such an illustration. With specific regulation and unusual investment practices, those funds are very particular in the financial industry and constitute a well-adapted source of study for the econometrician. In this perspective, Chapter 3 precisely reviews the biases and the specificities of the study of processes driving the presence of Hedge Funds in commercial databases. This step is mostly empirical but is crucial in the understanding of all the variables involved in the econometric study of such funds. Improving the current knowledge upon databases by adding new statistical findings to the existing literature is then a natural objective.

The clearest application of the theory developed in Chapter 2 will be provided in Chapter 4. The study of a non-conventional variable (called *lockup*) will be enlightening as we will show that this variable is endogenous and has an influence on the the lifetime of a fund. We develop a genuine framework of causal analysis to study the impact of this variable. Our approach and our findings are at the opposite of the results of the classical literature on the subject, but perfectly illustrates the endogeneity issue in a dynamic context. It also clarifies the ambiguity that may exist between conditional duration models and the structural approach we propose.

Our last chapter is quite disconnected from the previous chapters. It is however included in this manuscript as the link arises however through the use of Inverse Problem theory, already used in Chapter 2. We will show that this theory may be very useful to provide convenient readings of traditional financial problems such as the one presented in Chapter 5: portfolio allocation. If the endogenous topic is not present in this last chapter, some new findings will be however presented.

In this introduction, we will try to present and link the basics of the several problems tackled in this manuscript. The variety of the subjects (endogeneity, processes, inverse problems and Hedge Funds) will maybe make this introducing section a bit heterogeneous. We do not aim to be exhaustive on each topic, but we will mainly try to present each one along with a review of literature. In this respect, Section 1.1 tries to draw the connections between endogeneity, structural models, counterfactuals and causal analysis. In order to precise the field of application of our work, Section 1.2 will explore the classes of processes that are adapted to our analysis. Inverse Problems and regularization will be presented in Section 1.3. Finally, Section 1.4 focus on the presentation of Hedge Funds, their nature and specificities. Most of all, we try to highlight why their specific features lead to studies where endogeneity is omnipresent, and why the dynamic nature of the modelling is also a key point.

1.1 Endogeneity

1.1.1 Definition and examples

Giving an intuitive definition of endogeneity is difficult. The concept of exogeneity is better mastered and endogeneity is often presented as its alternative, or the situation where exogeneity assumptions do not hold. When studying the relation between a random variable Y and explanatory variables Z, econometric models are mainly interested in the mechanism that allows to analyze, predict or reproduce Y given Z, up to a random perturbation term U that impacts this mechanism. U may encompass the influence of unobservable disturbances and variables. Inference is possible only if some assumptions on U are made, as this perturbation is unobservable. Exogeneity specifies that Z and U must be sufficiently independent for inference to hold. Depending on the model, the strength of the independence assumption will also drive identification and unicity of the representation. In this case, the purpose of the analysis is to identify the conditional distribution of $Y|Z$ or a transformation of it.

Example 1.1.1. Linear regression model - The usual linear model assumes that $Y = Z\beta_0 + U$ and U is supposed to be orthogonal to Z i.e. $\mathbb{E}[ZU] = 0$. In the following examples we will assume that $\mathbb{E}[U] = 0$. In this case it is well known that the *Ordinary Least Squares* (henceforth OLS) estimate $\hat{\beta}_0$ of β_0 obtained as the minimizer of $||Y - Z\beta||^2$, is unbiased and equal to $\hat{\beta}_0 = (Z'Z)^{-1}Z'Y$.

Example 1.1.2. Nonlinear separable regression model - This is the nonlinear equivalent of Example 1.1.1 as it is assumed that $Y = \phi(Z) + U$ where $\mathbb{E}[ZU] = 0$ is now replaced by the stronger assumption $\mathbb{E}[U|Z] = 0$. In this case ϕ is equal to the conditional expectation of Y given Z:

$$\mathbb{E}[Y|Z] = \mathbb{E}[\phi(Z) + U|Z] = \mathbb{E}[\phi(Z)|Z] = \phi(Z).$$

Example 1.1.3. Nonlinear nonseparable regression model - Here the relationship between Y, Z and U is very general. Assuming $\mathbb{E}[U|Z] = 0$ is not sufficient. In addition to $Y = \phi(Z, U)$, it is required that $Z \perp\!\!\!\perp U$. In this case, ϕ describes the whole distribution of Y conditional on Z. A crucial assumption is that for a given $Z = z$, $u \mapsto \phi(z, u)$ has to be monotonous. This can also be related to quantile regression methods.

The endogeneity diagnosis is not easy: causes are various and sources of endogeneity often rely out of the control of the econometrician. Historically, the initial modelling of endogeneity has arisen in the context of simultaneous equations, where a system of equations jointly determines the outcome/dependent variable and the explanatory variables. Then, the explanatory variables are internally (or *endogenously*) determined. An economic example is straightforward in models of demand and supply where the price of a good is jointly determined by two equations. Other examples arise in social policy evaluation, or educational programs, where explanatory variables are affected or driven by choices, anticipations, or (unobservable) characteristics of individuals.

1.1.2 Resolution

1.1.2.1 Why solving for endogeneity?

Without a control of endogeneity, estimations are biased: it is quite difficult to isolate the effect of the endogenous, explanatory variable Z as the variations of Z affect the outcome also via the noise term. An other crucial consequence is that most of the time, with the estimated model, we cannot consider that Z plays an "arbitrary" role. Z cannot be conceived as being randomized or conceptually fixed in an observational study. We cannot make any *"what-if"* analysis on Y based on a description of its conditional distribution given Z. Mathematically, the data generating process that generates Z cannot be ignored. This is a major difference with the exogeneity situation, where the conditional distribution of $Y|Z$ incorporates all the necessary information for inference purposes. $Y|Z$ is easy to estimate, but insufficient in a "manipulation" perspective if we conceive Z as being assigned. For instance, for the policymaker in social experiments, this is a major drawback as evaluating the causal effect of a policy discarding endogeneity biases is nearly impossible.

In this respect, endogeneity is tightly linked to the treatment effect literature. A treatment effect may be estimated as soon as acting on a cause of an outcome of interest, the response in the outcome differs under two different level of treatment (determined at the same time, and under the same conditions). This finds a natural application in causal analysis ranging from medicine (the effect of a drug treatment) to social sciences (response of an individual to a social program). Evidently, this would be tempting for individuals with given characteristics, generating outcomes Y, to evaluate the whole response pattern in Y to a whole range of values of the treatment. Unfortunately, this is most of the time impossible in observational studies, because it will be unrealistic in most situations to repeat the experiment several times for various values of the treatment, for individuals with unchanged characteristics. Especially in medicine, it is not possible to analyze the effect of a treatment on the survival of a patient without affecting its status. Going back in the past to make a new experiment with a new treatment level is not possible. Practically, effective treatments take into account the current patient status and are therefore depending on its anticipated evolution.

The hypothetical assignment of a treatment to a conceptual value (or conceptual path if it's a time-varying process) is known as being a *counterfactual* approach. Counterfactuals are therefore related to causal relationships since a treatment effect is the difference between two conceptual situations (treatment and no treatment) i.e. two counterfactuals. They are the object of a heavy literature and their understanding is sometimes difficult. An introduction could be given by Heckman (2001). They are original in that they do not need to explicit any intervention or manipulation mechanism, and may be viewed as "*hypothetical worlds*". However, causal statements or treatment estimation must stem from a fully defined **structural model** with precise counterfactual issues, but also with a scheme allowing variations in the causes whose effect is under study. As explained in Heckman and Vytlacil (2007), "*models are descriptions of hypothetical worlds obtained by varying - hypothetically - the factors determining outcomes.*" When the schemes that drive the outcome selection are not part of the model, we are in a randomization framework. Consequently, the aim of the econometrician working with counterfactuals is to *"provide a forecast of effects of interventions that have never been experienced before"*. This being done by *"using relationships between unobservables in the outcome and the selection mechanisms"*.

We just gave some explanations concerning counterfactuals that will be reviewed later. A comprehensive discussion on the philosophical (and historical) aspects and the implications of the use of counterfactuals in causal evaluation is provided in Kluve (2004). A more general review of counterfactuals but also of related methods used for causal inference (with a focus on econometrics) is provided in Morgan and Winship (2007). We must retain that the curse of endogeneity is that we cannot consider the assignment mechanism of the explanatory variables as being random, as the explanatory variable is still correlated with the unobserved characteristics. But how to break this dependence?

1.1.2.2 Instrumental Variables (IV)

A central method in the literature is the *Instrumental Variables (IV)* framework. An IV (or alternatively an instrument) W is a set of auxiliary variables that affect the explanatory variable and the outcome, but only through the channel of the regressors. We have to formulate exclusion assumptions: independence assumptions between U and Z are therefore replaced by conditional independence assumptions of U and Z given W. IV are particularly suited when the relation between the outcome Y and the variable Z is more difficult to model than the relation between Y or Z and the chosen instrument. An other situation is when Z is not directly observable, but only through a proxy, or a function of it. The idea is to capture the effect on Y of variations on Z generated by variations in the instrument that only affect Z.

Example 1.1.1 (followed) - $\mathbb{E}[ZU] = 0$ may not be assumed: it is replaced by the orthogonality condition $\mathbb{E}[WU] = 0$[1]. The estimators that are derived are the *Indirect Least Squares* estimators. Discussions on identification are important and imply mainly conditions on the dimension of the in-

[1] A GMM view of the IV estimator is therefore possible through this condition.

struments and on the rank of $\mathbb{E}[WZ]$. In case of over-identification, the estimator is obtained by choosing a linear combination of IV of appropriate dimension. This leads to the setting of the *Two Stages Least Squares* estimator (henceforth 2SLS) where this combination is chosen in order to get the estimator with the smallest variance. For 2SLS, $\mathbb{E}[WU] = 0$ is exploited through the implied property that $\mathbb{E}[P_W(Z)U] = \mathbb{E}[\Pi_1 ZU] = 0$ with P_W the linear projection onto W and $\hat{\Pi}_1 = (W'W)^{-1}W'Z$. For the 2SLS estimator, the obtained expression is $\hat{\beta} = (\hat{Z}'Z)^{-1}\hat{Z}'Y$ with $\hat{Z} = Z\hat{\Pi}_1$. In practice, this estimator is obtained as the OLS regression of Y on the re-estimated version of Z obtained by a first OLS regression of Z onto W. If Z has exogenous components they must be incorporated in the set of instruments (in particular the constant term). More generally, the linear form for U may be replaced by a nonlinear, parametric form (providing nonlinear two-stage least squares and generalized instrumental variables approaches).

In the particular case of simultaneous equations, variables are interrelated so that endogenous variables may appear both as outcomes and as explanatory variables. Starting with a structural form of the system, the reduced form of the system is obtained by eliminating the endogenous variables between the several equations. Using OLS is only possible if the reduced form is recursive: the endogenous variable appears as an explanatory variable only if it's also the left-hand side of a former equation. This more simple writing is also the core of triangular systems. In all cases however, the method is the same: whatever the form of the model or its complexity, exclusion restrictions have to be formulated between variables, instruments and residuals. The estimators derive from those conditions.

Example 1.1.2 (followed) - Florens et al. (2009) consider a semi-parametric model:

$$Y = X\beta + \psi(Z) + U \tag{1.1}$$

where (X, Z) are endogenous. A more generic nonlinear separable model would be:

$$Y = \phi(Z) + U \tag{1.2}$$

where Z is endogenous. Darolles et al. (2010a) and Ai and Chen (2003) tackle such model (adopting in fact a semiparametric approach, in the sense that ϕ is more structured, possibly with parametric components such as $\phi = \phi(Z, \beta, \theta)$). Both 1.1 and 1.2 lead to a resolution implying to solve for an integral equation. Both models assume the existence of IV W such that $\mathbb{E}[U|W] = 0$. Consequently, ψ and β are the solution of:

$$\mathbb{E}[Y|W] = \mathbb{E}[\psi(Z)|W] + \mathbb{E}[X\beta|W]$$

whereas in Equation (1.2), ϕ is the solution of:

$$\mathbb{E}[Y|W] = \mathbb{E}[\phi(Z)|W].$$

Resolution of such equations involves in practice nonparametric estimation of conditional expectations.

Example 1.1.3 (followed) - The case where the model is nonseparable in the perturbation U as been discussed e.g. in Matzkin (2003) or Chernozhukov et al. (2007b). Working with endogenous regressors, the perturbation is assumed to be independent from the instruments W. The function of interest has to be monotonic in the perturbation. With a scalar perturbation, the model leads to conditional quantile restrictions. Assuming that $Y = \phi(Z, U)$ with $U \perp\!\!\!\perp W$ (ϕ monotonic in U) for a fixed $Z = z$, the ρ-quantile of $\phi(z, U)$ is equal to $\phi(z, q_\rho)$ where q_ρ is the ρ-quantile of the marginal of U. If U is assumed uniform on $[0; 1]$, and $q(z, \rho)$ is the ρ-quantile of the distribution of the outcome interest if Z could be assigned to z, the equation that must be solved is $\mathbb{P}[Y \leq q(Z, \rho)|W] = \rho$. Chernozhukov and Hansen (2005, 2006) adopt a more general framework as they work with a perturbation that is not restricted to the scalar case; they adapt their setting to the analysis of treatment effects, explicit the statistical properties of the estimators, and build tests. Identification is obtained through completeness conditions. Chernozhukov et al. (2007b) obtains local identification through a completeness condition involving the conditional density towards the instrument. Chernozhukov and Hansen (2005) are interested in global identification and develop sophisticated conditions, which rely again on completeness conditions.

Finally, a recent contribution of Hoderlein and Mammen (2007) is interested in describing functionals of the structural nonseparable model of interest (conditional expectations of the derivatives of ϕ) that remain identifiable in the scalar case with relaxed assumptions on ϕ which is no more assumed to be monotonic.

1.1.2.3 Other methods

Other methods exist to solve for endogeneity but they generally still involve additional variables that play a role similar to instruments (see e.g. Blundell and Powell (2003) for a complete presentation and the way they may be extended to semi- and nonparametric models). Among them are the *fitted value* and the *control functions* approaches. They share the same philosophy as in addition to IV, they may be seen as three interpretations of the 2SLS framework with supplementary variables W. Fitted values replace Z by its component that is uncorrelated with the perturbation. For a linear model the fitted values are the linear projection of Z unto W: $P_W(Z) = (Z'Z)^{-1}Z'W$. The residuals that are computed are $V = Z - P_W(Z)$. For the estimation of β in $Y = \beta Z + U$, the exclusion condition that is used is $\mathbb{E}[P_W(Z)(U + V'\beta)] = 0$. In a more general model, $P_W(Z)$ is replaced by $\mathbb{E}[Z|W]$ with the additional assumption that $\mathbb{E}[U|W] = 0$, potentially strengthened in $(U, V) \perp\!\!\!\perp W$. $\mathbb{E}[Z|W]$ may be estimated nonparametrically and is here again re-injected in the structural equation. The efficiency of the method depends in particular of the form of the model and of its linearity.

Control functions also use additional variables and first-stage residuals to control for the endogeneity of explanatory variables. The seminal idea of control functions is that in the 2SLS framework, the estimated coefficient is also the coefficient of Z in the regression of Y on $[Z; V]$ where V is the first stage residual of the regression of Z onto W. In the linear situation, the structural equations exploited here are the consequences of $\mathbb{E}[WU] = 0$ that are $P_{Z,W}(U) = P_V(U)$ or $P_{Z,W}(U) = P_{Z,V}(U)$. Those conditions are strengthened into $\mathbb{E}[U|Z, W] = \mathbb{E}[U, V]$ for a nonparametric separable model, and into $U|Z, W \sim U|V$for a nonparametric nonseparable model. The linear form of the residuals $V = Z - \mathbb{E}[Z|W]$ may also be replaced by any functional of Y, Z, W that respects the structural equations we just indicated. This makes possible to use such a method for instance with simultaneous equations models with no reduced form.

The most important are the common features shared by instrumental variables, control functions and fitted values. They require additional variables (the instruments) that are related to the outcome only through the explanatory variables. For this, estimation is then always possible after some independence assumptions are made between the instruments and the perturbation. The strength of the assumption depends on the model, its linearity and its sophistication. This assumption also defines the structural equation(s) that will be used to estimate the parameters of interest of the model. Those latter are then solutions of integral equations as conditional expectations are often involved. Finally, let's conclude by highlighting that any test of endogeneity should involve instruments. In the linear case, this is done by the Hausman test. More generally, the idea is to compare the conditional mean $\mathbb{E}[Y|Z]$ with $\psi(Z)$ solution of the moment equation $\mathbb{E}[Y - \psi(Z)|W] = 0$ for a given set of instruments W. $\psi(Z)$ and $\mathbb{E}[Y|Z]$ should be different in an endogenous situation. This is different from controlling for the omission[2] of a variable as an endogeneity test allows the conditional mean of $U|Z$ to vary with W. To test the validity of the instruments, in the linear case, the Sargan test regresses the residuals $e = Y - Z\beta_{IV}$ on the instruments and tests whether the corresponding coefficients are significative or not (null hypothesis if the instruments are valid).

1.2 Mathematical Framework

Our aim is to study a class of processes satisfying some minimal regularity conditions, while being sufficiently large to cover practical applications such as duration models. To do this, we will be interested in *special semi-martingales.*

1.2.1 Some definitions

Let $(X_t)_{t\geq 0}$ (t discrete or continuous) a process adapted to a filtration of σ-fields $\mathcal{F}_t$ such that (X_t) is càdlàg (its trajectories are right-continuous and have a left-limit) and that $\mathcal{F}_t$ is right-continuous (that is to say for a given $t \geq 0$, $\mathcal{F}_t = \bigcap_{s<t} \mathcal{F}_s$). For such processes that are adapted and càdlàg, the notion of semi-martingale, classical semi-martingale and decomposable processes are unified (see Protter (2003, chap.3, p.102)).

(X_t) is a *classical semi-martignale* w.r.t. filtration $\mathcal{F}_t$ if there exists two processes H_t and M_t such that:

$$X_t = X_0 + H_t + M_t \tag{1.3}$$

where H_t is $\mathcal{F}_t$-predictable and M_t is an $\mathcal{F}_t$-martingale.

We will retain this result under the name of **Doob-Meyer decomposition** even if a more general definition should assume that M_t is a *local martingale* and that H_t is only a *finite variation process.* However, we will restrict ourselves in the following to the case where M_t is a martingale, for sake of simplicity but extensions of the present work should be possible. We will also assume that $X_0 = 0$. The hypothesis that H_t is predictable ensures in fact that 1.3 is almost surely **unique**. These concepts are fundamental in the theory of stochastic processes, see in particular Dellacherie and Meyer (1971, vol.2,chap.7).

[2] An omitted variable test would require stronger assumptions on the error term, that is that $\mathbb{P}(\mathbb{E}[U|Z,W] = \mathbb{E}[U|Z]) = 1$ against $\mathbb{P}(\mathbb{E}[U|Z,W] = \mathbb{E}[U|Z]) < 1$ (see Blundell and Horowitz (2007)).

1.2.1.1 Example 1: discrete time models

When t is discrete, one may write for $t \in \mathbb{N}^*$ (Protter (2003, chap.3)):

$$X_t = H_t + M_t$$

$$H_t = H_{t-1} + (\mathbb{E}[X_t|\mathcal{F}_{t-1}] - X_{t-1}) \quad , \quad M_t = M_{t-1} + (X_t - \mathbb{E}[X_t|\mathcal{F}_{t-1}]) \quad , \quad M_0 = 0.$$

Moreover, if we note, $\Delta X_t = X_t - X_{t-1}$, we may also write:

$$\Delta X_t = (\mathbb{E}[X_t|\mathcal{F}_{t-1}] - X_{t-1}) + (X_t - \mathbb{E}[X_t|\mathcal{F}_{t-1}]).$$

1.2.1.2 Example 2: continuous time models

For continuous-time processes, we will concentrate our study on processes that may be written:

$$dX_t = h_t dt + dM_t$$

where $H_t = \int_0^t h_s ds$. This will give us access to two main classes of processes: diffusions and counting processes. These two particular cases will be analyzed in details. Diffusion encompasses the cases where $dM_t = \sigma_t dB_t$ where B_t is a Brownian motion (see e.g. Gard (1988)) and h_t and σ_t are respectively called the drift and the volatility. When X_t is a counting process, h_t will be called the stochastic intensity (and exists under specific assumptions on the nature of X_t.

1.2.2 Counting Processes

1.2.2.1 Definitions

A simple point process is a sequence of times $(T_n)_{n\geq 1}$ of random variables in $[0;\infty]$ defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such as $\mathbb{P}(\lim_{n\to\infty} T_n = \infty) = 1$, $\mathbb{P}(T_n < T_{n+1}, T_n < \infty) = 1$ for each $n \in \mathbb{N}$, and $\mathbb{P}(0 < T_1 \leq T_2 \ldots) = 1$. By definition, there are no event at the time origin and only a finite number of events can happen in a finite time interval. One may associate to a sequence $(T_n)_{n\geq 1}$ the *counting process* $(N_t)_{t\geq 0}$ of the number of events within $[0;t]$, which may be characterized by the following form:

$$N_t = \sum_{n=1}^{\infty} \mathbf{1}_{t \geq T_i}.$$

The internal history of the process is $\mathcal{F}_t^N = \{N_s | 0 \leq s \leq t\}$, potentially augmented with the null sets. As a $\mathcal{F}_t^N$-adapted process, a counting process is a sub-martingale and thus decomposable with the Doob-Meyer decomposition as:

$$N_t = H_t + M_t$$

where M_t is a martingale and H_t, $\mathcal{F}_t^N$-predictable, is the compensator process. We recover the framework of section 1.2.1 (see also Karr (1991, chap.2,p.59), Dellacherie and Meyer (1971)). The existence of the stochastic intensity process h_t (with $H_t = \int_0^t h_s ds$) is guaranteed as soon as the compensator H_t is absolutely continuous. The stochastic intensity is also the unique $\mathcal{F}_t$-progressive, positive process, equal to 0 at $t = 0$, and such as $N_t - \int_0^t h_s ds$ is a martingale. It is however often required for the stochastic intensity to be *predictable* (this being implied in particular by the continuity property). The minimal set of assumptions concerns the compensator that has to be positive, increasing, equal to zero at time-origin, predictable and right-continuous (see Jacobsen (2005)).

Protter (2003) explains the conditions for existence of the stochastic intensity : it exists when the jumps of the counting process are totally unreachables and the process is of locally integrable variation. This imposes conditions on the nature of the jump times and evidently on the filtration which is used (recall that the compensator is implicitly depending on the considered filtration). We have also the following characterization for stochastic intensity (see e.g. Brémaud (1991)):

Definition 1.2.1. - Stochastic intensity
If N_t is a point process, $\mathcal{F}_t$-adapted, and λ_t a $\mathcal{F}_t$-progressive process, positive, such as:

$$N_t - \int_0^t \lambda_u du$$

is a martingale. Then λ_u is the stochastic intensity of the process. When it exists, the intensity is unique.

□

It is worth noticing that $\Lambda_t = \int_0^t \lambda_s ds$ is both the compensator of N_t and M_t^2. We have also in Brémaud (1991) the following characterization:

Definition 1.2.2. - Stochastic Intensity (characterization)
If N_t is a point process, $\mathcal{F}_t$-adapted, and λ_t a $\mathcal{F}_t$-progressive process, positive, such as:

$$\forall t \geq 0 \quad \int_0^t \lambda_s ds < \infty \quad \mathbb{P} - a.s.$$

Then if for all positive, $\mathcal{F}_t$-predictable process ϕ_s, the following relation holds:

$$\mathbb{E}\Big[\int_0^\infty \phi_s dN_s\Big] = \mathbb{E}\Big[\int_0^\infty \phi_s \lambda_s ds\Big]$$

then N_t has the $\mathcal{F}_t$-intensity λ_t.

□

If λ_t is bounded and right-continuous, the stochastic intensity of a point process N_t has the following interpretation:

$$h(t) = \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{P}[N_{t+\Delta t} - N_t \geq 1 | \mathcal{F}_t^N]$$

and when $h(t)$ is right-continuous, we have moreover [3]:

$$\mathbb{P}[N_{t+s} - N_t | \mathcal{F}_t^N] = sh(t) + o(s).$$

This underlines the local nature of the modeling when using stochastic intensity. We will consider in the following **the class of processes with continuous compensators** to guarantee the existence of the stochastic intensity. An important theorem for those processes is the following:

[3] It may be found that $\mathbb{E}[N_{t+\Delta t} - N_t | \mathcal{F}_t^N]$ may be replaced by $\mathbb{P}[N_{t+\Delta t} - N_t = 1 | \mathcal{F}_t^N]$ or $\mathbb{E}[N_{t+\Delta t} - N_t | \mathcal{F}_t^N]$. Jacobsen (2005) however underlines that only expression involving $N_{t+\Delta t} - N_t \geq 1$ makes sense, and that one may be careful with other interpretations that must impose that $\lim_{h \to 0} \mathbb{P}[N_{t+h} - N_t \geq 2 | \mathcal{F}_t^N] = 0$.

Theorem 1.2.1. - *Change of time for an univariate point process*
Let X_t be a process with λ_t, a $\mathcal{F}_t$-intensity, and μ_t, a $\mathcal{G}_t$-intensity, where $\mathcal{F}_t$ and $\mathcal{G}_t$ are histories of Y_t with:

$$\mathcal{F}_t^N \subset \mathcal{G}_t \subset \mathcal{F}_t$$

We suppose that $\mathbb{P}$-as, we have $\lim_{t \longrightarrow \infty} X_t = \infty$. For $t \in \mathbb{R}$ we define $\phi(t)$ the $\mathcal{G}_t$-stopping time such as:

$$\int_0^{\phi(t)} \mu_u du = t$$

Then the process $\tilde{X}_t = X_{\phi(t)}$ is a Poisson process with an intensity equal to 1.

□

Finally, the **innovation theorem** states that for a counting process N_t and two filtrations $\mathcal{F}_t$ and $\mathcal{G}_t$ such that $\mathcal{F}_t^N \subseteq \mathcal{F}_t \subseteq \mathcal{G}_t$, we have $\lambda_t^{\mathcal{F}} = \mathbb{E}[\lambda_t^{\mathcal{G}} | \mathcal{F}_{t-}]$.

Examples of counting processes

The *dead process* is a trivial example where the counting process where no event occur in finite time. There are no jumps and $\mathbb{P}(T_1 = \infty) = 1$. The *canonical Poisson process* is characterized by waiting times $T_n - T_{n-1}$ that are independent and identically distributed following an exponential law of parameter $\lambda > 0$, that is to say for $n \in \mathbb{N}$:

$$\mathbb{P}(N_t - N_s = n) = \frac{\lambda(t-s)^n}{n!} e^{-\lambda(t-s)}.$$

In particular its compensator is deterministic. Finally, Cox processes are particular cases of doubly stochastic Poisson processes. Let L be a positive random variable which is $\mathcal{F}$-measurable. N counting process on $(\Omega, \mathcal{F}, \mathbb{P})$. N follows a Cox process if conditionally on $L = \lambda$, N follows a Poisson process of parameter λ. In this case, the $\sigma(L, \mathcal{F}_t^N)$-compensator is $\Lambda_t = Lt$. As a particular case, the homogenous Poisson process has a stochastic intensity which is constant ($\lambda_t = \lambda$, say). The relative events are totally independent one from the other and this process may be viewed as having no memory, as for two time-intervals with null intersection, the number of events of each are independent. The increments of the process are also independent and the number of events occurring in an interval follow a Poisson law of parameter proportional to the length of the considered interval up to λ.

1.2.3 Duration models

1.2.3.1 Introduction and definitions

A duration τ is the length of a time-period spent by an individual under study in a given state. Duration models are commonly linked to survival analysis. Generalizations are numerous and this extends to multi-individual studies, with possibly several states and several reasons for the individuals to leave their state. The range of applications of duration models is of course wide: medicine, econometrics, finance, sociology. One may wish to analyze the durations of marriages, the time between two transactions, time to effective treatment, etc. Lancaster (1992), van den Berg (2000) or Kalbfleisch and Prentice (2002) provide extensive introductions to duration study.

Before the occurrence of the event of interest that defines the duration, an underlying process is present. Thus the date of observation for instance, makes that observation is not complete for some individuals. This is censorship (i.e. incompletion of the underlying process that induces the observation of a duration for endogenous or exogenous reasons): selection bias, date of study, delayed birth, self-selection, etc. Depending on the nature of this censorship, the analysis will become more sophisticated: crucial assumptions have to be made upon the potential (in)dependence of the censorship mechanism towards the data generating process. When censorship occurs at the beginning of the duration (beginning of the life of the individual) we speak about left-censoring. In the opposite cas, this is right-censoring. In the following we will mainly focus on individual data, leaving aside tied events or multiple-spell data.

A duration τ must be understood as the event time for a process leaving one state. Throughout this thesis, we will assume that $\mathbb{P}(\tau = \infty) = 0$. The distribution function F of τ will be defined as $F(t) = \mathbb{P}(\tau \leq t)$ for $t \geq 0$. Alternatively, the survivor function S corresponds to $S(t) = \mathbb{P}(\tau \geq t) = 1 - F(t) + \mathbb{P}(\tau = t)$ for $t \geq 0$. The density f of τ is such that:

$$f(t) = \frac{dF}{dt} = \frac{-dS}{dt} \quad \text{and} \quad F(t) = \int_0^t f(u)du$$

Those functions are also clearly defined for any positive random variable with sufficient regularity. The integrated (alternatively cumulative) hazard function of τ is denoted $\Lambda(t) : \mathbb{R}^+ \mapsto \mathbb{R}^+$ and satisfies :

$$\Lambda(t) = \int_0^t \frac{f(u)}{S(u)} du = -\int_0^t \frac{dS(u)}{S(u)} = -ln(S(t)). \tag{1.4}$$

When Λ has a time-derivative, this latter is called the hazard rate $\lambda(t)$ defined via:

$$\lambda(t) = \frac{d\Lambda(t)}{dt} = \frac{f(t)}{S(t)} = -\frac{d\ln(S(t))}{dt}. \tag{1.5}$$

The following formulas hold trivially:

$$S(t) = e^{-\Lambda(t)} \quad \Lambda(t) = \int_0^t \lambda(u)du \quad f(t) = \lambda(t)S(t).$$

Some examples of parametric distributions for f are the Exponential distribution, Weibull, Gamma and Generalized Gamma, Singh-Maddala, Lognormal, or Log-logistic.

1.2.3.2 Main models

The hazard rate is intuitive as it models an instantaneous probability to leave the state at t. It is the natural object of interest of main econometric duration models. Sophistications include most of the time covariates that may be both static or time-dependent. Two main classes of models may be introduced. *Accelerated failure time models* suppose that the covariates will directly affect the event-time whereas *proportional hazard models* aim at describing directly on the hazard rate the impact of the modification of a covariate. Proportional hazard models are in fact a particular case of *mixed proportional hazard models* that allow to build models accounting for unobserved heterogeneity. Note in general that in presence of censoring, the natural filtration, potentially augmented with the covariates processes is not sufficient to describe the intensity process, unless information on the censoring process is “included” in the filtration.

Generally speaking, it is common when working with static covariates to estimate the model:

$$F_\tau(u|X) = exp(exp(\phi(X) + \psi(u)))$$

where ψ is a known monotonic transformation. This leads to the following regression:

$$\psi(\tau) = \phi(X) + V$$

where V follows an extreme value distribution, independent of X in the exogenous situation. This is obviously related to quantile regression (see e.g. Matzkin (2003)) but cannot however account for time-varying covariates or competing risks.

Accelerated failure time (AFT) models

In AFT models, the covariate re-scales the duration variable and modifies the time-scale. In terms of intensity, $\lambda_i(t) = \phi(X)\lambda_0(\phi(X)t)$ (with ϕ for instance parametric in X). It may be rewritten $log(\tau) = -\phi(X) + \epsilon$ with ϵ unspecified, having a 0 mean. For example when ϕ is simply linear, we recover a model of the form $log(\tau) = -X'\beta + \epsilon$. Such a model is particularly adapted to a regression framework and implies in particular that X is implicitly understood as being static.

Proportional hazard models

The main assumption is to specify for an individual i a general function for $\lambda(t)$ depending on some covariates X (potentially dynamic $X = X(t)$), risk indicators r and other variables U:

$$\lambda_i(t) = \phi(t, X_i, r_i, U_i).$$

$r_i(t)$ is a left-continuous process, observable, that includes information on the status of the individual ($r_i(t) = 1$ is the individual is at risk e.g. alive, 0 otherwise). In multiplicative intensity models, $\lambda(t) = \lambda_0(t)r(t)$ and can be either parametric or nonparametric for a whole group of individuals. Here $r(t) = \sum_i r_i(t)$ is the number of individuals at risk. In Additive regression model, $\lambda_i(t) = r_i(t)[\beta(t)X_i(t)]$. The effect on the covariates is for each component itself a function of time. In the Cox regression model parametric, non- or semi-parametric, for an individual, the individual hazard is given by:

$$\lambda_i(t) = \lambda_0(t)exp(\beta X_i(t))r_i(t).$$

λ_0 is a positive, deterministic function of time, called baseline intensity. This baseline intensity is multiplied by a scale factor depending on the covariates. Other forms are possible to describe the relation between the covariates $1+\beta X$, $log(1+exp(\beta X))$ for instance. As stated before, the censoring mechanism may depend on the conditioning covariates, unless all the mechanism is fully described by those covariates and depends on no other form of randomness.

Applying Theorem 1.2.1 we get that $\Lambda(\tau)$ is distributed as an exponential variable. This imply that with static variables, and a Cox intensity model, we get that $log(\Lambda_0(\tau)) = -X'\beta + \epsilon$ where Λ_0 is the integral of λ_0. This is different from the AFT model since in general β may be estimated with Λ_0 unknown, but ϵ is then specified, following an EV1 (extreme value) distribution ($f(\epsilon) = e^\epsilon e^{-exp(\epsilon)}$). Consequently, this approach is not a particular case of the AFT model. In particular, this makes it easy to test whether the chosen model for the intensity is appropriated.

Unobserved heterogeneity (mixed proportional hazard)

In fact, proportional hazard models are particular cases of mixed proportional hazard models (MPH). In this kind of model, a time-independent term, function of unobserved explanatory variables, multiplies the intensity model. For example: $\lambda(t) = \lambda_0(t)exp(\beta X(t))U$. In this case, $log(\Lambda_0(\tau)) = -X'\beta + \epsilon - log(U)$. U is an idiosyncratic unobservable variable that accounts for individual characteristics that cannot be observed. They are sometimes known also as *frailty* models. Those models are interesting in the sense that in an heterogenous population, even with a common monotonous baseline λ_0, the resulting hazard rate at the population level may show an inverted monotonicity. This model highlights that studying the hazard rate at the population level, it is sometimes difficult to understand if the behavior of the hazard rate is mainly driven by evolution of the covariates or by structural, unobservable differences between the individuals. However, conditioning upon U allows to deal with proportional hazard models in an unchanged way. That's why for practical use, both the hazard and the heterogeneity distribution are taken as being parametric under a multiplicative form. However, discussion on the nonparametric identification arising with unobserved heterogeneity is a difficult topic which is the subject of a wide stream of literature, and is particularly hard to achieve with single-spell data. This is for instance discussed in Heckman and Singer (1984) and van den Berg (2000).

Frailty models are mathematically convenient and conceptually appealing as all the randomness is left to the idiosyncratic component. As underlined by Heckman (2000) it is generally difficult to distinguish between effective heterogeneity among individuals, or causal dynamic effects. But the use of intensities and compensators (via the Doob-Meyer decomposition) makes that even if this model always relies on the considered filtration (i.e. the information set and the way it includes the past of the processes) assuming an intensity process is (nearly) always possible. Regardless of the existence of a possible frailty effect, the whole framework can be expressed thanks to a dynamic intensity model. Conversely, the opposite approach (from dynamic models to frailty) may also be plausible, but we must retain that frailty models have an alternative writing with dynamic models (see Aalen et al. (2008)) but without assuming that any dynamic expression will do it.

1.2.3.3 Endogeneity in duration models

Contrary to other kind of processes, some attempts have been made in order to model endogeneity or at least to define terminology, at least for time-series and duration models. Mostly for duration models, some techniques have also been built to deal with treatment effect evaluation in non-randomized experiments.

Causality in time-series

Concerning time-series, the main attempt of causality modeling is the Granger (or Granger-Schweder) causality which focus on the influence of the past on the future of the process. If a process Y_t is $\mathcal{G}_t$-adapted with $\mathcal{G}_t \subset \mathcal{F}_t$, a restatement of Granger's noncausality ($\mathcal{F}$ does not cause Y given $\mathcal{G}$) in discrete time would be that $\mathcal{F}_t^Z \perp\!\!\!\perp \mathcal{F}_{t-1}|\mathcal{G}_{t-1}$. This may be used in particular with $\mathcal{G} = \mathcal{F}^Y$ and $\mathcal{F} = \mathcal{F}^Y \vee \mathcal{F}^Z$ where Z_t is an other process. In essence, Z_t will not cause Y_t in the Granger sense if the prediction of Y with all predictors is not affected by the absence of presence of Z_t in the predictor set. The Granger causality implies the Sims' causality which is a stronger concept. See Florens and Mouchart (1982) for a precise review in discrete time. Extension in continuous time is provided by Florens and Fougère (1996). $\mathcal{F}$ does not weakly globally cause Z given $\mathcal{G}$ when for each $s, t \in \mathbb{R}$ the conditional expectation of Z_t given $\mathcal{F}_s$ or $\mathcal{G}_s$ is identical. Strong global causality is obtained when $\mathcal{F}_t^Z \perp\!\!\!\perp \mathcal{F}_s|\mathcal{G}_s$. This condition is met when $\mathcal{G}$-martingales are also $\mathcal{F}$-martingales An extension of the Granger causality concept is weak instantaneous (non)causality. This is verified when the Doob-Meyer decomposition of the Z process is unchanged for both filtrations $\mathcal{F}$ and $\mathcal{G}$. This is in particular equivalent to the $\mathcal{F}$-adaptability of the compensator. In terms of intensity, this can be stated under the following condition: if λ^Y is the intensity of Y respectively to $\mathcal{F}^Y \vee \mathcal{F}^Z$, Z_t does not cause Y_t is λ^Y is $\mathcal{F}^Y$ adapted. Strong noncausality is obtained when the Doob-Meyer decomposition also holds for $\mathcal{F}_{t^-}$.

An other approach to study causal analysis in a dynamic context is the framework of graphical models (via Directed Acyclical Graphs) that try to exhibit more explicit causal relations by making assumptions on their nature. They state a graphical, causal, scheme of interactions between variables, events, and potential treatment(s) of the overall model. The idea is then to estimate a model (with regressions e.g.) for each edge, and to deduce the corresponding total direct, or indirect effects of variables on others. This approach can evidently be thought with dynamic covariates. They lie far beyond the scope of this thesis, see Pearl (2000) for a comprehensive introduction.

Time-varying covariates in hazard models

When dealing with duration models, a specific terminology[4] has been set-up to deal with dynamic covariates. This is important since the term *endogenous* could be confusing. Usually, the following terminology is adopted. A covariate is *fixed* if this variable is known for all the time of study. In mathematical terms, this variable is assumed is $\mathcal{F}_0$-measurable ($\mathcal{F}$ being here a generic term that denotes the filtration the whole information set available at time t), or static. This is a particular case of *external* covariates that may depend on time and are sometimes called *exogenous* covariates. External covariates are *defined* when the whole relative process is known at the time origin (again, the covariate process Z_t is $\mathcal{F}_0$-measurable for each t). Additionally, an *ancillary* variable is an external time-dependent variable that is not affected by the occurrences of the underlying event under study. However, such covariates processes are only $\mathcal{F}_t$-measurable and not $\mathcal{F}_0$-measurable. Conversely, other time-dependent variables are called *internal* and may be viewed as endogenous. The simplest example is a variable whose observation is not available after the occurrence of the event (Kalbfleisch and Prentice (2002, chap.6), or Aalen et al. (2008, chap.4)).

Evaluation of dynamic treatment effects

When trying to understand causality, *marginal structural models* have been developed to estimate the causal effect on a time-to-treatment of a dynamically assigned treatment. For this the *sequential randomization assumption* is made: at a given date (of treatment assignment), conditionally on the history of the treatment and of the covariates, treated and untreated are exchangeable. This assumption is important (yet untestable) since even with unmeasured confounding, it is not possible to proceed to causal inference as risk-factors (covariates) may predict the treatment status (and reciprocally). All the information (necessary to treatment assignation and to individual status that may help to predict the evolution) is however assumed to be available for the econometric study. There are both Cox and AFT structural models. They are generally robust to the absence of the *rank-preservation hypothesis* which states that with a same history, the rank between time of survival of

[4]We mention here a work by Chesher (2002) that tries to identify in hazard models functions of interest concerning endogenous covariates, correlated with an unobserved heterogeneity term that appears under a multiplicative form in the intensity.

two individuals are preserved under the same treatment and under no treatment. In other words, the counterfactual time of survival must be conditionally independent[5] of the assigned treatment (which seems fairly natural). They have tight links with *G-computation* which allows to estimate the causal parameters. Given a structural model (Cox or AFT e.g.), a parametric quantity H_θ is then accessible which models the counterfactual survival time under no treatment when θ is equal to the true value of the causal effect. G-computation then necessitates to model the probability of treatment assignation conditional on the past of the treatment, the covariate and H_θ (by a probit regression framework e.g.). Then, under the sequential randomization assumption, H_θ must have no influence in this regression when θ equals the true value of the causal effect in the structural model. Practically, observations are weighted with weights estimated on discrete time intervals and that are inversely proportional to their probability of occurrence. As H expresses with an integral in time that may be sequentially split, those models are also called *nested models*. Assumptions are then mainly made on the difference between the treatment levels rather than directly on the outcome distribution. G-computation is then linked with counterfactual thinking and may be adapted to continuous-time problems (see also Gill and Robins (2001), Lok et al. (2004)). Finally, Lok (2008) shows that it is even possible to test for the presence of a treatment effect without specifying a model for treatment effect. A deep review of models and assumptions including G-computation and a survey on dynamic discrete choice models is available in Abbring and Heckman (2008).

Three successive papers of Abbring and van den Berg (2003a,b, 2004) have tackled the problem of estimating a treatment effect on duration outcomes and their nonparametric identification. They also deal with dynamic treatments, potentially binary. In this respect they clearly identify the necessity to deal with a dynamical analysis and to consider durations as resulting from a counting process. Their main objective is the construction of a whole causal model for the treatment for hypothetical manipulation of the treatment time assignment. This is the specificity of those papers as they mostly deal with a timing effect of an endogenous treatment assigned dynamically. Identification is however mostly handled with time-invariant covariates. Abbring and van den Berg (2003b) underlines that *"instrumental variables (or, better, processes), if available, may enable identification"* of causal effects. This is a supplementary argument for trying to model endogeneity through instrumental variables in a dynamic context.

Finally, our model will share some motivations with the work developed in Bijwaard (2007, 2008), whose aims are to propose an estimation procedure (called IVLR for *Instrumental Variable Linear Rank* estimator) in order to evaluate treatment effects in studies where outcomes are durations, and where dynamic treatments in continuous time are considered. A GAFT (Generalized Accelerated Failure Time) model, encompassing AFT and MPH models is used to model the hazard intensity and the treatment effect is estimated via the shift on the quantiles of the transformed durations (and then, shares in this aspect some similarities with the work of Lok (2008)). They derive an estimator by making independence assumptions between instrumental variables and the laws of integrated hazards among control and treated groups in the population. Our approach has in common with theirs that: durations are understood within a counting process context; continuous treatments are considered; independence assumptions with instruments involve the integrated hazard (compensator). However we differ in that: we consider general forms of endogenous filtrations, to redefine endogeneity under a broader form, not only for binary time-continuous processes; we do not model any control group and

[5] Lok (2008) strengthens this assumption in a time-continuous context.

do not make any assumption on rank preservation properties, which leads to different independence assumptions between compensators and instruments; the parametric form of intensities are relevant for estimation, not modeling; most importantly, our endogenous processes will not be assumed to be predetermined: deep discussions on adaptation and measurability conditions will be necessary, as an endogenous treatment may also be dynamically assigned, depending on anticipated (dynamic) outcomes.

1.3 Inverse problems in econometrics

Inverse problems are at the heart of many fields of scientific interest ranging from astronomy to quantum physics. They are particularly adapted to the resolution of econometric problems with non-parametric estimation, and then to the resolution of endogeneity in nonlinear models: this motivates the forthcoming introduction. Their usefulness will appear clear in the next chapters. In Chapter 2, we will see that the resolution of our problem of interest needs the theory of Inverse Problems. In Chapter 5, we will see that they provide a convenient way of interpreting specific financial problems. A complete mathematical presentation would be impossible: we present here the minimal yet necessary mathematical background and refer to Engl et al. (1996) and Kaltenbacher et al. (2008) for exhaustive readings. A complement is however proposed in Appendix A.2.

1.3.1 Motivations

1.3.1.1 Definition of inverse problems

Suppose that we dispose of indirect observations y that arise as the transformations of an object of interest f that we wish to identify. This object is usually a vector or a function. The transformations on f are represented by the application of an operator T to f:

$$Tf = y. \tag{1.6}$$

We wish to estimate f after having observed y, possibly through noisy observations. All the difficulties may come from what we really know, observe, or have to estimate. Either T, y, or both, may be unknown. A classical dichotomy between well-posed and ill-posed inverse problem has been stated by Hadamard. A problem is well-posed if three conditions are fulfilled: there is at least one solution to the problem (**existence**); this solution is unique (**unicity**); the solution f depends continuously of the data y (**stability**).

The two first conditions are not usually difficult to obtain. The last condition may be the hardest to verify in practice [6]. As soon as T^{-1} is unbounded any small perturbation in $y \to y + \delta$ is dramatically amplified. As the solution is estimated through $T^{-1}(y + \delta)$, if we do not have any control on δ the estimated solution may be far different from the true one. The definition of the statistical properties of the noise is then a topic of utmost importance. Depending on the problem, this perturbation may appear in y but also in T.

[6] Historically, solutions were believed to always depend continuously on the data. In case of instability, the blame was put on the mathematical model, assumed to be inappropriate, the problem being then *ill-posed.* Since y is observed, the existence of a solution is not the main obstacle. The unicity of the solution is less easy to deal with. If $\mathrm{Ker}(T) \neq \{0\}$ and f_0 is a solution, each $f_\epsilon \in \mathrm{Ker}(T)\backslash\{0\}$ is such that $f_0 + f_\epsilon$ is solution of (1.6) different from f_0. To assess unicity, one can first check the statistical conditions under which T is effectively injective, or one can add some conditions or information on the desired solution.

1.3.1.2 Inverse problems in econometrics

Numerous problems in econometrics appear to be inverse ones. A precise mathematical description of linear econometric inverse problems is provided in Carrasco et al. (2003b), including the generalized method of moments, deconvolution, instrumental variables, or general additive regression models. Other problems in adjacent domains such as economics express also as inverse ones (differential equations, game theoretic models, etc.). A large bulk of literature focuses on the study of inverse problems involving instrumental variables. As previously expressed, estimation using IV is possible through conditional restrictions or moment equations involving instruments. The object of interest is therefore defined through an integral equation, with the left-hand side acting on the object to be estimated, and the right-hand side known or not. The obtained equation, and the operator defined by the integral equation depends on the problem: for nonlinear and general problems, as the expectation operator (conditionally to the instrument) is often involved, nonparametric estimation is commonly needed.

In separable regression models, contributions are numerous. For the standard model $Y = \phi(Z) + U$ with Z endogenous and a set W of IV, Darolles et al. (2010a) clearly identifies ϕ as the solution of the inverse problem $\mathbb{E}[\phi(Z)|W] = \mathbb{E}[Y|W]$, as the primal condition that is exploited is $\mathbb{E}[U|W] = 0$. Operators $\mathbb{E}[...|W]$ are nonparametrically estimated, and the identification issue is tackled. A similar approach is developed in Newey and Powell (2003) and in Hall and Horowitz (2005). A review of semi- and nonparametric methods for (mainly separable) regression models with endogenous explanatory variables is available in Blundell and Powell (2003). Florens et al. (2009) extend this approach to partially linear models such as $Y = X\beta + \phi(Z) + U$ (X and Z endogenous) with instruments W. They study both consistency of the estimators of β and ϕ and their rate of convergence, depending on the potential specification error on the other parameter of interest. They also provide convincing simulations. For further mathematical reading on the link between IV and inverse problems in structural econometrics, see also Florens (2003). Finally, when the restriction $\mathbb{E}[U|W]$ is replaced by conditions such as $\mathbb{P}[U \leq 0|W] = q$, this define new integral equations which are at the heart of quantile regression models with endogenous variables. This kind of model is in particular studied by Horowitz and Lee (2007) and Chernozhukov et al. (2007a).

1.3.2 Operators and ill-posedness

1.3.2.1 Linear problems

Inverse problems may be either linear or nonlinear, depending on T. Linear problems are the most simple way to understand ill-posed inverse problems, but may also be at the heart of the resolution of nonlinear problems. Let's suppose that T is a linear operator between two Hilbert spaces $T : \mathcal{H} \to \mathcal{G}$ which is assumed to be known for the moment, and bounded (thus continuous). The problem is to estimate f from a noisy version of (1.6): $Tf = y + \delta$ where δ is a perturbative error and $y + \delta$ is observed. In a statistical inverse problem δ is considered as stochastic. The inversion of T is then at the heart of the control of the stability. Closely related to the inversion problem and to the stability topic is the singular value decomposition of the operator. Examining the spectrum of the operator (when it exists and may be obtained) allows a precise diagnosis of the magnitude of the stability problem. And as a particular case, the class of compact operators[7] is especially interesting for two reasons: their spectrum is easily accessible and they appear in numerous problems under mild conditions. In the following, when T is assumed compact, it is implicitly renamed $T = K$.

[7]See definition in Appendix A.2.1 p. 157.

1.3.2.2 Ill-posedness and compacity

Let's suppose that we work with a linear, compact operator $K : \mathcal{H} \to \mathcal{G}$ between two Hilbert spaces. In finite dimension ($dim(\mathcal{G}) < \infty$), the injectivity of K is linked to its surjectivity: invertibility and stability of the inverse are consequently well-mastered. When $dim(\mathcal{G}) = \infty$ and K is a finite range operator with $dim(\mathcal{R}(K)) = N_K$, its *singular value expansion*[8](henceforth SVD) $(\lambda_k, \phi_k, \psi_k)_{k\in\mathbb{N}}$ is finite (finite number of non-zero singular values). The inverse is then bounded and:

$$K^{-1}\phi = \sum_{k=1}^{N_K} \frac{1}{\lambda_k} < \phi, \psi_k > \phi_k.$$

In this case the problem is not ill-posed. $\mathcal{R}(K)$ is closed as a sub-space of finite dimension of an infinite dimension space. When $dim(\mathcal{R}(K)) = \infty$, $\mathcal{R}(K)$ is non-closed and:

$$K^{-1}\phi = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} < \phi, \psi_k > \phi_k.$$

As $\lambda_n \to_> 0$, we have that K is unbounded and the problem is ill-posed. We can state the following proposition available in Engl et al. (1996):

If K is compact, $dim(G) = \infty$, then the problem is ill-posed iff $\mathcal{R}(K)$ is non-closed.

The properties of the operator are related to the well/ill-posed nature of the problem. The asymptotic behavior of the sequence $\{\lambda_k\}$ allows to quantify the degree of ill-posedness of the problem. The faster the convergence of λ_k towards 0, the more instable is the problem, and greater is the ill-posedness[9].

1.3.2.3 Nonlinear problems

Equation 1.6 also applies for nonlinear problems. One needs in some situations to replace the operator by a linear counterpart, the Fréchet-derivative. Nonlinear problems are less frequent but we will see in Section 2.4.1.2 that their framework is sometimes a necessity. Let $T : \mathcal{H} \to \mathcal{G}$ be an operator between two Banach spaces.

Definition 1.3.1. Fréchet derivative T is Fréchet-differentiable in $\phi \in \mathcal{H}$ if there exists a bounded linear operator $T'_\phi : \mathcal{H} \to \mathcal{G}$ such that:

$$T(\psi) - T(\phi) - \frac{T'_\phi(\psi - \phi)}{||\psi - \phi||_{\mathcal{H}}} \to_{\mathcal{G}} 0.$$

when $\psi \to_{\neq} \phi$ in $\mathcal{H}$.

There are some elementary conditions to ensure that an operator has a Fréchet-derivative, which may be seen as being a more particular case of the Gateaux-derivative dT, defined as:

$$dT_\phi(\psi) = \lim_{\alpha\to 0} \frac{T(\phi + \alpha\psi) - T(\phi)}{\alpha}.$$

[8]See definition in Carrasco et al. (2003b) In many examples, the singular value decomposition is known.

[9]The terminology is usually the following: if $\lambda_j = \mathcal{O}(j^{-\beta})$, $\beta > 0$ the problem is said to be *mildly* ill-posed; if $\lambda_j = \mathcal{O}(exp(-\beta j^r))$, $\beta > 0$, $r > 0$ the problem is said to be *severely* ill-posed. The parameter β is the degree of ill-posedness of the problem.

To ensure that an operator $T : \mathcal{H} \to \mathcal{G}$ is Fréchet-differentiable, it may be sufficient to check some conditions on its Gateaux-derivative. For this, if ϕ is any element of $\mathcal{R}(T)$:

- (i) T'_ϕ is linear for any ϕ;
- (ii) T'_ϕ is continuous for every ϕ;
- (iii) the mapping $\phi \mapsto T'_\phi$ is continuous on $\mathcal{H}$ on a $||.||_{\mathfrak{L}(\mathcal{H},\mathcal{G})}$ sense (where $\mathfrak{L}(\mathcal{H},\mathcal{G})$ is the space of linear functions between $\mathcal{H}$ and $\mathcal{G}$).

1.3.3 Regularization

1.3.3.1 Definition

Regularization aims at solving the instability problem in ill-posed situations. The general idea is to build a sequence (in the sense of $n \to +\infty$ where n is the size of the data sample) of well-posed problems that approximates the initial ill-posed one. It requires parameters quantifying the distance between the regularized and the raw solution, and rules allowing to choose those parameters. The focal point in the study of the convergence of the solution is the trade-off between the precision and the regularity of the solution. In fact, any compact operator is the limit of a sequence of finite-rank operators (see Engl et al. (1996) or Appendix A.2.1 on p.157): this increases again the interest for compact operators. y will denote the exact data and y^δ is its observed, noisy version, where δ will quantify the noise level (in practice $\delta = \delta_n$).

Looking for the best-approximate solution[10] $f^\dagger = T^\dagger y$, we will have access to $f^\delta = T^\dagger y^\delta$. The approximations will be controlled by a regularization parameter α, which will provide stable solutions f_α^δ that depend continuously on y^δ. When $n \to +\infty$, we expect $\alpha = \alpha_n \to 0$ and the regularization technique to make that $f_{\alpha_n}^\delta$ converges towards the true solution. Generally speaking, the estimation of f_α^δ must take into account the nature of the operator T. In this respect, the problem $f^\delta = T^\dagger y^\delta$ is replaced by $f_\alpha^\delta = R_\alpha y^\delta$ where $\{R_\alpha\}$ is a family of continuous (but not necessarily linear) operators that depends on α. Intuitively, as the amount of noise δ diminishes, the regularizing operators chosen in $\{R_\alpha\}$ are "closer to T" (in a sense that has to be precised), and the estimated solution is closer to the true one.

In fact the choice of a regularization technique and of its parameters must also be linked to the total noise δ, the observed data y^δ, and the operator T. What should also be considered is *a priori* information available on the true data y and most of all on the nature/type of solution that we seek to obtain. Solutions with minimal norm may be viewed as a selection criterion among the set of solutions, thus being additional information added in the estimation procedure.

If α is chosen only according to the level of noise δ, and regardless of the observed data, $\alpha = \alpha(\delta)$ is called an *a priori parameter choice rule.* If $\alpha = \alpha(\delta, y^\delta)$ then it is an *a posteriori* parameter choice rule[11]. A family $\{R_\alpha\}$ (with R_α continuous $\forall \alpha > 0$) is a regularization for $T^\dagger$ if the convergence of

[10] The Moore-Penrose (generalized) inverse $T^\dagger y$ is the inverse of T obtained by restriction of its range and extension to $\mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}^\perp$. $T^\dagger y$ is the best L^2-approximation of the true solution, but in general it does not depend continuously on the data.

[11] Methods that depend only on y^δ but not on δ cannot be involved in convergent regularization methods. This does not imply however that such methods may be satisfying at a given, finite δ: this must be understood in an asymptotic perspective.

R_α towards $T^\dagger$ on $\mathcal{D}(T^\dagger)$ with $\alpha \to 0$ is pointwise. In this situation, for each $y \in \mathcal{D}(T^\dagger)$ there exists an *a priori* parameter choice rule $\alpha = \alpha(\delta)$ which allows that (R_α, α) is a convergent regularization method for solving the problem defined by 1.6.

Engl et al. (1996) state that if $\{R_\alpha\}$ is a family of regularizations, assumed to be linear, with $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ an a priori parameter choice rule for every $y \in \mathcal{D}(T^\dagger)$, then (R_α, α) is a convergent regularization method iff :

$$\lim_{\delta \to 0} \alpha(\delta) = 0$$

$$\lim_{\delta \to 0} \delta ||R_{\alpha(\delta)}|| = 0.$$

It is difficult to conclude on the speed of convergence of solutions of ill-posed problems without restricting oneself to specific subsets of solutions, defined via the smoothness of the true solution. Doing this, we get a precise and accurate control on the speed of convergence of the solution (which then depends on this condition, see Appendix A.2.2.2 p.158).

The best-approximate solution is also related to the normal equation (see Appendix A.2.2). Applying the dual of the operator, we recover the problem:

$$T^* T f = T^* y^\delta. \tag{1.7}$$

Then solution f could be computed if we could get the inverse of T^*T which cannot be obtained in ill-posed situations. Regularization procedures use functions F_α to provide an estimator $\hat{f}_\alpha$ of f defined as:

$$\hat{f}_\alpha^\delta = F_\alpha(T^*T) T^* y^\delta$$

where F_α depends on the so-called regularization parameter.

1.3.3.2 Regularization methods

We present briefly the main regularization methods in Appendix A.2.3 p.160. We do not aim at being exhaustive and refer to Engl et al. (1996) and Kaltenbacher et al. (2008) for a complete review. In a word, *spectral cut-off* is related to the spectral approach and eliminates the components corresponding to the lowest eigenvalues. The infinite development is replaced by a finite one. *Tikhonov regularization* adds a penalty to the norm of the solution to compensate the non-invertibility of T^*T. A closed form expression is obtained when deriving the first-order condition. Some variations exist depending on the chosen norm used to control the added penalty. Other methods are not direct but iterative. They apply to linear problems (*iterated Tikhonov*, *Landweber iteration*) and also to nonlinear problems (*nonlinear Landweber*). For this, the linear counterpart of the operator is used (via the Fréchet-derivative). Numerous extensions exist (*Newton-type methods*, etc.).

1.3.3.3 Ill-posedness in econometric problems

The need of regularization methods in econometric problems essentially comes from the frequent occurrence of ill-posedness situations is econometric problems. Practically, they arise often with the nonparametric estimation of conditional expectations. Gagliardini and Scaillet (2009) show precisely why the problem of instrumental separable regression with nonparametric estimation is ill-posed. To overcome this difficulty, Darolles et al. (2010a) using kernel estimation, proceed to Tikhonov regularization and obtain a rate of convergence. Gagliardini and Scaillet (2009) use a Sobolev penalty rather than a L^2 penalization for their regularized Tikhonov estimator. Newey and Powell (2003) and Ai and Chen (2003) restrict the solution to belong to a compact set of functions. Doing this is more stringent and constraining than using a strict regularization method. The solution is sought as a finite, linear combination of parametric functions but no rate of convergence is provided. Hall and Horowitz (2005) use an orthogonal series estimator. Finally, Blundell and Powell (2003) underline that other methods such as control functions may also lead to ill-posed problems.

We have presented the motivations of our forthcoming analysis, and highlighted the interest of using instrumental variables when solving for endogeneity. We have also described the framework of inverse problems that naturally arise as an appropriate tool in those situations. We now turn to the specification of the processes that will be under study.

1.4 Hedge Funds

Hedge Funds are speculative funds with restricted access / information, with limited regulation (up to now). Despite their opacity and the interrogations they raise, they manage however trillions of dollars. They are major participants of the financial industry, and may even impact the stability of markets in times of crisis. Understanding the mechanisms of Hedge Funds management and the stylized facts arising in econometric studies of their industry, and particularly the drivers of failure and survivorship, is then a topic of utmost importance. This section **does not aim** at providing a comprehensive study on the Hedge Fund universe (or to promote Hedge Fund investment!): numerous articles, papers, textbooks, courses, commercial presentations already exist in this perspective. Our objective is rather to highlight why their specific features build an ideal framework for the econometrician, under many aspects, and especially to illustrate endogeneity in a dynamic context. The aim of this section is then to provide a minimal background to understand Hedge Funds and their underlying econometric challenges.

1.4.1 A first glimpse

1.4.1.1 A difficult definition

Hedge Funds are speculative funds which, as Pension or Mutual Funds, invest on financial markets. Whereas Pension Funds manage assets of employees getting back money when retiring, Mutual Funds and Hedge Funds invest on behalf of their clients who pay them for this service. Mutual Funds are regulated by the *Investment Company Acts*, but Hedge Funds are private investment partnerships that have remained for long neither regulated nor monitored by the *SEC* (*Security Exchange Commission*, see p.170). This is consequently difficult to give any “legal” definition, the name of *Hedge Fund* being before all assessed by the financial industry.

The first recognized Hedge Fund has been set up by Alfred Winslow Jones in 1949: the first mention in the press was in *Fortune* in 1966. The term *hedge* is a bit confusing as their main vocation is to speculate for their clients; those clients being funds (Pension Funds, other Hedge Funds, etc.), financial institutions, endowments or high net worth individuals. Yet, this term reflects the initial strategy of Winslow Jones who took relative bets by buying under-priced and selling over-priced securities, while trying to be *hedged* in the same time from directional market movements. The use of specific financial techniques (leverage, short-selling, see definitions p.170) is another appealing feature of Hedge Funds, that differentiates them again from Mutual Funds (see below). This approach remains at the heart of many Hedge Fund strategies and leads to an empirical/statistical definition of Hedge Funds, that are assumed to generate specific return distributions. Those distributions are mainly characterized by high returns, low volatility, and low correlation with traditional financial indices.

1.4.1.2 Hedge Funds strategies

Hedge Funds invest in various asset classes with the help of various financial tools. The population of funds is an heterogeneous group that may however be separated in several main categories, each one representing a type of strategy employed by the fund. Let us precise that those categories are not "registered" in the sense that this classification results from studies and descriptions of academics and practitioners monitoring the global industry. This explains why the differences between them may be subtle or vague, since for each fund, strategy is a self-described feature. Categories are described in more details in Appendix A.4.2. We give however in the paragraphs some insights on the more important classes of strategies that are represented in the Hedge Fund industry. Five main categories may be identified[12].

(i) Long-Short Equity (or also *Equity Hedge*) is a category of funds that buy and sell stocks to take (relative) benefit from over or under-priced assets. Funds are market neutral when the investment is made in order to immunize the portfolio from market movements. Market neutrality is of course not confined to Long-Short Equity funds.

(ii) (Global) Macro funds take directional bets on various financial instruments (such as bonds, equities, stocks, commodities, credit, etc). Long-Short Equity and Global Macro are the two main historical categories of Hedge Funds. Macro funds are now maybe not so different from *Multi-Strategy* funds, and their approach is close to "tactical asset allocation". Macro funds are generally highly dynamic, and use a high leverage, with a top-down approach.

(iii) Relative Value is a group of strategies whose objectives are to be immunized against market risk with various kinds of assets and techniques. Some Relative Value strategies could then be considered as Long-Short Equity strategies.

(iv) Event-Driven funds are specialized in identifying major events for companies such as merging, and try to take benefit from it.

According to HFR (2009), those four main strategies represent respectively (and approximatively) 35%, 20%, 25%, and 20% of the industry in terms of assets under management (henceforth AUM, see definition p.169), and 50%, 18%, 20%, 12% in terms of number of funds (only for single funds).

The last category, *(v) Managed Futures*, is rather special. They trade futures, using mostly trend-following strategies. The distinction with *CTA* (*Commodity Trading Advisor*) is now blurred. CTA are trading organizations that have to register with the *CFTC* (*Commodities and Futures Trading Commission*, see p.169), and that make trading decisions for clients, operating on futures or options. CTA correspond to highly technical trading strategies, use widely technology, and try to exploit inefficiencies across assets and markets. They are usually separated from the other single funds in databases.

[12]Of course a fund can combine several strategies: those are *Multi-Strategy funds*

At this point, we must also mention *Funds of Funds* (henceforth FoF) that are Hedge Funds investing in other Hedge Funds. This may be understood as a portfolio where stocks are replaced by Hedge Funds. They constitute a special category, as opposed to single funds. FoF are diversified, generally bigger and more diversified (by definition), then claimed to be "safer" than single funds. FoF represent approximatively 25% of the total number of Hedge Funds.

1.4.2 Specific features

We highlight in this paragraph the specific features of Hedge Funds in their organization (structure, fees, regulation) and the statistical properties of Hedge Fund returns, that make them so specific among other market participants.

1.4.2.1 Regulation

A Hedge Fund is often a limited partnership or limited liability company. Formerly, the number of investors was limited (to one hundred) but this rule has disappeared in the nineties. For long, the regulation of Hedge Funds has been scarce. The report to the SEC was possible but voluntary. Contrarily to Mutual Funds, Hedge Funds have a greater flexibility to invest in whatever securities they want: no standard limitation of sector, industry, or country. They can use leverage and short-selling: the use of leverage and short-selling is often taken for the empirical definition of Hedge Funds as they appear to be strong proper features. Hedge Funds that trade futures and options must however register as a *"commodity pool operator"* under the CEA (*Commodity Exchange Act*) by the CFTC (*Commodity and Futures Tranding Commission*) and NFA (*National Futures Association*).

Regulation requirements are currently changing and subject to modification. A former rule set up by the SEC in 2004 stated that Hedge Funds were expected to register as investment advisors. This rule applied to US Hedge Funds and to non-US with more than 14 US investors. Funds with less than 25 millions of dollars did not have to comply (see Getmansky (2004)). Regulatory constraints for Hedge Funds have deeply evolved. In Europe, the *AIFMD* (Alternative Investment Fund Manager Directive, November 2010) now rules all the funds that are not concerned by the traditional and common UCITS norms (see definition p.171): Hedge Funds, Private Equity Funds, Real Estate Funds. This directive concerns big funds (over 100 millions euros) or funds of more that 500 millions of euros but with no leverage. Now the concerned funds will have to set a limit on their leverage level and comply from then. The required capital will be enhanced, so will be risk management requirements and additional market participants (depositaries) will be involved in their overseeing. In the United States, the equivalent regulating directive will be played by the *Dodd-Frank act.* SEC and domicile state will share the registration role depending on the size of the fund. The rule set up by the SEC in 2004 is still valid. A main evolution is that funds will have to provide the regulator with their trades and portfolios.

Mutual and Hedge Funds differ also by their reporting constraints. Indeed, it is not compulsory for Hedge Funds to publish their performance. Things are changing fast and in a near future they will not only have to report their performance, but also their positions, this depending on the size of the fund. But until now, each Hedge Fund was defining its own rule of reporting, and was potentially publishing its performances in a database. As the information provided in databases is often missing, limited or subject to caution, this induces several classical biases that are listed below. This report relies on their willingness to provide figures, which is not compulsory and is related to the objectives of the manager. As we will see in Section 1.4.2.4, Hedge Funds have moreover a limited liquidity, and clients that enter in a Hedge Fund agree to have limited rights. As underlined by Khandani and Lo (2007), it must be noted that regulation perspectives are in general more focused on the investor's interest and protection.

1.4.2.2 Return distribution

Empirically, Hedge Funds have specific return distributions. This depends of the strategy and the period but some stylized facts are however commonly accepted in the financial industry. Essentially, Hedge Funds have dynamic trading strategies (sometimes with a very high frequency) and thus a time-dependent exposure to risk. This creates very unusual profiles of returns. In practice, their return distributions are deeply non-gaussian and have option-like properties i.e. there exists a non-linear relationship between Hedge Fund returns and market returns. This stylized fact is particularly well documented and often constitutes a marketing argument for funds in a commercial perspective. Depending on the strategy, different option payoffs appear (see Fung and Hsieh (1997a), Boyson et al. (2010) for further reading).

An other well-known fact has been highlighted by Fung and Hsieh (1997a): proceeding to a factor analysis on Hedge Fund returns, they identify 5 mutually orthogonal principal components that explain 43% of the cross sectional return variance. The 5 style factors have been identified through the self-declared strategies. This kind of analysis is now widespread among practitioners and has led to a very important semantic for Hedge Funds. A related approach consists in replacing PCA by regression models where given risk factors (benchmarks of strategies) are taken for regressors. Regressing returns on risk-factors allow to identify through the constant term (*alpha*) the specific value added by the manager, and to the factor coefficients of the regression (*beta*) the sensitivity of the fund to other risks. Those coefficients are of course depending on the considered risk-factors (which can be investable or not) but the terms *alternative alpha* and *alternative beta* have remained to qualify the value added individually by Hedge Funds and the way they are exposed to the market.

As explained by Darolles et al. (2010b), Hedge Funds challenge in fact the usual assumptions on risk-return trade-off. It is commonly assumed in usual financial semantics that in order to get a superior performance, a greater risk should be accepted. For stocks, e.g., returns volatility is a common proxy for risk evaluation. For main Hedge Funds styles, minimizing volatility could lead to higher risk. Striking examples exist with funds that presented high returns with little volatility just before failure. But funds manager are also personally concerned with reputation risk (see also Aragon and Nanda (2009) and Section 1.4.2.3). They are not induced to take more risk, as if they increase their volatility, they also increase their probability of failure, and then to lower their reputation.

Moreover, Hedge Funds returns show easily some persistence: this serial correlation is for instance studied in details in Getmansky et al. (2004a). This persistence appears to be more evident for a quarterly horizon and decreases with time (see Agarwal and Naik (2000)). But if it was possible to exploit such an auto-correlation, as this is a very well-known empirical fact, fund managers would have tried for long to make profit with it. They would have taken huge leveraged directional bets, and this auto-correlation would have disappeared. But it's not the case and the main argument to explain this is Hedge Fund (il)liquidity. This argument is developed further in Section 1.4.2.4.

1.4.2.3 Fees and manager's reputation

Hedge Funds are said to offer a great performance only if the investor is ready to pay for it. This "price" for the investor is both in terms of fees but also in terms of impediments of the investment (opacity and liquidity). Fees may be of several natures. Darolles and Gouriéroux (2009) give precise details on the various sophisticated schemes that may exist, and beyond usual presentations, provide mathematical insights on their consequences in terms of risk and related stylized facts. Fees schemes vary from one fund to an other, and are contractually defined by Hedge Funds prospectus. They share however some common patterns, and are proportional to the fund size. Generally, funds take *management fees* that are around $1\% - 2\%$. It is a fixed amount asked whatever the performance. Management fees are used to pay employees, technological investments, computers, and common expenses. *Performance fees* appear when there is a positive performance: they are around 10 or 20% of the resulting performance after management fees. As an alternative, *over-performance fees* often exist and are generally greater (between 20 and 50% commonly) and are calculated on the surplus of performance above a given threshold (that may be evolutive) called the *hurdle rate*.

This is the general picture. However, some particularities (see Appendix A.4.1) in the way they are collected make them very special. Fees are estimated each month and put on special accounts; but they are still part of the investment capacity (see definition p.169) and reinvested each month. Generally, *class A* accounts are accounts of external clients, and *class B* accounts are accounts of the company on which fees are accumulated. Clients know the fees calculation method but are only provided with the net value of the accounts of class A. Regularly (annually, semi-annually, etc.) the accounts are reset and fees are collected by the fund. Commonly, a *Highwater Mark* (see p.169) is included: this indicates that fees are not collected as long as the level of the investment is below its historic maximal value. An other possible scheme of fees collection is the *Loss Carry Forward* scheme (see p.170). Darolles and Gouriéroux (2009) highlights the asymmetric nature of this remuneration scheme as the way accounts of class B are constituted may induce an increase of risk taken by the manager, but are detrimental to client performance. Finally, more confidential schemes implying *Provision Accounts* also exist (see p.171). They are special in the sense that they use a third account that acts as a buffer in difficult times. The account is regularly feeded when the fund has positive returns, but money is taken out to be transferred to the client account in case of bad performances. Such schemes help in fact to reduce the fund's volatility and hide its real underlying risk. Such a mechanism does not explain short-term persistence of Hedge Funds returns but helps to recover results of Bollen and Pool (2009) on negative auto-correlation of slight positive returns. This explains why fees collection schemes are really crucial to understand.

The remuneration of the manager is directly linked to the performance of the fund. This relation is however asymmetric as the manager earns nothing when the fund is a bad performer, and has positive earnings in case of good performance. This is a way to ensure the manager to get the best possible performance. Plantin and Makarov (2011) develop a model that seeks the optimal contract between the fund manager and the clients, depending on the rights allowed to each one. They justify in particular the presence of a highwater mark. This asymmetry from the manager's point of view is also studied by Holland et al. (2010), who underline the potential hazard of this form of remuneration. When the value of the fund is below a threshold that does not allow the manager to be paid, there is a temptation to increase the fund volatility as long as the highwater mark is not reached. More risk is then taken which links this fact to survivorship. See also Darolles et al. (2010b).

The bigger the fund, the more important the fees, as they are estimated proportionally to the size of the fund. The temptation for the manager to collect even more capital could be natural. However, the size of a fund is also linked to its survivorship. Agarwal et al. (2009) show that there is a decreasing return to scale. Fund's size and large money-flows can decrease future performance. It's natural to think that investors seek good performance, and that they try to invest in good performers, hoping persistence of positive returns. Managers' incentives and fees are also often thought as being positive signals for the investors, as their objectives will then be in-line with those of the manager. It is also shown that larger size and higher money-flows will induce poor future performance. In the model of Berk and Green (2004) for active management, performance is deteriorated with inflow of capital: fund's size is not compulsory a positive factor for survivorship.

1.4.2.4 Liquidity

"If you thought getting into a HF was tough, try to get out of one."
The Wall Street Journal[13]*, 10 April 2008).*

Fees are not the only constraint in the Hedge Fund universe. A major drawback of Hedge Funds from the investor's point of view is their opacity and their lack of transparency, but it's also their illiquidity. When investing, the client is given few details on the overall strategy of the fund (which does not prevent him from bad surprises). To give superior performance, Hedge Funds managers claim to use illiquid securities, that is assets with long maturities, scarce resource, infrequent pricing, seeking new sources of value.

It is straightforward that if funds invest in illiquid assets, it will be hazardous to allow investors to get their money back at any moment they want. That's why funds nearly systematically set *share restrictions*. Their effect and their implications for econometric studies will be intensively studied in Chapter 4. Share restrictions are of several kinds. A *lockup* is the period, just after investment, during which it is impossible for the investor to redeem its initial shares. After this period, the investor must notify her redemption sufficiently in advance before the redemption is processed: this is the *redemption notice*. Processing to the redemption is possible only at a given frequency which is the *redemption frequency*. It is contractually defined for each fund and may be days, weeks, or months. Finally, *gates* give the manager the possibility to discretionary block or fractionate the redemption in particularly hard times.

Two forms of illiquidity exist: first the illiquidity from the investor's point of view for Hedge Fund investment; second the illiquidity of some securities that are difficult to purchase for the fund, that imply a mismatch between demand and supply on some assets. Illiquidity for the investor is commonly proxied by share restrictions (see Aragon (2007), Derman (2007), or Sadka (2010)). It implies a premium for the investor that cannot get out from a low performing fund to enter in a good one. Derman (2007) estimates that this premium for a 3 years lockup is about 0.16%, comparatively to a one-year lockup. Aragon (2007) and Teo (2010) underline the role of gates, that diminish *ex-post* performance deterioration. But they highlight that *ex-ante*, this could be an incentive for the manager to take more risks in terms of liquidity that the manager should have done. Teo (2010) also founds that portfolios of funds with high liquidity sensitivities outperform others from about 4.63% a year. This is also confirmed by Sadka (2010) that shows that funds taking more liquidity risks outperform

[13] This sentence is quite famous and is often quoted in some papers dealing with Hedge Funds.

other up to 6% annually. This paper also shows that in periods of liquidity crises with scarce supply of assets, strong negative performances can appear. This clearly helps to understand liquidity as a risk-factor. The lack of liquidity of some traded securities may also explain the stylized fact of correlation of returns. This is the argument of Getmansky (2004) and Bollen and Pool (2009), who explain that serial correlation has primarily for cause illiquidity and smoothed returns. It is difficult to get a fair and quick market price for securities with few trading activity and few supply capacity. Funds containing illiquid instruments proceed to an estimation of their returns and sometimes skip little negative returns. Consequently, their returns appear to be smoother than real economic returns, and their variance is upwardly biased, with more serial correlation.

The liquidity topic also asks the question of the role of Hedge Funds at an aggregated level, and their influence on the financial industry and the global economy. Fung and Hsieh (1997a), Ben-David et al. (2010), and Bernanke (2006) all consider Hedge Funds as major liquidity providers for financial markets. As they invest in illiquid and risky securities, they naturally help to correct mispricing, and upgrade channels of risk transfer. As they seek superior performance, they also influence research of new financial techniques and often push for technological innovations. Those features were arguments for Bernanke (2006) to justify their low level of regulation, as Hedge Funds were seen as helping in times of crisis. Unfortunately, with the events of 2007 (see below), one must accept the paradoxical view of Ben-David et al. (2010) that present Hedge Funds in the same time as liquidity providers but also as destabilizing forces on financial markets. During crises, main contagion channels are essentially the various forms of (il)liquidity. Boyson et al. (2010) define contagion as the dependence that appears between the probabilities of extreme events, between various asset classes and Hedge Fund styles that cannot be explained by linear correlation. This paper also shows that contagion mechanisms are reinforced when liquidity is reduced. This phenomenon is also sometimes called *phase-locking* (see Getmansky (2004)). For a comprehensive study and review on the potential links between Hedge Funds and systemic risk, the work of Chan et al. (2007) constitutes however a major reference.

1.4.3 The industry

1.4.3.1 Databases: a particular case

As a consequence of the absence of regulation and the lack of information on Hedge Funds, commercial databases have appeared, raising the interest of investors. TASS, HFR, Eurekahedge, MSCI, InvestHedge, and CISDM are the main examples of commercial databases. From one database to another, the nature and the quality of the data may vary. They share common drawbacks (see below), and the imperfect information they carry is of great interest for econometric studies. Additionally, it is claimed that funds usually report to one (or two) commercial databases at a time. The proportion of Hedge Funds reporting to the five or six main databases is said to be equal to only few percents. No database can then pretend to provide a complete picture of the industry.

There is of course a trade-off for reporting or not for funds, as this report is voluntary. As opacity is a claimed key-factor of the success of Hedge Funds, reporting is in a way a loss of privacy from the manager's point of view. This lose of secrecy is in fact weaker for higher frequency or more diversified strategies (as they manage a great amount of securities, with a quick management of the positions). Agarwal et al. (2010) identify it but shows that in average, returns distributions of reporting and non-reporting are not so different. In a multi-database study, with extended sources of information, they show that self-reporting funds show a deterioration of performance at initial date and final date.

They underline moreover that reporting funds may be not interested by being in contact with all the kinds of investors that subscribe to databases. Incentives for reporting are in fact stronger for young funds and medium-sized funds.

1.4.3.2 Famous biases

We summarize here some major biases induced by commercial databases for econometric studies. All those biases will be reviewed in details in the following chapters as they induce interesting statistical difficulties through various mechanisms (endogeneity, missing information, self-selection, etc.).

First of all, the *selection bias* is the more intuitive: as reporting is not compulsory, funds that are not part of the database will never be observed. Very few studies (except Agarwal et al. (2010)) tackle this issue as it requires a huge work on data and individual contact to non-reporting funds. Non-reporting funds cannot be thought as being part of a counterfactual study: an unidentifiable part of the funds' population is missing in databases that only provide information on funds that desire a commercial exposition and having survived sufficiently long to contact a database and provide figures. This will also be discussed in Chapter 3.

This evidently imply an other bias called *backfilling*: Hedge Funds communicate *a posteriori* past performances to databases. For instance, an existing fund in 2007 survives until 2008 and publishes then a performance relative to the 2007-2008 period. The bias is that we dispose in 2010 of a past track of its performances, conditionally on the fact that we know that the fund has survived. First, we miss the funds that have not entered yet the database. Second, the fund controls the length of the published track: the fund may have selected attractive past returns, or have produced proforma (i.e. virtual or historically re-estimated) figures. This may also be known as *incubation* bias.[14].

An other related bias is the *survivorship bias.* Several definitions are possible (Brown et al. (1999), Fung and Hsieh (2000), Amin and Kat (2002)). One is the difference between average return of all existing funds at the end of the sample period and the average return of a portfolio containing all the funds of the sample. A second and more restrictive definition is the difference between the average return of funds that have specifically survived to all the sample period and the mean return of a portfolio containing all the funds of the sample. Amin and Kat (2002) shows that estimating quantities only on survivors induces an upward bias of 2% for average returns (Fung and Hsieh (2006) estimates this bias to be around 2.5%). The potential consequence for investors interested in Hedge Funds is then to over-allocate in this asset class. This bias is even more pronounced when considering small funds (bias that may be close to 5%). For large funds, this bias is close to zero. This confirms that funds with huge AUM that stop to report are probably not liquidated. Some studies estimate this bias depending on the investment style, or try to quantify similar effects on higher order moments. For more developments, see Ackermann et al. (1999) or Liang (2000).

Concomitantly, the *liquidation bias* is explained by the fact that reporting is sometimes stopped before the real end of the fund. Bad-performing funds hide their last (poor) returns. The exit reason from a database will be explored in great details in Chapter 3 but they cannot be all explained by

[14]Fung and Hsieh (2006) estimates the incubation period as the instant of the highest dropout rate in databases, equal to 14 months. So they suppress systematically the 14 first months of the return history for statistical studies. We will review this topic in great details in Chapter 3

pure liquidation. The argument of Getmansky (2004) is for instance that funds with higher AUM can decide to stop reporting and to close the fund to new investors, rather to attract new clients. A fund that is too big can erode its performance: the manager may decide to limit its capacity rather than to experiment a decrease of the returns, and an increase of the probability of liquidation. Even pure liquidation may be of several kinds, as it may be explained by failure (caused by fraud or by a massive capital withdrawal) or by closure (the manager thinks that there is no more profit to take with the current strategies).

Finally, the *look-ahead bias* arises when studies are made conditional on survival on further consecutive periods (see Baquero et al. (2005)). This may be a problem also related to the construction of some Hedge Funds benchmarks whose composition at a given date is made of top-performing funds.

1.4.3.3 A quick overview of the industry

After the initial attempt of Alfred Winslow Jones, interest for Hedge Funds stagnated during the seventies, because of bear markets and mild results. Some top-performing funds in the eighties made that Hedge Funds experienced a significant resurgence in popularity. This interest has kept growing at the beginning of the nineties with a consequent tremendous increase of AUM. Of course, with the turmoils faced by financial markets since 2007, the exponential growth of the total AUM has been deeply tampered. At the end of 1994, there were between 1000 and 2000 Hedge Funds, with a global AUM between 100 and 160 billions of dollars (Fung and Hsieh (1997a)). According to a Tremont study also quoted in Getmansky (2004), there were in 2003 nearly 5000 single funds (and 1400 FoF) with 700 billions under management, versus 610 funds and 39 billions under management in 1990. In 2007, Hedge Funds operations were representing around one third of the equity trading volume in the United States, for a total amount of AUM between 1.5 and 2.0 trillions of US dollars. 2008 has been the year of worst performances ever, facing negative returns. In June 2008, this amount was estimated around 1.93 trillions (Sadka (2010)), 1.47 trillions at the end of December 2008 and 1.35 trillions at the end of July 2009 (Eurekahedge). In 2009, liquidations have accelerated but this will probably not imply any dramatic change in the fees structure or in the liquidity terms.

With the uprising interest for the world of Hedge Fund, some products or techniques have appeared that claim to give access to Hedge Fund performances without their drawbacks. Among them are the Hedge Fund benchmarks and indices. The non-investable indices are in fact more a way to assess the health of the industry, depending on the strategy. They work as a mean index that proxies the average behavior of precise kind of funds. Investable indices try to mimic non-investable indices and are commercialized. In practice, they are not so different from an index based on a basket of funds, and fail to give access to the same performance promoted by non-investable indices. This is because non-investable indices do not take into account the potential limited capacity of funds, their liquidity, or transaction costs. An other trend, appeared some years ago, is replication, which is a set of techniques that claim to give access to Hedge Fund performance by replicating a given profile of return distribution, without entering any Hedge Fund. The commercialized products are in fact based on classical multi-investable factor regression of the historical returns of a given fund and aim at reproducing its performance month by month. More obscure techniques aim at reproducing any given return profile in average, in a distributional way. This last procedure has many statistical difficulties that are not tackled in the papers advertising for this technique. Details on this are far beyond this manuscript, we refer to Beare (2009) for a remarkable work of statistical modelling on this subject.

Hedge Funds are not always appealing. It would be too long to explain the circumstances of the two big crises that have struck the Hedge Fund industry but we can briefly explain their mechanisms and their consequences. In 1998, Long Term Capital Management (LTCM) encountered severe losses that led it to liquidation. Managers expected differences in interest rates from several countries to tighten. Contrarily to this view, default of Russian bonds induced a flight to quality, increasing the gap between safe and risky bonds. The losses were multiplied by the amount of the leverage, equal here to nearly 20. Precise explanations are available in Edwards (1999). LTCM failure helped to understand for the first time that Hedge Funds could have an influence on the global financial stability. The summer 2007 crisis and its stream of consequences has its sources in the subprime mortgage markets but it shares with the LTCM crisis a common point: the channels of contagion lie in the leveraged margin calls. Khandani and Lo (2007) made an impressive contribution to understand the whole mechanism of the crisis. During summer 2007, banks and funds realized that the probabilities of default of mortgages used in the pricing of subprime products were deeply under-evaluated. This changed instantaneously the face-value of the products, creating important margin calls, amplified by the leverage taken on those positions. To satisfy this sudden need of liquidity, funds have unwinded huge position on the more liquid assets they got: equities. With massive sell orders, some stocks of companies that had nothing to do with subprimes saw their price falling suddenly. Other funds with quantitative automatic trading strategies, mis-interpreted those signals and began to buy stocks that were falling; this increased the losses of other kind of funds. Case studies of several individual fund failures (including frauds) are available in Gupta and Kazemi (2008).

1.4.4 Motivations and open questions

Figure 1.1 sums up the empirical observations we have just exposed. Plain arrows describe the assessed links between structural features, investors and managers behavior. Dashed arrows represent incentives or constraints both for investors and fund managers. A "+" indicates a positive (or increasing) effect of the variable at the origin of the arrow, on the variable at the opposite side of the arrow. Conversely, a "-" indicates a negative (or decreasing) effect. Dashed arrows with question marks represent links where current analysis is missing or insufficient. Our motivation will be to precise the nature of those edges. It is straightforward that the main sophistication of this challenge is that this mechanism is dynamic through time.

In practice, a fund delegates the management of its assets to a fund manager. This manager has a personal perspective: she is interested in her remuneration but seeks to preserve her reputation, for the sake of her future career. Her remuneration is proportional to assets and performance. Past performance help in improving reputation, and future performance is the objective of investors. Investments may be potentially constrained by share restrictions. Potential performance, size and fund reputation drive the inflows of money whose accumulation defines the AUM.

The dynamical link between AUM and performance is more subtle: when the fund is too big, this may decrease future performance, as it is a limit to take profit from any arbitrage. The liquidation risk is the major concern, both for the investor, the fund and the fund manager. A manager whose only focus is remuneration may potentially increase the investment risks to improve performance. This would increase the liquidation risk, with a potential influence on the reputation risk in case of fund termination.

The questions that still remain are then the following. What is the true influence of fund's assets and fund performance through time, on the liquidation risk? Moreover, is liquidation identical to a stop of report in databases (that are the only mean for the regulator or statisticians to collect information)? Those two questions will be addressed in Chapter 3. Finally, what is the role played by share restrictions and what is their influence of fund liquidation? May share restrictions be considered as exogenous? This will be the aim of Chapter 4. Again, Hedge Funds are a perfect field of application to study endogeneity in a dynamic context.

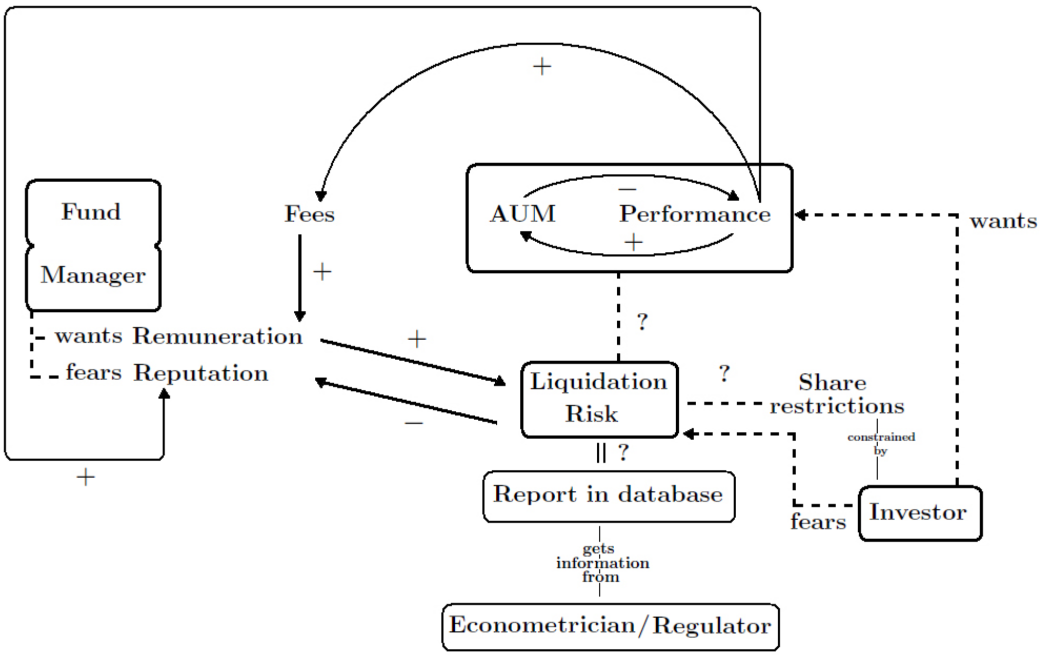

Figure 1.1: A mechanism for Hedge Funds dynamics?

Chapter 2

Endogeneity and Instrumental Variables in dynamic models

joint work with Jean-Pierre Florens[1].

Abstract

The objective of the chapter is to draw the theory of endogeneity in dynamic models in discrete and continuous time, in particular for diffusions and counting processes. We first provide an extension of the separable set-up to a separable dynamic framework given in term of semi-martingale decomposition. Then we define our function of interest as a stopping time for an additional noise process, whose role is played by a Brownian motion for diffusions, and a Poisson process for counting processes.

[1]This chapter is adapted from Florens, J.P., Simon, G. (2010), *Endogeneity And Instrumental Variables In Dynamic Models*, submitted to *Econometrica*. Sections 2.4 and 2.5 are however additional developments that do not appear in the original paper.

2.1 Introduction

2.1.1 Motivations

An econometric model has often the form of a relation where a random element Y depends on a set of random elements Z and a random noise U. If Z is exogenous (see for precise definition of this concept Engle et al. (1983) or Florens and Mouchart (1982)) some independence or non correlation property is assumed between the Z and the U in order to characterize uniquely the relation. For example, if the relation has the form $Y = \phi(Z) + U$ the condition $\mathbb{E}[U|Z] = 0$ characterizes ϕ as the conditional expectation and if $Y = \phi(Z, U)$ with ϕ monotonous in U, U uniform, the condition that Z and U are independent characterizes ϕ as the conditional quantile function. This exogeneity condition is usually not satisfied (as for instance in market models, treatment effect models, selection models...) and the relation should be characterized by other assumptions.

The instrumental variables approach replaces the independence between Z and U by an independence condition between U and another set of variables W called the instruments. For example, in the separable case $Y = \phi(Z) + U$ the assumption becomes $\mathbb{E}[U|W] = 0$ (see for a recent literature Florens (2003), Newey and Powell (2003), Hall and Horowitz (2005)). In the nonseparable model, it is assumed that $U \perp\!\!\!\perp W$ (see contributions of Horowitz and Lee (2007), Chernozhukov et al. (2007a), or Chernozhukov et al. (2007b)). In these cases the characterization of the relation is not fully determined by the independence condition but also by a dependence condition between the Z and the W. This dependence determines the identifiability of the relation: in a nonparametric framework, this impacts the speed of convergence of the estimators.

The objective of this chapter is to analyze dynamic models with endogenous elements. The goal is concentrated on the specification of the models in such a way that the functional parameter of interest appears as the solution of a functional equation (essentially linear or nonlinear integral equation). Using this equation, identification or local identification conditions may be discussed. This chapter is not concerned by statistical inference but shows how the functional parameter may be derived from objects which may be estimable using data. The theory of nonparametric estimation in these cases belongs to the theory of ill-posed inverse problems (see Darolles et al. (2010a), Carrasco et al. (2003a), Carrasco (2008)) and will be treated in specific cases in other papers.

We address the question of endogeneity in dynamic models in two ways. First we consider a separable case which extends the usual model $Y = \phi(Z) + U$ with $\mathbb{E}[U|W]$. However, this case is not sufficient to cover the endogeneity question in models such as counting processes or diffusions. In this case, we analyze the impact of endogenous variables through a change of time depending on the endogenous variables. This approach covers the example of the duration models, the counting processes, the diffusion with a volatility depending on the endogenous for example. It will be shown that those change of time models give an interesting extension of non-separable models in the dynamic case. These two approaches will be treated in Section 2.2 and 2.3 of the chapter and will be illustrated by examples. We first recall in the next paragraph the main mathematical tools we will use.

2.1.2 Mathematical framework

In this chapter we essentially analyze a large class of stochastic processes verifying a decomposition property. Let $(X_t)_{t\geq 0}$ (t may be discrete or continuous) and $\mathcal{F}_t$ a filtration of σ-fields such that X_t is càdlàg (its trajectories are right-continuous and have a left-limit) and that $(\mathcal{F}_t)_t$ is right-continuous (that is to say that $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$). In the usual terminology of the general theory of stochastic processes we will say that $\mathcal{F}_t$ satisfies the "*conditions habituelles*".

A process X_t is a special semi-martingale w.r.t. $(\mathcal{F}_t)_t$ if there exists two processes H_t and M_t such that:

$$X_t = X_0 + H_t + M_t; \tag{2.1}$$

- M_t is an $\mathcal{F}_t$-martingale;
- H_t is $\mathcal{F}_t$-predictable.

A more general definition only assumes that M_t is a local martingale but for sake of simplicity only the martingale case is treated in this chapter. We also simplify the expressions by always assuming $X_0 = 0$. Extension to local martingales and to cases where $X_0 \neq 0$ requires more technicalities (in particular in Section 2.3). Let us note that the decomposition (2.1.2) is a.s. unique. These concepts are fundamental in the theory of stochastic processes (see in particular Dellacherie and Meyer (1971) - Vol II - Chap VII). is the underlying expression of the usual Doob-Meyer decomposition.

We may easily illustrate this definition in the case of discrete time models. In that case we have: $M_0 = 0$, $M_t = M_{t-1} + (X_t - \mathbb{E}[X_t|\mathcal{F}_{t-1}])$ and $H_t = H_{t-1} + [\mathbb{E}[X_t|\mathcal{F}_{t-1}] - X_{t-1}]$ (see Protter (2003) - Chap III). Equivalently $\Delta X_t = X_t - X_{t-1}$ may also be written:

$$\Delta X_t = X_t - X_{t-1} = (\mathbb{E}[X_t|\mathcal{F}_{t-1}] - X_{t-1}) + (X_t - \mathbb{E}[X_t|\mathcal{F}_{t-1}]).$$

In case of continuous time processes, we also restrict our study to cases where H_t is differentiable and we have the expression:

$$dX_t = h_t dt + dM_t$$

where $H_t = \int_0^t h_s ds$. Some particular cases will be analyzed in details. The first one is the single duration model with endogenous cofactors possibly time-dependent. More generally, we analyze counting processes and an example of Markovian transition model is also discussed. Finally, we also applied our approach to diffusion models.

2.2 The additively separable case: the Instrumental Variables decomposition of semi-martingales

2.2.1 The framework

Let us consider a multivariate stochastic process $X_t = (Y_t, Z_t, W_t)$ (with $Y_t \in \mathbb{R}$, $Z_t \in \mathbb{R}^p$, $W_t \in \mathbb{R}^q$) and $\mathcal{X}_t$ the filtration generated by X_t i.e. $\mathcal{X}_t$ is the σ-field generated by $((Y_s, Z_s, W_s)_{s \leq t})$. We consider different subfiltrations of $\mathcal{X}_t$:

1. $\mathcal{Y}_t$, $\mathcal{Z}_t$, $\mathcal{W}_t$ are the filtrations generated by each subprocess;

2. we call the *endogenous filtration* the filtration generated by $\mathcal{Y}_t$ and $\mathcal{Z}_t$, and the *instrumental filtration* the filtration $\mathcal{Y}_t \vee \mathcal{W}_t$ generated by $\mathcal{Y}_t$ and $\mathcal{W}_t$.

We first extend the usual decomposition of semi-martingales in the following way.:

Definition 2.2.1. The process Y_t has a Doob-Meyer Instrumental Variable (DMIV) decomposition if:

$$Y_t = \Lambda_t + U_t$$

where:

1. **A1 -** Λ_t is $\mathcal{Y}_t \vee \mathcal{Z}_t$ predictable ;

2. **A2 -** $\mathbb{E}[U_t - U_s | \mathcal{Y}_t \vee \mathcal{W}_t] = 0$ for $0 \leq s < t$.

□

Equivalently we may say that Y_t is an IV semi-martingale w.r.t. $(\mathcal{Y}_t \vee \mathcal{Z}_t)_t$ and $(\mathcal{Y}_t \vee \mathcal{W}_t)_t$. First we can note that if $\mathcal{W}_t = \mathcal{Z}_t$ this definition reduces to the usual Doob-Meyer decomposition. If the filtration $(\mathcal{Y}_t \vee \mathcal{Z}_t)_t$ is included into $(\mathcal{Y}_t \vee \mathcal{W}_t)_t$ the problem becomes a problem of enlargement of filtrations and preservation of the martingale property. This question is central in the theory of non-causality treated e.g. by Florens and Fougère (1996).

We consider then the more general case where $(\mathcal{Y}_t \vee \mathcal{Z}_t)_t$ and $(\mathcal{Y}_t \vee \mathcal{W}_t)_t$ have no inclusion relation. Moreover, the two filtrations do not need to be generated by processes and $(\mathcal{Y}_t \vee \mathcal{Z}_t)_t$ and $(\mathcal{Y}_t \vee \mathcal{W}_t)_t$, and may be replaced by more general filtrations $\mathcal{F}_t$ and $\mathcal{G}_t$ under the condition that Y_t has to be adapted to each of them. Assumption **A1** means that the predictable process "only depends" on the past of Y_t and on the past of Z_t. Assumption **A2** is the independence condition between the "noise" U_t and the instruments W_t. Equality in $A2$ is a mean independence only (like in the static separable model $Y = \phi(Z) + U$) and looks like a martingale property. It's not strictly speaking a martingale property because U_t is not assumed to be adapted to $(\mathcal{Y}_t \vee \mathcal{W}_t)_t$. The usual decomposition when $(\mathcal{Y}_t \vee \mathcal{Z}_t)_t = (\mathcal{Y}_t \vee \mathcal{W}_t)_t$ is unique a.s. but in the general case, it should be noted that this unicity result is not true: this will be precisely the object of the identification condition analyzed below.

2.2.2 Identification

Let us first consider the characterization of the decomposition in term of conditional expectation.

Theorem 2.2.1. *Let us assume that Y_t is a special semi-martingale w.r.t. $\mathcal{Y}_t \vee \mathcal{W}_t$ and that :*

$$dY_t = h_t dt + dM_t$$

where $H_t = \int_0^t h_s ds$ is $\mathcal{Y}_t \vee \mathcal{W}_t$-predictable and M_t is a $\mathcal{Y}_t \vee \mathcal{W}_t$-martingale.

If the following family of integral equations:

$$h_t = \mathbb{E}[\lambda_t | \mathcal{Y}_t \vee \mathcal{W}_t] \quad t \geq 0 \tag{2.2}$$

with λ_t $\mathcal{Y}_t \vee \mathcal{Z}_t$-measurable and integrable

has a sequence of solutions λ_t, then Y_t is an IV semi-martingale and $\Lambda_t = \int_0^t \lambda_s ds$.

□

Roughly speaking, Equation (2.2) means that we have to solve:

$$\begin{aligned} h_t dt &= \mathbb{E}[dX_t | (Y_s, W_s)_{0 \leq s \leq t}] \\ &= \left[\int \lambda_t((Y_s, Z_s)_{0 \leq s \leq t}) f((Z_s)_{0 \leq s \leq t} | (Y_s, W_s)_{0 \leq s \leq t}) d(Z_s)_{0 \leq s \leq t} \right] dt. \end{aligned}$$

This expression is not mathematically rigorous because the arguments of the functions are infinite dimensional but it shows how our definition extends the static separable case.

A DMIV decomposition exists if and only if h_t belongs to the range of the "instrumental" conditional expectation operator. If we restrict our attention to square integrable variables, this operator is defined on $L^2(\mathcal{Y}_t \vee \mathcal{W}_t)$. Note that the conditional expectation operator is compact under minor regularity conditions. Its range is then a strict subspace of $L^2(\mathcal{Y}_t \vee \mathcal{W}_t)$ and the existence assumption is an over-identification condition on the model. The main question concerns the unicity of the solution, which is equivalently the identifiability problem. Given the distribution of the process X_t, the function h_t, and the conditional expectation operator $\mathbb{E}[...|\mathcal{Y}_t \vee \mathcal{W}_t]$ defined on $L^2(\mathcal{Y}_t \vee \mathcal{Z}_t)$ are identifiable. The DMIV decomposition is then unique (or equivalently Λ_t is identifiable) if and only if the conditional expectation operator is one-to-one. The following concept extends the (fully known) case of static models.

Definition 2.2.2. The filtration $(\mathcal{Y}_t \vee \mathcal{Z}_t)_t$ is strongly identified by the filtration $(\mathcal{Y}_t \vee \mathcal{W}_t)_t$ (or Z_t is strongly identified by $\mathcal{W}_t$ given $\mathcal{Y}_t$) if and only if for $t \geq 0$:

$$\forall \psi \in L^2(\mathcal{Y}_t \vee \mathcal{Z}_t), \mathbb{E}[\psi | \mathcal{Y}_s \vee \mathcal{W}_s] = 0 \Rightarrow \psi = 0 \quad a.s.$$

□

Corollary 2.2.1. *The DMIV is unique if $(\mathcal{Y}_t \vee \mathcal{Z}_t)$ is strongly identified by $(\mathcal{Y}_t \vee \mathcal{W}_t)$.*

□

For a good treatment of conditional strong identification and its relation with the completeness concept in statistics, see Florens et al. (1990) - Chap 5. Then if $\mathcal{Z}_t$ is strongly identified by $\mathcal{W}_t$ given $\mathcal{Y}_t$, the conditional expectation operator is one-to-one and Λ_t is identified. Several papers give more primary conditions which link this property to the conditional expectation operator (see a recent contribution of d'Haultfoeuille (2010)). We want to illustrate this concept in two examples : discrete-time models and diffusions.

2.2.3 Examples

2.2.3.1 Example 1 : discrete time model

Suppose that we have a discrete time model such as:

$$y_t = \lambda(\xi_t) + \epsilon_t$$

with $\mathbb{E}[\epsilon_t | y_{t-1}, \ldots, \xi_{t-1}, \ldots] = 0$. In our framework, we have then:

$$Y_t = y_0 + ... + y_t \qquad \Lambda_t = \lambda(\xi_0) + ... + \lambda(\xi_t) \qquad U_t = \epsilon_0 + ...\epsilon_t.$$

Moreover if we define:

$$\mathcal{Z}_t = \sigma\{\xi_{t+1}, \xi_t, \ldots\} \quad \mathcal{Y}_t = \sigma\{y_t, y_{t-1}, \ldots\} \quad \mathcal{W}_t = \sigma\{\xi_t, \xi_{t-1}, \ldots\}$$

then we have the following properties:

- Y_t is $\mathcal{Y}_t \vee \mathcal{Z}_t$-adapted and $\mathcal{Y}_t \vee \mathcal{W}_t$-adapted;
- Λ_t is $\mathcal{Y}_t \vee \mathcal{Z}_{t-1}$-measurable and $\mathcal{Y}_t \vee \mathcal{Z}_t$-predictable;
- $\mathbb{E}[U_t | W_{t-1}] = U_{t-1}$ as $\mathbb{E}[\epsilon_t | \mathcal{Y}_t \vee W_{t-1}] = 0$ and $\mathbb{E}[\epsilon_s | \mathcal{Y}_t \vee W_{t-1}] = \epsilon_s$ if $s \leq t-1$.

In that case, $\mathcal{Y}_t \vee \mathcal{W}_t \subset \mathcal{Y}_t \vee \mathcal{Z}_t$, $\lambda_t = \lambda(\xi_t)$, $h_t = \mathbb{E}[\lambda_t | \xi_{t-1}, ..., y_{t-1}, \ldots]$, i.e. :

$$h_t = \mathbb{E}[\lambda(\xi_t) | \xi_{t-1}, \ldots, y_{t-1}, \ldots, ...].$$

If (y_t, ξ_t) is Markovian, then we have moreover:

$$\mathbb{E}[y_t | \xi_{t-1}, y_{t-1}] = \mathbb{E}[\lambda(\xi_t) | \xi_{t-1}, y_{t-1}].$$

One can then proceed to nonparametric estimation:

- for weakly dependent stationary processes, we face inverse problems as in the usual i.i.d. case;
- when studying unit root processes, we can use ordinary kernel estimation but there is a second order bias. Wang and Phillips (2009) treats it with a control function, but this does not address the case of the second order bias of instrumental variables. This is therefore an argument for pure IV in non-stationary models.

More generally, we could consider:

$$y_t = \lambda_t(z_t, z_{t-1}, \ldots, y_{t-1}, \ldots) + \epsilon_t$$

with $\mathbb{E}[\epsilon_t|w_t, w_{t-1}, \ldots, y_{t-1}, \ldots] = 0$, $\mathcal{Z}_t = \sigma(z_t, z_{t-1}, \ldots, y_t, \ldots)$ and $\mathcal{W}_t = \sigma(w_{t+1}, \ldots, y_t, \ldots)$. The decomposition of $Y_t = y_0 + \ldots + y_t$ w.r.t. $\mathcal{W}_t$ writes:

$$Y_t = \sum_{j=1}^{t} \underbrace{\mathbb{E}[y_j|\mathcal{W}_{j-1}]}_{=h_j} + \sum_{j=1}^{t} (y_j - \mathbb{E}[y_j|\mathcal{W}_{j-1}])$$

with $h_t = \sum_{j=1}^{t} h_j$. We must then solve:

$$h_t = \mathbb{E}[\lambda_t(z_t, \ldots, y_{t-1}, \ldots)|\mathcal{W}_t].$$

2.2.3.2 Example 2 : diffusions

Let us assume that the structural model has the following form :

$$dY_t = \lambda_t(Y_t, Z_t)dt + \sigma_t(Y_t)dB_t \tag{2.3}$$

where B_t is a Brownian motion. This means that if Z_t is fixed (or randomized, and not generated by the distribution mechanism), Y_t follows a diffusion process with a drift equal to λ_t and a volatility equal to $\sigma_t(Y_t)$. Note that we assume that Z_t does not appear in the volatility term. Let us assume that:

$$\mathbb{E}[dB_t|\mathcal{Y}_t \vee \mathcal{W}_t] = 0.$$

In that case Equation (2.3) characterizes the DMIV decomposition of Y_t. In order to identify λ_t we need to construct the decomposition of Y_t w.r.t. the filtration $\mathcal{Y}_t \vee \mathcal{W}_t$ (that we write $dY_t = h_t dt + dM_t$) and to solve:

$$h_t = \mathbb{E}[\lambda_t|\mathcal{Y}_t \vee \mathcal{W}_t]. \tag{2.4}$$

Note that the "reduced form" model $dY_t = h_t dt + dM_t$ has no reason to be a diffusion. Conditionally on $\mathcal{W}_t$, the process may be non Markovian and M_t maybe different from a Brownian motion. This general framework may be applied to particular cases, and simplifies the estimation problem. For example, let's assume that the structural model has an Ornstein-Uhlenbeck form:

$$\lambda_t(Y_t, Z_t) = \theta(\mu(Z_t) - Y_t) \quad \text{and} \quad \sigma_t(Y_t) = \sigma^2$$

where θ is a constant. In that case the model becomes a semi-parametric model:

$$h_t = \theta(\mathbb{E}[\mu(Z_t)|\mathcal{Y}_t \vee \mathcal{Z}_t] - Y_t).$$

We may project this equation under the σ-field generated by Y_t and W_t, only having then to solve:

$$\mathbb{E}[h_t|\mathcal{Y}_t \vee \mathcal{Z}_t] = \theta(\mathbb{E}[\mu(Z_t)|\mathcal{Y}_t \vee \mathcal{Z}_t] - Y_t).$$

This construction may not be generalized if the volatility depends on Z_t because $\mathbb{E}[\sigma(Y_t, Z_t)dB_t|\mathcal{Y}_t \vee \mathcal{W}_t]$ does not cancel. The change of time models we will present in the next section, will solve this problem as it will be shown in paragraph 2.3.4.5. An other approach may be however adapted in the same direction as the DMIV decomposition. We briefly introduce this approach which will be treated in an other paper.

Let us start with a structural model (z fixed or assigned).

$$dY_t = \lambda(Y_t, z)dt + \sigma(Y_t, z)dB_t$$

which is assumed to be stationary and Z, the endogenous element is assumed to be not time-dependent. Following a method[2] presented by Aït-Sahalia (2002), let us introduce the transformation $\tilde{Y}_t = \gamma(Y_t, z) = \int_0^{Y_t} \frac{du}{\sigma(u,z)}$. This leads to the equation:

$$d\tilde{Y}_t = \mu(\tilde{Y}_t, z)dt + dB_t$$

where:

$$\mu(\eta, z) = \frac{\lambda(\gamma^{-1}(\eta, z), z)}{\sigma(\gamma^{-1}(\eta, z), z)} - \frac{1}{2}\frac{\partial \sigma}{\partial u}(\gamma^{-1}(\eta, z), z)$$

The model may be completed by two assumptions:

$$\mathbb{E}[dB_t|\mathcal{Y}_t \vee \mathcal{W}_t] = 0$$

$$\mathbb{E}[(dB_t)^2|\mathcal{Y}_t \vee \mathcal{W}_t] = 1$$

which are satisfied in particular if dB_t is independent of $\mathcal{Y}_t \vee \mathcal{W}_t$. This two equations may be used to characterize λ and σ. The main difficulty is coming from the fact that $\tilde{Y}_t$ depends on the parameters and this type of model may be viewed as a dynamic extension to transformation models. In fact, such a change in variables is originally closely related to the transformation established by Doss (1977), whose principle is also used by Detemple et al. (2005) for representation of Malliavin derivatives. We also refer to a recent work of Park (2010) that tries to estimate a conditional mean model in continuous time. This paper, where endogeneity is not tackled, is more focused on estimation, but a time change is also used to get a unit-volatility Brownian residual term.

2.3 The non-separable case : the time-change models

The DMIV decomposition is not sufficient to cover models like counting processes models, or diffusion models with volatility dependent on endogenous variables. We need to propose an other concept for instrumental variables analysis, which we will extend to dynamic models in the non-separable case (treated in the static case by e.g. Horowitz and Lee (2007)). In order to motivate our presentation, we start by the basic example of duration models.

2.3.1 Duration models: a motivating example

Let τ a be duration, i.e. a positive random variable. The distribution of τ is characterized by its survivor function $S(t) = \mathbb{P}(t \leq \tau)$ assumed to be differentiable. Let $\lambda(t)$ denotes the hazard function i.e. (that is $\lambda = -S'/S$) and $\Lambda(t)$ the integrated hazard function ($\Lambda(t) = \int_0^t \lambda(s)ds = -ln(S(t))$). We assume that Λ is strictly increasing. Such a duration model has a counting process representation

[2]We thank Nour Meddahi and Eric Renault for very helpful discussions and comments on this topic.

through the process $N_t = \mathbf{1}\{t \geq \tau\}$. This process is a sub-martingale and then a semi-martingale that may be represented w.r.t. the filtration generated by the history of N_t through:

$$N_t = \int_0^t \lambda(s)\mathbf{1}(s < \tau)ds + M_t. \tag{2.5}$$

The intensity of N_t (relatively to its history) is equal to $\lambda_t \mathbf{1}\{\tau > t\} = \lambda_t(1 - N_{t^-})$ (see e.g. Karr (1991)). A fundamental property we will use in the following is that $\Lambda(\tau)$ has an exponential distribution with parameter 1. Then, if U_t is the counting process $\mathbf{1}\{t \geq \Lambda(\tau)\}$ we have:

$$U_t = \int_0^t \mathbf{1}\{s < \Lambda(\tau)\}ds + M_t^U \tag{2.6}$$

because the hazard function of the exponential is constant equal to 1. Equivalently, these relations imply that:

$$N_{\Lambda^{-1}(t)} = U_t \tag{2.7}$$

and the given N becomes the process U via a change of time.

We want now to introduce a random endogenous factor Z in the duration model and an instrument W. For sake of simplicity, both Z and W are not time-dependent in this paragraph. An important literature analyzes endogenous variables in duration models (see van den Berg (2000)) and is in particular motivated by treatment models where outcomes are durations (see Abbring and van den Berg (2003b)). Our approach does not depend on any specific statistical models and extends the instrumental variable analysis to this problem. It is natural to assume that the integrated hazard function Λ becomes a function $\Lambda(t, Z)$ of Z (also noted $\Lambda_t(Z)$); the "noise" of the model, equal to $\Lambda_\tau(Z)$, is assumed to be independent of the instruments W, and has an exponential distribution with parameter 1. The model may be written in the usual way:

$$\tau = \Phi(U, Z) = \Lambda^{-1}(U, Z) \tag{2.8}$$

where $\Lambda(., Z)$ is strictly increasing, $U \perp\!\!\!\perp W$ and the distribution of U is given. This model becomes an example of non-separable IV model and generates a non-linear integral equation which characterizes Λ or equivalently $\Phi_t(Z) = \Lambda_t^{-1}(Z)$. Let us consider the following function :

$$S(t, z|w) = \frac{\partial}{\partial z}\mathbb{P}(\tau \geq t, Z \leq z | W = w) \tag{2.9}$$

which may be seen as the joint survivor of τ and density of Z conditionally on $W = w$, identified by the joint observation of (τ, Z, W). Then the independence condition between U and W implies:

$$\int_Z S(\Phi_t(Z), z|w)dz = P(U \geq t) = e^{-t} \tag{2.10}$$

because U is exponential with parameter 1. We will discuss later the identification of Φ i.e. the unicity of the solution of this equation.

We may wish to apply the DMIV decomposition to the N_t process considering the two filtrations $\mathcal{N}_t \vee \mathcal{Z}$ and $\mathcal{N}_t \vee \mathcal{W}$, generated by the history of N_t and respectively the endogenous and the instrumental variable. We then obtain a decomposition $N_t = \tilde{\Lambda}_t + \tilde{U}_t$ where $\tilde{\Lambda}_t = \int_0^t \tilde{\lambda}(s,z)ds$ and $\tilde{\lambda}(s,z) = \tilde{\lambda}_0(s,z)\mathbf{1}\{\tau > t\}$. In this context, the function $\tilde{\lambda}_0(s,z)$ should then become the solution of:

$$\frac{f_\tau(t|W)}{S_\tau(t|W)} = \int \tilde{\lambda}_0(s,z) f_\tau(z|w,\tau \geq t)dz \tag{2.11}$$

where the left hand-side is the hazard function of τ given $W = w$ (in that case $h_t = \frac{f_\tau(t|W)}{S_\tau(t|W)}\mathbf{1}\{\tau > t\}$) and $f_\tau(z|w,\tau \geq t)$ is the conditional density of Z given $W = w$ and the event $\{\tau \geq t\}$. However the $\tilde{\lambda}_0(s,z)$ function is the derivative of $\tilde{\Lambda}_t$ but not the derivative of $\Lambda(t,z)$ we have introduced above, and is not in general the hazard rate of the counting process associated to the duration.

The counting process version of the non-separable model (2.8) follows from the previous remarks. We may consider $N_t = \mathbf{1}\{\tau \leq t\}$ and assume that there exists a time-change function $\Phi_t(Z)$ strictly increasing and depending on Z such that $N_{\Phi_t(Z)} = U_t$ where U_t is a counting process associated to an exponential distribution of parameter 1 and such that U_t is independent of W. We will see later that these assumptions generates a non-linear integral equation deriving from semi-martingale decompositions which is equivalent in this particular case to Equation (2.10).

2.3.2 Time-change models

We use the notations introduced at the beginning of Section 2.2. We consider a stochastic process Y_t and two filtrations $\mathcal{F}_t = \mathcal{Y}_t \vee \mathcal{Z}_t$ (the "endogenous filtration") and $\mathcal{G}_t = \mathcal{Y}_t \vee \mathcal{W}_t$ (the "instrumental filtration") such that Y_t is adapted to both. We also introduce $\mathcal{H}_t = \mathcal{F}_t \vee \mathcal{G}_t$ generated by the three processes, Y_t, Z_t and W_t.

Definition 2.3.1. The process Y_t has an instrumental variable non-separable representation if there exists a stochastic process Φ_t such that:

1. $(\Phi_t)_t$ is an increasing sequence of stopping times relatively to the filtration $\mathcal{F}$;
2. $(Y_{\Phi_t})_t$ (the process Y stopped at time Φ_t) is equal to a process U_t independent of the W_t process;
3. U_t is a semi-martingale w.r.t. to its own history ($U_t = H_t^U + M_t^U$) with a given compensator H_t^U.

□

Remember that the property that for $t \geq 0$, Φ_t is a stopping time w.r.t. $\mathcal{F}$ means that $\forall s \geq 0$, $\{\Phi_t \leq s\} \in \mathcal{F}_s$. In the introducing example of the duration model of Section 2.3.1, Z is not time-dependent and this property only means that Φ_t is measurable w.r.t. Z for any t. The property that Φ_t is a $\mathcal{F}$-stopping time formalizes the idea that Φ_t only depends on Z and on the past of Y but not on W. However, Assumption (1) of Definition 2.3.1 implies that Φ_t is also a stopping time for the filtration $\mathcal{H}_t$. An important literature exists in abstract probability theory about the increasing sequences of stopping times and about the properties of processes stopped at these stopping times and the authors usually look at the properties (martingale, local martingale, ...) preserved by the change of time. Examples of this (not very recent) literature are Kazamaki (1972), El-Karoui and Weidenfeld (1977), El-Karoui and Meyer (1977), LeJan (1979).

2.3.3 Identification

Our objective is now to characterize the function Φ_t (depending also on the Z_t process) from objects identified through the joint process (Y_t, Z_t, W_t). We adopt a strategy based on the decomposition of the Y_t process w.r.t. the larger filtration $\mathcal{H}$.

Theorem 2.3.1. *Let us assume that:*

1. *Y_t is a semi-martingale w.r.t. filtration $\mathcal{H}$ and that we have:*

$$dY_t = k_t dt + dE_t$$

where $K_t = \int_0^t k_s ds$ is an $\mathcal{H}_t$-predictable process and E_t is an $\mathcal{H}_t$ martingale.

2. *Y_t has an instrumental variable non-separable representation as defined in Definition 2.3.1 when Φ_t is assumed to be continuous and differentiable (possibly except at a discrete set of points).*

3. *The distribution of $(Z_t)_t$, conditionally on σ-fields $\mathcal{Y}_s \vee \mathcal{W}_s$ for any s, is dominated by a measure Q and has a density denoted $g(z|\mathcal{Y}_s \vee \mathcal{W}_s)$.*

Then:

$$\int Q(dz) \int_0^{\Phi_t} k_s g(z|\mathcal{Y}_s \vee \mathcal{W}_s) ds = H_t^U \tag{2.12}$$

□

This equation shows that Φ_t is the solution of a non-linear integral equation where the right-hand side term is given and all the left-hand side (k and g) are identified by the distribution of the process (Y_t, Z_t, W_t). We assume that the model is well specified or equivalently that a solution exists to the Equation (2.12). The identification question is concerned with the unicity of the solution. As the problem is non-linear it is natural to look at local unicity of the solution. Let us assume that Φ_t is the true process and we compute the Gateaux-derivative of the left hand-side, taken in Φ_t, in direction of a function $\tilde{\Phi}_t$: $T'_{\Phi_t}(\tilde{\Phi}_t)$. We get obviously for any t:

$$T'_{\Phi_t}(\tilde{\Phi}_t) = \int \tilde{\Phi}_t k_{\Phi_t} g(z|\mathcal{Y}_{\Phi_t} \vee \mathcal{W}_{\Phi_t}) Q(dz).$$

We note that $T'_{\Phi_t}(\tilde{\Phi}_t)$ is linear and we assume that it is equal to the Frechet-derivative. Local unicity is then obtained through the condition :

$$T'_{\Phi_t}(\tilde{\Phi}_t) = 0 \quad \Rightarrow \quad \tilde{\Phi}_t = 0 \quad a.s. \tag{2.13}$$

If Φ'_t (the derivative w.r.t. t) does not cancel, this implication is true as soon as:

$$\int R_s k_s g(z|\mathcal{Y}_s \vee \mathcal{W}_s) Q(dz) = 0 \quad \Rightarrow \quad R_s = 0 \quad a.s.$$

where $R_s = \tilde{\Phi}_{\Phi^{-1}(s)}$.

2.3.4 Examples

2.3.4.1 Duration model with constant covariates

We take here the example of Section 2.3.1 in the case where variables $Z_t = Z$ and $W_t = W$ are fixed and known at time-origin. We have $Y_t = \mathbf{1}\{t \geq \tau\}$, and we suppose that there exists a sequence $\Phi_t(Z)$ of stopping-times such that $Y_{\Phi_t(Z)} = U_t$ with $U_t = \mathbf{1}\{t \geq U\}$ where U follows an exponential of parameter 1, $U \perp\!\!\!\perp W$. In this framework, we want to use Equation (2.12) of Theorem (2.3.1). In this context, k_s is the intensity of Y_t w.r.t. to $\mathcal{H}_t$ (with $\mathcal{H}_t$ equal here to $\sigma(Y_t, Z, W) = \sigma(\{\tau \geq t\}, Z, W)$). Equivalently, $g(z|\mathcal{Y}_s \vee \mathcal{W}_s) = g(z|\tau \geq t, W)$. As Z is a random variable and not a process, $\int Q(dz)$ will then be an integral over the support of Z relatively to the Lebesgue measure. As U is an exponential variable, the compensator of $U_t = \mathbf{1}\{t \geq U\}$ is trivially equal to $H_t^U = t \wedge U$. If we apply Theorem 2.3.1, we get:

$$\int dz \int_0^{\Phi_{U \wedge t}(z)} k(s|\tau \geq s, z, W) g(z|\tau \geq s, W) ds = U \wedge t. \tag{2.14}$$

As we will work with a fixed, arbitrary t, we can therefore conceptually eliminate U in all calculations and replace $U \wedge t$ with t. We already had the result of Equation (2.10) and we want to show that it leads to the same equation than Equation (2.14). Now, we write $f(t, Z|W)$ the joint law of (τ, Z) conditional on W. Having $\tau = \Phi_u(Z)$, if we note $g(U, Z|W)$ the joint law of (U, Z) conditional on W, we have:

$$g(U, Z|W) = \Phi'_U(Z) \times f(\Phi_U(Z), Z|W).$$

Our main assumption was that $U = \Lambda(\tau, z) \sim \mathcal{E}xp(1)$ conditionally on W. Then, this leads to :

$$e^{-U} = \int g(U, z|W) dz = \int \Phi'_U(z) f(\Phi_U(z), z|W) dz.$$

Then we have the two following expressions, holding $\forall u \geq 0$:

$$\begin{cases} \int \Phi'_u(z) f(\Phi_u(z), z|W) dz &= e^{-u} \\ \int S_\tau(\Phi_u(z), z|W) dz &= e^{-u}. \end{cases}$$

If we divide the first equation by the second, we get:

$$\begin{aligned} 1 &= \int \frac{\Phi'_u(z) f(\Phi_u(z), z|W)}{\int S_\tau(\Phi_u(z'), z'|W) dz'} dz \\ &= \int \Phi'_u(z) \underbrace{\frac{f(\Phi_u(z), z|W)}{S_\tau(\Phi_u(z), z|W)}}_{I_1} \times \underbrace{\frac{S_\tau(\Phi_u(z), z|W)}{\int S_\tau(\Phi_u(z'), z'|W) dz'}}_{I_2}. \end{aligned}$$

I_1 is the hazard function of the process $\{Y_t\}$ taken in $\Phi_u(z)$ conditional on $Z = z$, W. Indeed:

$$\frac{f(t, z|W)}{S(t, z|W)} = \frac{f(t|z, W) f(z|W)}{S(t|z, W) f(z|W)} = \frac{f(t|z, W)}{S(t|z, W)} = k(t|z, W).$$

I_2 is the law of Z conditional to W and $U \geq u$. Finally:

$$\int \Phi'_u(z) k(\Phi_u(z)|Z = z, W) g(z|U \geq u, W) dz = 1.$$

If we integrate in u for u varying from 0 to t we get:

$$\int_0^t du \int_z \Phi'_u(z)k(\Phi_u(z)|Z=z,W)g(z|W,U \geq u)dz = t.$$

If we commute the integral terms and make the change of variable $s = \Phi_u(z)$, and remark that $\{U \geq u\}$ is equivalent to $\{\Phi_U(z) \geq \Phi_u(z)\} = \{\tau \geq s\}$, then we recover Equation (2.14):

$$\int dz \int_0^{\Phi_t(z)} k(s|\tau \geq s, z, W)g(z|\tau \geq s, W)ds = t.$$

2.3.4.2 An example of duration model with process covariate

Let $N_t = \mathbf{1}(t \geq \tau)$ the explained process associated to the duration τ and Z_t be an endogenous covariate process assumed to be a jump process: $Z_t = \mathbf{1}\{t \geq \epsilon\}$. The process Z_t may be a treatment equal to 0 up to a random time ϵ and to 1 after. The structural model may be interpreted in the following way: if Z_t is "fixed" or assigned we assume that N_t has a structural hazard function equal to $\lambda_t = \alpha + \beta Z_t$ with $\alpha, \beta > 0$ and its compensator is $\Lambda_t = \int_0^t \lambda_s(Z)ds = \alpha t + \beta(t-\epsilon)\mathbf{1}\{t \geq \epsilon\}$ and then:

$$\Phi_t(Z) = \Lambda_t^{-1}(Z) = \frac{t}{\alpha}\mathbf{1}\{t < \alpha\epsilon\} + \frac{t+\beta\epsilon}{\alpha+\beta}\mathbf{1}\{t \geq \alpha\epsilon\}$$

which is an increasing sequence of stopping times adapted[3] to $\mathcal{Z}_t$.

In that case the model is parametric and the structural parameters are α and β. Let us now consider an instrument constant in time W and we assume that $N_{\Phi_t} = \mathbf{1}\{t \geq u\}$ with $u \sim Exp(1)$ and $U \perp\!\!\!\perp W$. Now consider ρ the hazard rate of τ given the Z_t process and the W variable. We have:

$$\rho(t) = \rho_1(t|\epsilon, W)\mathbf{1}\{\epsilon \leq t\} + \rho_2(t|\epsilon \geq t, W)\mathbf{1}\{\epsilon > t\}$$

where ρ_1 and ρ_2 are the hazard rates of τ given W and respectively ϵ or $\epsilon \geq t$. Then α and β are characterized as the solution of:

$$\int d\epsilon \int_0^{\Phi_t} \rho(s)g(\epsilon|\tau \geq s, W)ds = t \tag{2.15}$$

where $g(\epsilon|\tau \geq s, W)$ is the conditional density of ϵ given $\tau \geq s$ and W. In this equation ρ and g are identified and α and β follows from the resolution of Equation (2.15).

2.3.4.3 Counting process with endogenous cofactor

Let us assume that Y_t is a counting process, i.e. a process valued in $\mathbb{N}$ such that $Y_0 = 0$ and with càdlàg trajectories which are step functions having jumps of size 1 i.e. there exists a sequence of (τ_j) such that:

$$Y_t = \sum_{j \geq 1} \mathbf{1}\{t \geq \tau_j\}.$$

[3] Indeed let's consider for a given t the event $E = \{\Phi_t(Z) \leq s\}$. If $s \geq \epsilon$, $\mathcal{Z}_s = \sigma\{\epsilon\}$ and then $E \in \mathcal{Z}_s$ for any s. If $s < \epsilon$, $\mathcal{Z}_s = \sigma\{\mathbf{1}(\epsilon > s)\}$. In that case if $t < \alpha\epsilon$, E is always true and if $t \geq \alpha\epsilon$, E is always false.

If Z is assumed first to be fixed or assigned at a value $Z = z$ the process Y_t is modelled by its stochastic intensity $\lambda_t(z)$ or by its compensator $\Lambda_t(z) = \int_0^t \lambda_s(z)ds$. It is clear that if $\Lambda_t(z)$ is invertible and if we define:

$$\Phi_t(z) = \Lambda_t^{-1}(z)$$

the process $Y_{\Phi_t(z)} = U_t$ is an homogenous Poisson process. Indeed we have the decomposition:

$$Y_t = \Lambda_t(z) + M_t$$

and

$$Y_{\Phi_t(z)} = t + M_{\Phi_t(z)}.$$

Therefore the compensator is equal to t, which fully characterizes the Poisson process.

If Z is now randomly generated but not necessarily independent of U_t but if U_t is independent of W, we face the situation described in Definition 2.3.1 . We limit ourself in the following to the case where Z and W are time-independent for sake of simplicity.

We first rewrite in that case the integral equation characterizing $\Phi_t(z)$. Note that the intensity k_t verifies $k_t = \sum_{j\geq 1} k_t^{(j)} \mathbf{1}\{\tau_{j-1} \leq t < \tau_j\}$ with:

$$k_t^{(j)} = \frac{f_j(s + \tau_{j-1}|W, z, \tau_1, \ldots, \tau_{j-1})}{S_j(s + \tau_{j-1}|W, z, \tau_1, \ldots, \tau_{j-1})}$$

where f_j and S_j are the density and the survivor function of the (difference in) durations $\tau_j - \tau_{j-1}$ conditional to W, Z, and the past of the durations. The equation (2.12) becomes the following sequence of integral equations:

$$\begin{aligned}\sum_{l=1}^{j} \int dz \int_0^{\tau_l - \tau_{l-1}} \frac{f_l(s + \tau_{l-1}|W, z, \tau_1, \ldots, \tau_{l-1})}{S_l(s + \tau_{l-1}|W, z, \tau_1, \ldots, \tau_{l-1})} g(z|W, \tau_1, \ldots, \tau_{l-1}, \tau_l > s + \tau_{l-1})ds &\quad + \\ \int dz \int_0^{\Phi_t(z) - \tau_j} \frac{f_j(s + \tau_{j-1}|W, z, \tau_1, \ldots, \tau_j)}{S_j(s + \tau_{j-1}|W, z, \tau_1, \ldots, \tau_j)} g(z|W, \tau_1, \ldots, \tau_{j-1}, \tau_j > s + \tau_j)ds &\quad = \quad t\end{aligned} \tag{2.16}$$

One may add that

$$\frac{f_l(s + \tau_{l-1}|W, z, \tau_1, \ldots, \tau_{l-1})}{S_l(s + \tau_{l-1}|W, z, \tau_1, \ldots, \tau_{l-1})} g(z|W, \tau_1, \ldots, \tau_{l-1}, \tau_l > s + \tau_{l-1}) = \frac{f_l(s + \tau_{l-1}, z|W, \tau_1, \ldots, \tau_{l-1})}{S_l(s + \tau_{l-1}|W, \tau_1, \ldots, \tau_{l-1})}.$$

All the elements inside the integral may be estimated and this sequence of integral equations characterizes $\Phi_t(z)$ by intervals. Let us now analyze in more details the nature of the function $\Phi_t(z)$ and come back to the structural model where Z is fixed or assigned. In this structural model the $\lambda_t(z)$ function takes the form $\lambda_t(z) = \lambda_{t-\tau_{j-1}^{(j)}}(z$ for $t \in]\tau_{j-1}; \tau_j]$ where $\lambda_t^{(j)}(z)$ is the hazard rate of $\tau_j - \tau_{j-1}$ conditional on the past $(\tau_1, \ldots, \tau_{j-1})$ and given z. Then $\Lambda_t(z) = \int_0^t \lambda_s(z)ds$ which implies that $\Lambda_t(z) = \Lambda_{\tau_{j-1}}(z) + \int_0^{t-\tau_{j-1}} \lambda_{s-\tau_{j-1}}^{(j)}(z)ds$ if $\tau_{j-1} < t \leq \tau_j$. From this follows:

$$\text{if } \Lambda_{\tau_{j-1}}(z) < t \leq \Lambda_{\tau_j}(z) \text{ , then } \Phi_t(z) = \tau_{j-1} + (\Lambda_{t-\Lambda_{\tau_{j-1}}(z)}^{(j)})^{-1}(z)$$

where $\Lambda_t^{(j)}(z)$ is the integral of $\lambda^{(j)}(z)$.

In practice, $\Phi_t(Z)$ will be selected such that some properties are satisfied in the model when Z is fixed. For example, Y_t may be in that case an accelerated life non homogenous Poisson process i.e.:

$$Y_t = F(\psi(Z)t) + M_t$$

where $\psi(Z)$ is a function depending on the variables Z and F is a baseline, cumulative function on $\mathbb{R}^+$. In that case we have obviously:

$$\Phi_t(Z) = \frac{F^{-1}(t)}{\psi(Z)}$$

depending on the functional parameters ψ and F. Note however that this assumption **does not imply** that Y_t given Z and W is a Poisson process.

An other example of structural modelling is given by the Hawkes process. Let us assume that for Z fixed Y_t is an Hawkes process whose intensity is:

$$\lambda_t(z) = \mu + \int_0^t g_z(t,s)dY_s$$

where the parameters are μ and g function of Z, t and s. For example g may take the semi-parametric form:

$$g_z(t,s) = e^{-\beta(Z)(t-s)}$$

where β is an unknown positive function of Z. More generally, Z may be a stochastic process and g may be equal to $g_z(t,s) = e^{-\beta(Z_t)(t-s)}$ or $e^{-\beta(Z_s)(t-s)}$. For simplicity, we concentrate our presentation to the case where Z is constant w.r.t. the time index. The compensator of Y_t for any fixed value of $Z = z$ is equal to:

$$\begin{aligned}
\Lambda_t(z) &= \mu t + \int_0^t du \int_0^u g_z(u,s)dN_s \\
&= \mu t + \int_0^t dN_s \int_s^t g_z(u,s)du \\
&= \mu t + \sum_{j=1}^{N_t} \int_{\tau_j}^t g_z(u,\tau_j)du\mathbf{1}\{t \geq \tau_j\}
\end{aligned}$$

The inverse function $\Phi_t(z) = \Lambda_t^{-1}(z)$ has not an explicit form but may be easily numerically computed if g is given and Theorem 2.3.1 gives the way to estimate $\Phi_t(z)$ and then $g_z(t,u)$. As in the Poisson case let us note that Y_t given Z and W is not in general an Hawkes process.

2.3.4.4 Markovian transition models

An other application could concern Markov processes with multiple states. We begin by considering a Markov process Y_t with two states $\{1,2\}$. We write I^Y the generator of Y and suppose that I^Y has the form $q_t(Z)I$ where I is the following matrix:

$$I = \begin{bmatrix} -1 & 1 \\ a & -a \end{bmatrix}$$

where $a \in \mathbb{R}_+^*$. We denote $Q_t(Z) = \int_0^t q_s(Z)ds$. Z is assumed here to be static, endogenous. We assume that there exists a change of time $\Phi(Z) = \Lambda^{-1}(Z)$ such that $Y_t = U_{\Lambda_t(Z)}$ where U_t is a homogenous Markov process with two states and with a generator I. We make the assumption that U_t is independent from given instruments W. In the following we will skip the indexation in Z for simplicity (Z will be assumed to be fixed or assigned). It is possible to show[4] that:

$$\Phi(t) = \Lambda^{-1}(t) = Q^{-1}\Big(\frac{1-e^{(1+a)t}}{1+a}\Big).$$

We verify easily that this function is increasing in t.

We now consider the counting processes $N^{12}(t)$ and $N^{21}(t)$ that jump when respectively the process Y_t jumps from state 1 to 2 conditional on the fact that Y_t is in 1, and when Y_t jumps from state 2 to 1 conditional on the fact that Y_t is in state 2. We remark that in the general case, Y_t conditional on Z, W has no reason to remain Markovian. We note k_s^{12} (respectively k_s^{21}) the intensity of N^{12} (resp. N_s^{21}) conditional on Z, W, $\mathcal{H}_t$. Applying Theorem 2.3.1 we get:

$$\int_z\int_0^{\Phi_t(z)} k_s^{12} g(z|W,\mathcal{Y}_t)ds = \int_0^t \mathbf{1}\{U_s = 1\}ds,$$
$$\int_z\int_0^{\Phi_t(z)} k_s^{21} g(z|W,\mathcal{Y}_t)ds = \int_0^t a\mathbf{1}\{U_s = 2\}ds.$$

These equations are not useful because U_t is not observed and this right-hand side cannot be computed. But dividing the second equation by a, summing both and remarking that for each s, $\mathbf{1}\{U_s = 1\} + \mathbf{1}\{U_s = 2\} = 1$, then we get:

$$\int_z\int_0^{\Phi_t(z)} (k_s^{12} + \frac{1}{a}k_s^{21})g(z|W,\mathcal{Y}_t)ds = t.$$

2.3.4.5 The diffusion model

Let us first consider a structural model which generates a zero-mean diffusion process for Z (assumed to be time-independent) fixed:

$$dY_t = \sigma(Y_t, Z)dB_t \tag{2.17}$$

We simplify our presentation by assuming σ independent from t. Let us consider the quadratic variation of Y_t:

$$\Lambda_t(Z) =< Y_t >= \int_0^t \sigma^2(Y_s, Z)ds.$$

[4] See Appendix .

We define $\Phi_t(Z)$ the inverse function of $\Lambda(Z)$ (which is invertible because σ is assumed not null for any Y_t). This function characterizes an increasing sequence of stopping times (the event $\Phi_t(Z) \leq s$ is equivalent to $t \leq \Lambda_s(Z)$, and only depends on the past of Y until s). The process:

$$Y_{\Phi_t(Z)} = U_t$$

is then a Brownian motion (see Protter (2003)). We now consider that Z is randomly generated and that W is an instrument. The model still assumes Equation (2.17) and that the process U is independent of the filtration $\mathcal{W}_t$ generated by W and the past of W. In order to characterize σ or Φ we applied Theorem 2.3.1 to the relation:

$$Y^2_{\Phi_t(Z)} = U_t^2. \tag{2.18}$$

The compensator of U_t^2 is equal to t. Let k the stochastic intensity of Y_t^2 w.r.t. $\mathcal{Z}_t \vee \mathcal{W}_t$. We have:

$$\int dz \int_0^{\Phi_t(z)} k_s g_s(z|\mathcal{W}_s) ds = t.$$

In this expression k and g are identifiable for the DGP and Φ is obtained by solving this nonlinear, integral equation.

This approach may be generalized by considering a process Z_t instead of a fixed value Z if we assume σ depending only on the past up to t of Z (e.g. $\sigma(Y_t, Z_t)$) and $\mathcal{W}_t$ may be a filtration generated by a process W_t and Y_t. Let us underline that even if the structural model 2.17 is a zero-mean diffusion, this is in general not the case for the process Y_t given the filtration $\mathcal{Z}_t \vee \mathcal{W}_t$ and even if k is identifiable, its estimate may be complex.

An other extension is to consider a structural model with drift; if Z is fixed or assigned we assume the model:

$$dY_t = m_Z(Y_t, Z)dt + \sigma(Y_t, Z)dB_t.$$

We consider the same stopping time as before and the sequence of equations:

$$Y_{\Phi_t(Z)} = \int_0^{\Phi_t(Z)} m(Y_s, Z)ds + U_t.$$

The parameters of the model are $\Phi_t(Z)$ and $m(Y_s, Z)$ and in the case where Z is random we assume that U_t is a Brownian motion independent of W_t. We then apply twice Theorem 2.3.1: we compute the stochastic intensities $k_s^{(1)}$ of $Y_t - \int_0^t m(Y_s, Z)ds$ w.r.t. $\mathcal{Z}_t \vee \mathcal{W}_t$ and $k_s^{(2)}$ of $\left[Y_t - \int_0^t m(Y_s, Z)ds\right]^2$ w.r.t. $\mathcal{Z}_t \vee \mathcal{W}_t$ also and we derive from Theorem 2.3.1 that:

$$\int dz \int_0^{\Phi_t(z)} k_s^{(1)} g_s(z|\mathcal{W}_s) ds = 0$$
$$\int dz \int_0^{\Phi_t(z)} k_s^{(2)} g_s(z|\mathcal{W}_s) ds = t.$$

The functional parameters σ and m are solution of this system of nonlinear equations.

2.4 Operators, identification and estimation

We try here to state the estimation problems arising when trying to solve Equation 2.3.1 with time-independent covariates in duration models (like Equation 2.10). This framework has also been developed in Florens (2005).

2.4.1 Operators

2.4.1.1 Definition of the problem

We first assume without loss of generality that Z and W have a compact support and take their values on $[0;1]^{dz}$ and $[0;1]^{dw}$. The object of interest is a function:

$$\phi : \left\{ \begin{array}{rcl} \mathbb{R}^+ \times [0;1]^{dz} & \to & \mathbb{R}^+ \\ (u,z) & \mapsto & \phi(u,z) \end{array} \right.$$

that we assume to belong to $L^2(U,Z)$ where:

$$L^2(U,Z) = \left\{ \psi : \mathbb{R}^+ \times [0;1]^{dz} \mapsto \mathbb{R} \quad \text{measurable} \quad \Big| \quad ||\psi||_{L^2(U,Z)} = \left(\int |\psi(u,z)|^2 f_U(u) f_Z(z) dudz \right)^{\frac{1}{2}} < +\infty \right\}.$$

$$L^2(U,W) = \left\{ \psi : \mathbb{R}^+ \times [0;1]^{dw} \mapsto \mathbb{R} \quad \text{measurable} \quad \Big| \quad ||\psi||_{L^2(U,W)} = \left(\int |\psi(u,w)|^2 f_U(u) f_W(w) dudw \right)^{\frac{1}{2}} < +\infty \right\}.$$

In the following we will work with a fixed $U = u$; $L^2(U,Z)$, $L^2(U,W)$ will be noted L^2_Z, L^2_W. We define:

$$T : \left\{ \begin{array}{rcl} L^2_Z & \to & L^2_W \\ \phi & \mapsto & T(\phi) \end{array} \right.$$

where $T(\phi) : (u,w) \mapsto (u,w) \to \int_z \int_0^{\phi(u,z)} a(s,z,w) dsdz - u$ with:

$$a(s,z,w) = \lambda_\tau(s|Z = z, W) g(z|\tau \geq s, w) = f_\tau(\phi(u,z), z|\tau \geq \phi(u,z), W).$$

Resolving $T(\phi) = 0$ is a nonlinear inverse problem.

2.4.1.2 Fréchet-derivative of the operator

We would like to check the Fréchet-differentiability of the problem. The Gateau-derivative of the operator is given by taking $\alpha = 0$ in:

$$\frac{\partial}{\partial \alpha} \Big[\int_z \int_{s=0}^{\phi(u,z)+\alpha\tilde{\phi}(u,z)} a(s,z,w) - u \Big].$$

Then the Gateau-derivative in ϕ is the operator:

$$\tilde{\phi} \mapsto T'_\phi(\tilde{\phi}) \quad = \quad \int_z \tilde{\phi}(u,z) a(\phi(u,z), z, w) dz.$$

We will give sufficient assumptions on function $a()$ in order to ensure that the Gateau-derivative is also the Fréchet-derivative. As we have $a(s,z,w) = f_\tau(\phi(u,z), z|\tau \geq \phi(u,z), W)$, conditions on $a()$ can be translated on conditions on the joint density $f_\tau(t,z,w)$ (see Florens (2005)).

Assumption 2.4.1. We assume that:

1. $a()$ satisfies a Hilbert-Schmidt condition ie:

$$\int_{u,w}\int_z sup_t \Big|\frac{a(t,z,w)}{f_Z(z)}\Big|^2 f_U(u)f_Z(z)f_W(w)dudwdz < +\infty.$$

2. $a(t,z,w)$ is continuous and continuously differentiable with respect to t for all z, w, fixed u, and that $\int_{u,w}\sup_{t,z}|a'(t,z,w)|dudw < +\infty$ where $a'(t,z,w)$ denotes the partial derivative of a with respect to the first argument.

Proposition 2.4.1. Let Assumption 2.4.1 be satisfied. Then the operator T is Fréchet differentiable at every point $\phi \in L^2_Z$, and the Fréchet derivative T'_ϕ is given by:

$$T'_\phi(\tilde{\phi}) = \int_z \tilde{\phi}(u,z)a(\phi(u,z),z,w)dz.$$

□

Proof: we only need to verify that:

- T'_ϕ is linear (which is straightforward);
- T'_ϕ is continuous for every ϕ;
- the mapping $\phi \mapsto T'_\phi$ is continuous on L^2_Z in a $||.||_{\mathfrak{L}(L^2_Z,L^2_W)}$ sense (where $\mathfrak{L}(L^2_Z,L^2_W)$ is the space of linear functions between L^2_Z and L^2_W*).

With Assumption 2.4.1, for any ϕ_0 if we pose $k(u,w,z) = \frac{a(\phi_0(u,z),z,w)}{f_Z(z)}$, $k(.)$ is then the kernel of the Hilbert-Schmidt operator T'_ϕ since:

$$T'_{\phi_0}\tilde{\phi} = \int_z k(u,w,z)\tilde{\phi}(z)f_Z(z)dz$$

with:

$$\int_u\int_w\int_z |k(u,w,z)|^2 f_Z(z)f_W(w)f_U(u)dudwdz.$$

Under those conditions, T'_{ϕ_0} is Hilbert-Schmidt and therefore compact, then continuous for any ϕ_0.

The second part of the assumption helps to obtain the third criterion and the continuity of $\phi \mapsto T'_\phi$ as for ϕ_0,$\phi_1 \in L^2_Z$

$$(T'_{\phi_0} - T'_{\phi_1})(\tilde{\phi}) = \int_z \tilde{\phi}(u,z)\left\{a(\phi_0(u,z),z,w) - a(\phi_1(u,z),z,w)\right\}dz.$$

□

2.4.1.3 Adjoint operator

We now compute the expression of the adjoint operator of the Fréchet derivative we just obtained. For functions $\tilde{\phi} \in L^2_Z$ and $\tilde{\psi} \in L^2_W$, this is the linear operator $T^{'*}_{\phi_0}$ from L^2_W to L^2_Z such that:

$$< T'_{\phi_0}(\tilde{\phi}), \tilde{\psi} >_{L^2_W} = < \tilde{\phi}, T^{'*}_{\phi_0}(\tilde{\psi}) >_{L^2_Z} .$$

Writing explicitly:

$$\begin{aligned}
< T'_{\phi_0}(\tilde{\phi}), \tilde{\psi} >_{L^2_W} &= \int_u \int_w (T^{'*}_{\phi_0}(\tilde{\phi})(u,z))(\tilde{\psi}(u,w)) f_U(u) f_W(w) du dw \\
&= \int_u \int_w \left\{ \int_z \tilde{\phi}(u,z) \lambda_T(\phi_0(u,z)|z,w) g(z|\tau \geq \phi_0(u,z), w) dz \right\} \\
&\quad \times (\tilde{\psi}(u,w)) f_U(u) f_W(w) du dw \\
&= \int_u \int_z \left\{ \int_w \tilde{\psi}(u,w) \lambda_T(\phi_0(u,z)|z,w) g(z|\tau \geq \phi_0(u,z), w) f_W(w) \right\} \\
&\quad \times (\tilde{\phi}(u,z)) f_U(u) dz du dw \\
&= \int_u \int_z \left\{ \int_w \tilde{\psi}(u,w) \lambda_T(\phi_0(u,z)|z,w) g(w|\tau \geq \phi_0(u,z), z) \right\} \\
&\quad \times f_Z(z))(\tilde{\phi}(u,z)) f_U(u) dz du dw \\
&= \int_u \int_z \left\{ \int_w \tilde{\psi}(u,w) \lambda_T(\phi_0(u,z)|z,w) g(w|\tau \geq \phi_0(u,z), z) du dw \right\} \\
&\quad \times (\tilde{\phi}(u,z)) f_Z(z) f_U(u) dz \\
&= < \tilde{\phi}, T^{'*}_{\phi_0}(\tilde{\psi}) >_{L^2_Z}
\end{aligned}$$

where T'_{ϕ_0} is consequently defined by:

$$\boxed{T^{'*}_{\phi_0}(\tilde{\psi}) = \int_w \tilde{\psi}(u,w) \lambda_T(\phi_0(u,z)|z,w) g(w|\tau \geq \phi_0(u,z), z) dw.}$$

2.4.2 Ill-posedness

As underlined by Proposition 10.1 of Engl et al. (1996), **the characterization of the ill-posedness of an operator through conditions on its linearization is sometimes difficult and no general conditions can be given**. We could use this proposition or its local version given in Gagliardini and Scaillet (2009), and try to work directly on T. Alternatively, one can try to show that T' computed on the true solution is compact, which is ensured by condition 2.4.1. This condition will be sufficient only in the case of infinite dimension of the range of T' (which is easily verified since the arrival space of T'_ϕ for $\phi \in L^2_Z$ is L^2_W). An other approach could be to show that the image $T'_\phi(S)$ of bounded sets S is relatively compact (i.e. the closure of $T'_\phi(S)$ is also compact)and use for this the characterization of Alt (2006) used by Gagliardini and Scaillet (2009).

2.4.3 Estimation

Estimation is not the purpose of this work. We give however here some insights on how the objects of interest (operators, functions) could be estimated. Suppose that a sample (τ_j, z_j, w_j) of size N is available. For instance $\hat{g}(z|\tau \geq t, w)$ may be estimated with classical kernel density estimation such as:

$$\hat{g}(z|\tau \geq t, w) = \frac{1}{h_z} \sum_{j=1}^{N} \frac{K(z - z_j)\bar{K}(t - \tau_j)K(w - w_j)}{\sum_{j'=1}^{N} \bar{K}(t - \tau_{j'})K(w - w_{j'})}.$$

$K()$ is here taken as the standard notation for an appropriated kernel, $\bar{K}$ its survivor version, and h the appropriate bandwidth adapted to variable Z. Despite the generic notation, kernels K may depend on z, w, etc. Concerning the intensity, parametric and nonparametric estimation procedures are described in Appendix A.3. For instance, a Cox model could be helpful using $\lambda(t|z, w) = \lambda_0(t) exp(\beta' z + \gamma' w)$.

We need to express $T\phi_0$, $T'_{\phi_0}\phi$ for given $\phi_0, \phi \in L^2_Z$. In the following, as we will work with fixed u, we will omit u in the expressions of the estimated quantities. To obtain an estimator for ϕ, this assumes that for each z_j in the sample, a corresponding value $\phi^{(j)}$ is available. Such a value is not observed and is only obtained by the resolution algorithm (and potentially updated). With a given set $(z_j, \phi^{(j)})$, with $\phi^{(j)} = \phi(u, z_j)$, ϕ is estimated through:

$$\hat{\phi}(z) = \sum_{j=1}^{N} \frac{\phi^{(j)} K(z - z_j)}{\sum_{j=1}^{N} K(z - z_j)}.$$

With two estimators $\hat{\lambda}$ and $\hat{g}$ of λ and g, we have then:

$$\hat{T}(\hat{\phi}_0) = \int_z \int_{t=0}^{\hat{\phi}_0(z)} \hat{\lambda}(\hat{\phi}_0(z)|z, w)\hat{g}(z|\tau \geq t, w)dz.$$

$$\hat{T}'_{\hat{\phi}_0}\tilde{\phi} = \int_z \sum_{j=1}^{N} \tilde{\phi}(z)\hat{\lambda}(\hat{\phi}_0(z)|z, w)\hat{g}(z|\tau \geq \hat{\phi}_0(z), w)dz.$$

2.5 Regularization

We now turn to the study of the regularization aspects of the problem. In the following, we will note $\hat{T} = \hat{T}_n$ an estimator of T. Generically, $\hat{T}$ would be defined through:

$$\hat{T} : \phi \to \int_z \int_{t=0}^{\phi(s,z)} \hat{a}(z, W, t)\ dzdt - s.$$

In our problem, the operator is nonlinear and we cannot assume that we know quantities such as singular values or functions. T and even T' are not known (as our object of interest appears in the expression of T'). This implies the use of a regularization technique: we present the regularization in Hilbert Scales and the study of Tikhonov regularization as a particular case, and regularization by an iteration method (Landweber-Fridman).

2.5.1 Regularization in Hilbert scales

2.5.2 Problem definition

We recall that we want to solve in ϕ the following problem:

$$\min_{\phi} ||\hat{T}(\phi)||^2 + \alpha_n||\phi - \phi^*||_s^2 \quad \text{with} \quad \phi(u,z) \in L_Z^2 \tag{2.19}$$

where α_n is a regularization parameter and ϕ^* is an arbitrary function. In particular, we will have to control the behavior of the parameter α_n as n goes to infinity. We will adapt the approach of Engl et al. (1996) p.245 although we face a different problem. Additional conditions on the Fréchet-derivative of T will be needed to obtain the speed of convergence of our solution. We first examine the convergence of the sequence of solutions of problem 2.19 towards the true solution of the initial problem $T(\phi) = 0$.

Assumption 2.5.1. Assume that:

- if ϕ_0 is a solution of the problem $T(\phi) = 0$, then there exists a sequence δ_n such as
$$||\hat{T}(\phi_0) - T(\phi_0)|| = ||\hat{T}(\phi_0)|| \leq \delta_n;$$
- δ_n, α_n, and δ_n^2/α_n tends to 0 as n increases to infinity;
- we assume that $\hat{T}$ is an estimator such that $||\hat{T} - T|| \to 0$.

We can show that under Assumption 2.5.1 we have:

Lemma 2.5.1. *If $(\hat{\phi}_n^\alpha)$ is a sequence of solutions of the related minimization problems (2.19), then there exists a subsequence $(\hat{\phi}_{n,2}^\alpha)$ of $(\hat{\phi}_n^\alpha)$ which converges towards a function ϕ_l (in a L_W^2 sense). Moreover ϕ_l is a solution of the problem $T(\phi) = 0$.*

See proof in Appendix 2.7.4.

2.5.3 Unicity

Some conditions may be examined to ensure unicity of ϕ_l. If the problem $T(\phi) = 0$ is identified, ϕ_0 is unique and $\phi_0 = \phi_l$. As soon as ϕ_l is unique, we have that there is only one limit for any convergent subsequence of $(\hat{\phi}_n^\alpha)$, so $(\hat{\phi}_n^\alpha)$ is itself convergent and tends to ϕ_0. However, even when the initial problem is not identified, it is possible to restrict our problem to some classes of solutions. Engl et al. (1996) uses the concept of ϕ^*-minimal norm solutions. Then ϕ_0 is taken as the function, among the set of solutions ϕ of $T(\phi) = 0$, which minimizes the quantity $||\phi - \phi^*||_s$. Then, we would have that $\phi_0 = \phi_l$. Indeed:

$$\begin{aligned} ||\phi_l - \phi^*|| &\leq \limsup_{n\to\infty} ||\hat{\phi}_n^{\alpha_n} - \phi^*|| \\ &\leq ||\phi_0 - \phi^*|| \\ &\leq ||\phi_l - \phi^*||. \end{aligned}$$

The first inequality comes from the lower-semi continuity of the norm. The second comes from the definition of $(\hat{\phi}_n^\alpha)$, and the third from the fact that ϕ_0 is a ϕ^*-minimal norm solution and that ϕ_l is itself a solution. This ensures unicity as soon as the solution has been restricted to be a ϕ^*-minimal norm solution.

2.5.4 Speed of convergence

We suppose that we have a solution to the initial problem and that this solution is identified. We want to derive the speed of convergence of the solutions $(\hat{\phi}_n^\alpha)$ to the true solution ϕ_0 under this identification assumption. We mainly need conditions that are similar to those of Engl et al. (1996), with additional assumptions concerning the Fréchet derivative of $\hat{T}$ and T in ϕ_0.

Assumption 2.5.2. We suppose that:

- (i) - the problem $T(\phi) = 0$ is identified with a true solution ϕ_0
- (ii) - T and $\hat{T}$ are continuous and Fréchet-differentiable with convex domains;
- (iii) - there exists $C > 0$ such as $||\hat{r}_n|| \leq C||\hat{\phi}_n^\alpha - \phi_0||^2$ where $\hat{r}_n = \hat{T}(\hat{\phi}_n) - \hat{T}(\phi_0) - \hat{T}'_{\phi_0}(\hat{\phi}_n - \phi_0)$;
- (iv) - there exists γ_n such as $||\hat{T}'_{\phi_0} - T'_{\phi_0}|| \leq \gamma_n$;
- (v) - there exists $\beta \in \mathbb{R}$ such as $\phi_0 - \phi^* \in H_{-\beta}$ where $(H_s)_{s\in\mathbb{R}}$ is a Hilbert scale (source condition);
- (vi) - there exists a such as $||T'_{\phi_0}(\hat{\phi} - \phi_0)||^2 \sim ||\hat{\phi} - \phi_0||^2_{-a}$;
- (vii) - $a \leq s$ and $s \leq \beta \leq a + 2s$;
- (viii) - $\gamma_n^{2(a+s)/a}/\alpha_n \to 0$ when $n \to +\infty$.

Under Assumptions 2.5.1 and 2.5.2 we have the following lemma:

Lemma 2.5.2.

$$\begin{aligned} ||\hat{\phi}_n^\alpha - \phi_0||^2_{-a} + \alpha_n||\hat{\phi}_n^\alpha - \phi_0||^2_s \quad \leq \quad & \delta_n^2 + \alpha_n||\hat{\phi}_n^\alpha - \phi_0||_{2s-\beta} + ||\hat{\phi}_n^\alpha - \phi_0||_{-a}(\delta_n + ||\hat{\phi}_n^\alpha - \phi_0||^2) \\ & + \gamma_n||\hat{\phi}_n^\alpha - \phi_0||(\delta_n + ||\hat{\phi}_n^\alpha - \phi_0||_{-a} + ||\hat{\phi}_n^\alpha - \phi_0||^2) \end{aligned}$$

See the proof in Appendix 2.7.5.

At this point we obtain the same expression than in the case of nonlinear ill-posed inverse problems, see Engl et al. (1996)). But with the additional term related to γ_n and the convergence of the Fréchet-derivative taken on the true solution. If $\gamma_n = O(\delta_n)$ this term is likely to be negligible. However when it's not the case, this term has to be taken into account for the study of the speed of convergence of the solution. In our case, as T is estimated through two integrals, and T' only one, γ_n will probably be slower than δ_n and thus cannot be considered as systematically negligible. Yet, we will need in the following the next assumption:

Assumption 2.5.3. We suppose that:

$$\gamma_n^{\frac{a+\beta}{a}} \delta_n^{-1} \to 0$$

This assumption will help us to derive the speed of convergence of the solution. It appears that γ_n must not be too slow compared to δ_n and that there is a minimal power for γ_n, equal to $(a+\beta)/a$ which is greater than 1, to be at least faster than δ_n.

Lemma 2.5.3. *The best choice for α_n is:*

$$\alpha_n \sim \delta_n^{\frac{2a+2s}{a+\beta}}.$$

If moreover we make assumption 2.5.3, we get a speed of convergence equal to:

$$\mathcal{O}(\delta^{\frac{\beta-s}{a+\beta}}).$$

α_n is chosen according to the usual case and leads to a speed of convergence that is similar to the standard situation. However, if Assumption 2.5.3 is not verified, the terms in γ_n are too slow compared to the others, and the result does not hold, those terms driving then the speed of convergence of the sequential solutions towards the true solution.

2.5.5 Tikhonov regularization

We reformulate problem 2.19 with a traditional quadratic penalization. We study the new following minimization objective:

$$\min_{\phi} ||\hat{T}(\phi)||^2 + \alpha_n ||\phi - \phi^*||^2 \quad \text{for} \quad \phi(u,z) \in L_Z^2 \tag{2.20}$$

where α_n is the regularization parameter and ϕ^* an arbitrary function. The study of the convergence is not affected by this, since a L^2 norm is a particular case of the Hilbert space where $s = 0$. Then we make the same assumption than 2.5.1 and the demonstration is not affected by the change of penalization and result 2.5.1 still holds (see Appendix 2.7.7). However, the result concerning the speed of convergence is a bit different since we cannot simply replace s by 0 in the demonstration. However, we can make some similar assumptions to derive the speed of convergence, and replace Assumption 2.5.2 by:

Assumption 2.5.4. We suppose that:

- (i) - the problem $T(\phi) = 0$ is identified with a true solution ϕ_0;
- (ii) - T and $\hat{T}$ are continuous and Frechet differentiable with convex domains;
- (iii) - there exists $C > 0$ such as $||\hat{r}_n|| \leq C||\hat{\phi}_n^{\alpha} - \phi_0||^2$;
- (iv) - there exists γ_n such as $||\hat{T}'_{\phi_0} - T'_{\phi_0}|| \leq \gamma_n$;
- (v) - there exists w such as $\phi_0 - \phi_0^* = {T'_{\phi_0}}^*.w$;
- (vi) - $\gamma_n = o(\sqrt{\delta_n})$;
- (vii) - $2||w||C < 1$.

Under hypothesis 2.5.1 and 2.5.4, we can show that :

Lemma 2.5.4. *The best choice for α_n is:*

$$\alpha_n \sim \delta_n$$

and the resulting speed of convergence is :

$$||\hat{\phi}_n^\alpha - \phi_0|| = \mathcal{O}(\sqrt{\delta_n})$$
$$||\hat{T}(\hat{\phi})|| = \mathcal{O}(\delta_n).$$

2.5.6 Iterative regularization

Rather than defining sequences of minimization problems, we could also try to regularize our problem with an iteration method.

Generalities

The principle of an iterative method is to linearize the operator around the current solution, and to update it (see Kaltenbacher et al. (2008)). The interest is to compute linear versions of the operator rather than the initial one, and then to reduce computation costs. This is the idea of the Newton-type methods. If we have a solution ϕ_k at the iteration k, we could try to solve the linearized problem:

$$T'_{\phi_k}(\phi_{k+1} - \phi_k) = -T\phi_k. \tag{2.21}$$

At a given step k, $T'_{\hat{\phi}_k}\phi$ is a function of w. The matrix formulation of the problem stated in 2.21 is $A_k\phi = v_k$ where:

- ϕ is the vector of the $(\phi^{(j)})_{j\in[1;n]}$;
- v_k is the vector of the $((-T\phi_k)(w_j))_{j\in[1;n]}$;
- $A_k = (a_{ij})_{i,j}$ is the matrix such that:

$$a_{ij} = \int_z \frac{K(z - z_j)}{\sum_{j=1}^n K(z - z_j)} \hat{\lambda}(\hat{\phi}(z)|z, w_i)\hat{g}(z|\tau \geq \hat{\phi}(z), w_i)dz.$$

However in practice, this linear problem may also be ill-posed and has to be regularized. If the Tikhonov regularization is applied to solve 2.21, then this is the Levenberg-Marquardt method. With an additional penalty term to solve 2.21, this is the iteratively regularized Gauss-Newton method. Finally, the Levenberg-Marquardt method regularizes 2.21 with a regularization parameter α_k, modifying the iteration through:

$$\phi_{k+1} = \phi_k + (\alpha_k I + T'^*_{\phi_k}T'_{\phi_k})^{-1}T'^*_{\phi_k}(-T\phi_k).$$

A stopping criterion has to be chosen. The Tikhonov regularization parameter α_k has also to be controlled and depends on the iteration parameter k. When we use the iteratively regularized Gauss-Newton method, the iteration becomes:

$$\phi_{k+1} = \phi_k + (\alpha_k I + T'^*_{\phi_k}T'_{\phi_k})^{-1}T'^*_{\phi_k}(-T\phi_k) + \alpha_k(\phi^\dagger - \phi_k)$$

where $\phi^\dagger$ is an *a-priori* chosen function.

Landweber-Fridman regularization

We will study here in details a fourth method, the Landweber-Fridman regularization scheme. First, let's recall that the scheme of the Landweber iteration is given by:

$$\phi_{k+1}^{\delta} = \phi_k^{\delta} + (T'_{\phi_k^{\delta}})^*(y^{\delta} - T\phi_k^{\delta})$$

We assume first that:

Assumption 2.5.5. $T\phi = 0$ is solvable in $B(0;\rho)$.

Assumption 2.5.6. T is of the form $T = \tilde{T}\frac{f_U f_W}{\tau}$ where $\tilde{T}$ is an operator from $L^2(U,Z)$ to $L^2(U,W)$ and τ is a probability density. For $\phi \in L^2(U,Z)$, $T'_{\phi} = \tilde{T}'_{\phi}\frac{f_U f_W}{\tau}$. So is T'^*_{ϕ} the dual of T'_{ϕ}, $T'^*_{\phi} = \tilde{T}'^*_{\phi} \times \frac{f_U f_Z}{\pi}$ where $\tilde{T}'^*_{\phi}$ is also the dual of $\tilde{T}'_{\phi}$ and π a probability density. T'^*_{ϕ} is then defined from $L^2_{\Pi}(\mathbb{R}^+ \times \mathbb{R}^{dz})$ to $L^2_{\tau}(\mathbb{R}^+ \times \mathbb{R}^{dw})$. The norms used here are to be understood for the corresponding probability densities depending of the arrival space.

$\tilde{T}'_{\phi}$ and $\tilde{T}'^*_{\phi}$ must be understood for our problem as being again estimators of respectively:

$$T'_{\phi}(\tilde{\phi}) = \int_Z \tilde{\phi}(u,z)\lambda_{\tau}(\phi(u,z)|z,w)g(z|\tau \geq \phi(u,z),w)dudz$$

$$T'^*_{\phi}(\tilde{\psi}) = \int_W \tilde{\psi}(u,w)\lambda_{\tau}(\phi(u,z)|z,w)g(w|\tau \geq \phi(u,z),z)dudw.$$

Assumption 2.5.7. T is such that:

- (i) - $||T'_{\phi}|| \leq 1$ for $\phi \in B(0;2\rho) \subset \mathcal{D}(T)$;
- (ii) - $||T\phi - T\tilde{\phi} - T'_{\phi}(\phi - \tilde{\phi})|| \leq \eta ||T\phi - T\tilde{\phi}||$ with $\eta < \frac{1}{2}$ and $\phi, \tilde{\phi} \in B(0;2\rho) \subset \mathcal{D}(T)$;
- (iii) - $||T|| < +\infty$ henceforth T continuous.

Assumption 2.5.8. We have estimators $\hat{T}_n$ of T, $\hat{T}()'_n$ of $T()'$ and $\hat{T}'()^*_n$ of $T'()^*$ such that:

- (i) - $||\hat{T}_n - T|| \leq \delta_n$;
- (ii) - $||\hat{T}'^*_{n,\phi} - T'^*_{\phi}|| \leq \gamma_n$ for $\phi \in B(0;2\rho)$;
- (iii) - $||\hat{T}_n\phi - \hat{T}_n\tilde{\phi} - \hat{T}'_{n,\phi}(\phi - \tilde{\phi})|| \leq \eta ||\hat{T}_n\phi - \hat{T}_n\tilde{\phi}||$ with $\eta < \frac{1}{2}$ and $\phi, \tilde{\phi} \in B(0;2\rho) \subset \mathcal{D}(\hat{T}_n)$.

where δ_n and γ_n converges towards 0 and are assumed to be monotonously decreasing without loss of generality.

Assumption 2.5.9. Discrepancy Principle : $k_* = k_*(\delta_n, \hat{T}_n)$ is chosen such that the iteration is stopped for the first k_* such that:

$$|| - \hat{T}_n\hat{\phi}_{k_*}|| \leq \tau\delta_n < || - \hat{T}_n\hat{\phi}_k||$$

for $0 \leq k \leq k_*$ and $\tau > 1$ fixed, such as $\tau > 2\frac{1+\eta}{1-2\eta} > 2$.

Theorem 2.5.1. *Under Assumptions 2.5.5, 2.5.6, 2.5.7, 2.5.8, and with the stopping criterion proposed in 2.5.9, the nonlinear Landweber iteration with an unknown, estimated operator, converges to a solution of* $Tf = 0$.

A proof of Theorem 2.5.1 is given in Section 2.7.8

2.6 Conclusion

We have presented two classes of models for stochastic processes with endogenous variables treated with the instrumental variables method. Dynamic extension of separable models gives a generalization of the standard Doob-Meyer decomposition of semi-martingales and some probabilistic aspects of this model should be developed (extension for example to the case when martingales are only local). In the two kinds of approaches the functional parameters of interest are characterized as solutions of integral equations and their identification (unicity of the solution) is discussed. We have illustrated these concepts to many kinds if stochastic processes used in many fields of applied econometrics. All these examples need to be developed in connection with the infinitesimal generator.

This chapter only treats modelling and not the practical aspects and the theoretical properties of the inference. In practice, many objects we have introduced depend on infinite past and cannot be estimated under this form. We have introduced models where the specification is made on the structural form and reduced forms are implicitly left unconstrained for the estimation. Tractable approximations for the reduced form should be selected in order to implement the presented methods. In the considered cases, the parameters are solutions of ill-posed inverse problems and their statistical properties have to be analyzed.

2.7 Proofs

2.7.1 Proof of Theorem 2.2.1

Let us define U_t by $Y_t - \int_0^t \lambda_s ds$. As $\Lambda_t = \int_0^t \lambda_s ds$ is predictable by construction, we just have to proove that U_s satisfies condition **A2** of Definition 2.2.1. We have:

$$\begin{aligned} \mathbb{E}[U_t - U_s|\mathcal{Y}_s \vee \mathcal{W}_s] &= \mathbb{E}[Y_t - Y_s - \int_s^t \lambda_u du|\mathcal{Y}_s \vee \mathcal{W}_s] \\ &= \mathbb{E}[\int_s^t h_u du - \int_s^t \lambda_u du|\mathcal{Y}_s \vee \mathcal{W}_s] \end{aligned}$$

because $\mathbb{E}[M_t - M_s|\mathcal{Y}_s \vee \mathcal{W}_s] = 0$. We can commute the integration and the conditional expectation terms, and we get:

$$\int_s^t \mathbb{E}[h_u - \lambda_u|\mathcal{Y}_s \vee \mathcal{W}_s]du = \int_s^t \mathbb{E}[h_u - \mathbb{E}(\lambda_u|\mathcal{Y}_u \vee \mathcal{W}_u)|\mathcal{Y}_s \vee \mathcal{W}_s]du$$

because $\mathcal{Y}_s \vee \mathcal{W}_s \subset \mathcal{Y}_u \vee \mathcal{W}_u$ for each $s \leq u$. The second assumption allows then to conclude and to obtain the desired result $\mathbb{E}[U_t - U_s|\mathcal{Y}_s \vee \mathcal{W}_s] = 0$ and Y_t has a DMIV decomposition.

$\square$

2.7.2 Proof of Theorem 2.3.1

Let us start with the decomposition of Y_t w.r.t. $\mathcal{H}_t$:

$$Y_t = K_t + E_t. \tag{2.22}$$

We consider $(\mathcal{H}_{\Phi_t})_t$ the filtration where for any t, $\mathcal{H}_{\Phi_t}$ is the stopping-time sub σ-field of $\mathcal{H}_\infty$ associated to Φ_t, i.e. :

$$\mathcal{H}_{\Phi_t} = \sigma\{A \in \mathcal{H}_\infty | \{\Phi_t < s\} \cap A \in \mathcal{H}_s\}. \tag{2.23}$$

Note that $(\mathcal{H}_{\Phi_t})_t$ is a filtration because Φ_t is increasing. Equivalently (see Protter (2003) - Chap. I - Theorem 6):

$$\mathcal{H}_{\Phi_t} = \sigma\{\mathcal{Y}_{\Phi_t}, \mathcal{Z}_{\Phi_t}, \mathcal{W}_{\Phi_t}\}. \tag{2.24}$$

Then:

$$Y_{\Phi_t} = K_{\Phi_t} + E_{\Phi_t}$$

is the semi-martingale decomposition of $(Y_{\Phi_t})_t$ w.r.t. the filtration $(\mathcal{H}_{\Phi_t})_t$. This result follows from Proposition 1 of Kazamaki (1972) which implies that E_{Φ_t} remains a martingale w.r.t. $(\mathcal{H}_{\Phi_t})_t$ and K_{Φ_t} is predictable under our assumption $K_t = \int_0^t k_s ds$. The continuity condition of Φ_t is obviously satisfied under our assumptions. Under the model specification $Y_{\Phi_t} = U_t$ and :

$$\mathcal{H}_{\Phi_t} = \mathcal{U}_t \vee \mathcal{Z}_{\Phi_t} \vee \mathcal{W}_{\Phi_t} \tag{2.25}$$

where $\mathcal{U}_t$ is the σ-field generated by $(U_s)_{0 \leq s \leq t}$. Then the decomposition (2.24) rewrites :

$$U_t = K_{\Phi_t} + E_{\Phi_t} \tag{2.26}$$

and is also the semi-martingale decomposition of U_t w.r.t. $(\mathcal{U}_t \vee \mathcal{Z}_{\Phi_t} \vee \mathcal{W}_{\Phi_t})$. Equivalently Equation (2.26) becomes:

$$U_t = \int_0^t \Phi'_s k_{\Phi_s} ds + E_{\Phi_t} \tag{2.27}$$

where Φ'_t is the derivative w.r.t. t of Φ_t.

If Φ_t is not differentiable at some point, we partition the integral before and after the point. For simplicity we assume here Φ to be differentiable.

The next step is to derive from Equation (2.27) the decomposition of U_t w.r.t. the sub-filtration $(\mathcal{U}_t \vee \mathcal{W}_{\Phi_t})_t$. We have (see Karr (1991)):

$$U_t = \int_0^t \mathbb{E}[\Phi'_s k_{\Phi_s} | \mathcal{U}_s \vee \mathcal{W}_{\Phi_s}] ds + \tilde{E}_t$$

where $\tilde{E}_t$ is a martingale adapted to $(\mathcal{U}_t \vee \mathcal{W}_{\Phi_t})_t$. The computation of the conditional expectation inside the integral may be conventionally written as an integral w.r.t. a conditional density of Z_t process given $\mathcal{U}_s \vee W_{\Phi_s}$, noted $g(z|\mathcal{U}_s \vee \mathcal{W}_{\Phi_s})$:

$$U_t = \int_0^t ds \int \Phi'_s k_{\Phi_s} g(z|\mathcal{U}_s \vee W_{\Phi_s}) dz + \tilde{E}_t$$

We commute the integrals and, after a change of variable $v = \Phi_s$, we get:

$$U_t = \int_0^t Q(dz) \int_0^{\Phi_t} k_v g(z|\mathcal{Y}_v \vee W_v) dv + \tilde{E}_t.$$

Finally let us consider the decomposition of U_t w.r.t. its own filtration:

$$U_t = H_t^U + E_t^U.$$

As $(U_t)_t$ and $(W_t)_t$ are independent, (U_t) and (W_{Φ_t}), are also independent and this last decomposition is also the decomposition w.r.t. $(\mathcal{U}_t \vee \mathcal{W}_{\Phi_t})$. By unicity of the decomposition we get:

$$\int dz \int_0^{\Phi_t} k_v g(z|\mathcal{Y}_v \vee \mathcal{W}_v) dv = H_t^U.$$

□

2.7.3 Expression of Φ_t for Markov models with two states

By definition, we have that $\mathbb{P}[U_{t+s}|U_t] = e^{Is}$. We remark that I has for eigenvalues 0 and $-(1+a)$ and U writes $U = DLD^{-1}$ with D the matrix of eigenvectors which are respectively $(1;1)'$ and $(-1;1)'$. It follows that for $t, s \geq 0$:

$$\mathbb{P}[Y_{t+s}|Y_t] = \mathbb{P}[U_{\Lambda(t+s)}|U_{\Lambda(t)}] = e^{I(\Lambda(t+s)-\Lambda(s))}.$$

Matrix $e^{I(\Lambda(t+s)-\Lambda(s))}$ rewrites:

$$D \begin{bmatrix} 1 & 0 \\ 0 & e^{-(1+a)(\Lambda(t+s)-\Lambda(s))} \end{bmatrix} D^{-1}.$$

The generator I^Y of Y is then given by the derivative in s, taken in $s = 0$ of the former expression. That is:

$$I^Y = \begin{bmatrix} 0 & 0 \\ 0 & -(1+a)\Lambda'(t)e^{-(1+a)\Lambda(t)} \end{bmatrix} D^{-1} = \Lambda'(t)e^{-(1+a)\Lambda(t)} DLD^{-1} = \Lambda'(t)e^{-(1+a)\Lambda(t)} I.$$

Consequently, the generator matrix I^Y is of the form $q(t)I$ with $q(t) = \Lambda'(t)e^{-(1+a)\Lambda(t)}$. Then we have that

$$Q(t) = \frac{1}{1+a}\Big(1 - e^{-(1+a)\Lambda(t)}\Big).$$

The expression of Φ follows.

□

2.7.4 Proof of Lemma 2.5.1

We note $\hat{\phi}_n^\alpha$ the sequence of solutions of problems 2.19 (with ϕ^* an arbitrary function) and ϕ_0 a true solution of the initial problem. We want now to show that under Assumption 2.5.1, there exists a subsequence of solutions which converges to a function which is solution of $T(\phi) = 0$. For each n and by definition of $\hat{\phi}_n^\alpha$ we have :

$$\begin{aligned} ||\hat{T}_n(\hat{\phi}_n^\alpha)||^2 + \alpha_n ||\hat{\phi}_n^{\alpha} - \phi^*||_s^2 &\leq ||\hat{T}_n(\phi_0)||^2 + \alpha_n ||\phi_0 - \phi^*||_s^2 \\ &\leq \underbrace{\delta_n^2}_{\to 0} + \underbrace{\alpha_n}_{\to 0} \underbrace{||\phi_0 - \phi^*||}_{\text{fixed}} \to 0. \end{aligned}$$

It's easy to see that:

$$||\hat{\phi}_n^\alpha - \phi^*||_s^2 \leq \underbrace{\frac{\delta_n^2}{\alpha_n}}_{\to 0} + ||\phi_0 - \phi^*||_s^2.$$

Consequently, the sequence $(\hat{\phi}_{\alpha_n})$ is bounded. We can therefore extract a convergent subsequence, that we will note $\hat{\phi}_{\alpha_n,2}$. We note ϕ_l the limit of this subsequence. We have to show that ϕ_l is itself a solution of $T(\phi) = 0$. First let's remark that $\hat{T}_n(\hat{\phi}_{\alpha_n,2})$ tends to 0 since:

$$||\hat{T}_n(\hat{\phi}_{\alpha_n,2})||^2 \leq \underbrace{||\hat{T}_n(\phi_0)||^2}_{\to 0} + \underbrace{\alpha_n ||\phi_0 - \phi^*||_s^2}_{\to 0}.$$

Then we decompose:

$$T(\phi_l) = T(\phi_l) + \hat{T}(\phi_l) - \hat{T}(\phi_l) + \hat{T}(\hat{\phi}_{\alpha_n,2}) - \hat{T}(\hat{\phi}_{\alpha_n,2})$$

and consequently:

$$||T(\phi_l)|| \leq \underbrace{||\hat{T}(\hat{\phi}_{\alpha_n,2})||}_{\to 0} + \underbrace{\sqrt{||\hat{T}(\phi_l) - \hat{T}(\hat{\phi}_{\alpha_n,2})||}}_{\leq ||\hat{T}||\ ||\phi_l - \hat{\phi}_{\alpha_n,2}||_s} + \underbrace{\leq ||(\hat{T} - T)(\phi_l)||}_{\sqrt{||\hat{T}-T||\ ||\phi_l||_s}}$$

with the right hand side tending to 0. Indeed, $||\phi_l||_s$ is fixed, $||\phi_l - \hat{\phi}_{\alpha_n,2}||_s \to 0$, and with additional assumption that $||\hat{T} - T|| \to 0$, $||\hat{T}|| = O(1)$ will also follow. We conclude from this that $(\hat{\phi}_{\alpha_n,2})$ (which is a subsequence of $(\hat{\phi}_{\alpha_n})$) tends to ϕ_l, which is solution of the initial problem.

□

2.7.5 Proof of Lemma 2.5.2

In the following, for sake of simplicity, we will skip indexation by n, using $\hat{\phi}$, α, $\hat{T}$, and δ instead of $\hat{\phi}_n^\alpha$, α_n, $\hat{T}_n$, and δ_n. Those objects however still depend on the size of the sample. The problem is to minimize :

$$||\hat{T}(\phi)||^2 + \alpha||\phi - \phi^*||_s^2.$$

By definition of the solution $\hat{\phi}$, we have in particular:

$$||\hat{T}(\hat{\phi})||^2 + \alpha||\hat{\phi} - \phi^*||_s^2 \leq ||\hat{T}(\phi_0)||^2 + \alpha||\phi_0 - \phi^*||_s^2.$$

Having:

$$||\hat{\phi} - \phi^*||_s^2 = ||\hat{\phi} - \phi_0||_s^2 + ||\phi_0 - \phi^*||_s^2 + 2 < \hat{\phi} - \phi_0, \phi_0 - \phi^* >_s$$

we get:

$$||\hat{T}(\hat{\phi})||^2 + \alpha||\hat{\phi} - \phi_0||_s^2 \leq ||\hat{T}(\phi_0)||^2 + 2\alpha < \hat{\phi} - \phi_0, \phi_0 - \phi^* >_s .$$

Equivalently, we assume without loss of generality that $\phi^* = 0$. We write the decomposition of $\hat{T}$ in ϕ_0 with the Fréchet-derivative:

$$\hat{T}(\hat{\phi}) = \hat{T}(\phi_0) + \hat{T}'_{\phi_0}(\hat{\phi} - \phi_0) + \hat{r}.$$

Then:

$$||\hat{T}(\hat{\phi})||^2 = ||\hat{T}(\phi_0) + \hat{r}||^2 + ||\hat{T}'_{\phi_0}(\hat{\phi} - \phi_0)||^2 + 2 < \hat{T}(\phi_0) + \hat{r}, \hat{T}'_{\phi_0}(\hat{\phi} - \phi_0) >$$

and as $||\hat{T}(\phi_0)||^2 \leq \delta^2$ we finally obtain:

$$\begin{aligned} ||\hat{T}'_{\phi_0}(\hat{\phi}-\phi_0)||^2 + \alpha||\hat{\phi}-\phi_0||_s^2 &\leq \delta^2 + \alpha < \hat{\phi}-\phi_0, \phi_0 >_s -2 < \hat{T}(\phi_0) + \hat{r}, \hat{T}'_{\phi_0}(\hat{\phi}-\phi_0) > \\ & -2 < (\hat{T}'_{\phi_0} - T'_{\phi_0})(\hat{\phi}-\phi_0), T'_{\phi_0}(\hat{\phi}-\phi_0) > \\ & -||\hat{T}(\phi_0) + \hat{r}||^2 - ||(\hat{T}'_{\phi_0} - T'_{\phi_0})(\hat{\phi}-\phi_0)||^2. \end{aligned}$$

The two last terms of the right hand side equation are negative, we also have:

$$\begin{aligned} ||\hat{T}'_{\phi_0}(\hat{\phi}-\phi_0)||^2 + \alpha||\hat{\phi}-\phi_0||_s^2 &\leq \delta^2 + \alpha < \hat{\phi}-\phi_0, \phi_0 >_s -2 < \hat{T}(\phi_0) + \hat{r}, \hat{T}'_{\phi_0}(\hat{\phi}-\phi_0) > \\ & -2 < (\hat{T}'_{\phi_0} - T'_{\phi_0})(\hat{\phi}-\phi_0), T'_{\phi_0}(\hat{\phi}-\phi_0) > . \end{aligned} \tag{2.28}$$

Using Assumption 2.5.2-(vi), we have that $||T'_{\phi_0}(\hat{\phi}-\phi_0)||^2 \sim ||\hat{\phi}-\phi_0||^2_{-a}$. Moreover the source condition (Assumption 2.5.2-(v)) implies that there exists w such as $\phi_0 = L^{-\beta}w$. We can also transform $< \hat{\phi}-\phi_0, \phi_0 >_s$ as $<,>_s$ is a scalar product in a Hilbert scale: there exists an operator L (self-adjoint, unbounded, strictly positive, densely defined) such that:

$$\begin{aligned} < \hat{\phi}-\phi_0, \phi_0 >_s &= < L^s(\hat{\phi}-\phi_0), L^s\phi_0 > \\ &= < L^2 s(\hat{\phi}-\phi_0), \phi_0 > \\ &= < L^2 s(\hat{\phi}-\phi_0), L^{-\beta}w > \\ &= < L^{2s-\beta}(\hat{\phi}-\phi_0), w > \\ &= \mathcal{O}(||\hat{\phi}-\phi_0||_{2s-\beta}) \end{aligned}$$

as L^s and $L^{-\beta}$ are self-adjoint, the last equality coming from the Cauchy-Schwartz inequality. We still have to transform the two scalar products in the right-hand side equation. First if we write:

$$\hat{T}'_{\phi_0} = T'_{\phi_0} + (\hat{T}'_{\phi_0} - T'_{\phi_0})$$

we simply get by Cauchy-Schwartz inequality:

$$< \hat{T}(\phi_0) + \hat{r}, T'_{\phi_0}(\hat{\phi}-\phi_0) >= \mathcal{O}(\underbrace{||\hat{T}(\phi_0) + \hat{r}||}_{\mathcal{O}(\delta+||\hat{r}||)} \underbrace{||T'_{\phi_0}(\hat{\phi}-\phi_0)||}_{=\mathcal{O}(||(\hat{\phi}-\phi_0)||_{-a})}).$$

Using Assumption 2.5.2-(iii) we have consequently:

$$< \hat{T}(\phi_0) + \hat{r}, T'_{\phi_0}(\hat{\phi}-\phi_0) >= \mathcal{O}\Big((\delta + ||\hat{\phi}-\phi_0||^2)||\hat{\phi}-\phi_0||_{-a}\Big)$$

which implies:

$$< \hat{T}(\phi_0) + \hat{r}, (T'_{\phi_0} - T'_{\phi_0})(\hat{\phi}-\phi_0) >= \mathcal{O}\Big((\delta + ||\hat{\phi}-\phi_0||^2)\underbrace{(||T'_{\phi_0} - T'_{\phi_0}||\ ||\hat{\phi}-\phi_0||)}_{=\mathcal{O}(\gamma||\hat{\phi}-\phi_0||)}\Big).$$

Then:

$$< \hat{T}(\phi_0) + \hat{r}, (T'_{\phi_0} - T'_{\phi_0})(\hat{\phi}-\phi_0) >= \mathcal{O}\Big((\delta + ||\hat{\phi}-\phi_0||^2)(||\hat{\phi}-\phi_0)||_{-a} + \gamma||\hat{\phi}-\phi_0||)\Big).$$

Moreover it's easy to obtain:

$$< (\hat{T}'_{\phi_0} - T'_{\phi_0})(\hat{\phi}-\phi_0), T'_{\phi_0}(\hat{\phi}-\phi_0) >= \mathcal{O}(\gamma||\hat{\phi}-\phi_0||\ ||\hat{\phi}-\phi||_{-a})$$

by applying the Cauchy-Schwartz inequality and Assumptions 2.5.2-(iv) and (vi). Equation 2.28 becomes finally:

$$\begin{aligned}||\hat{\phi}_n^\alpha - \phi_0||_{-a}^2 + \alpha_n||\hat{\phi}_n^\alpha - \phi_0||_s^2 \quad &\leq \quad \delta_n^2 + \alpha_n||\hat{\phi}_n^\alpha - \phi_0||_{2s-\beta}^2 + ||\hat{\phi}_n^\alpha - \phi_0||_{-a}(\delta_n + ||\hat{\phi}_n^\alpha - \phi_0||^2) \\ &\quad +\gamma_n||\hat{\phi}_n^\alpha - \phi_0||(\delta_n + ||\hat{\phi}_n^\alpha - \phi_0||_{-a} + ||\hat{\phi}_n^\alpha - \phi_0||^2)\end{aligned}$$

□

2.7.6 Proof of Lemma 2.5.3

We will see two useful results. First, the interpolation inequality in Hilbert scales which expresses that for x in H_s and $-\infty < q < r < s < \infty$ we have:

$$||x||_r \leq ||x||_q^{\frac{s-r}{s-q}} ||x||_s^{\frac{r-q}{s-q}}. \tag{2.29}$$

Second, Engl et al. (1996) recalls that for any $c, d, e \geq 0$ and $p, q \geq 0$ we have that:

$$c^p \leq e + dc^q \Rightarrow c^p = \mathcal{O}(e + d^{\frac{p}{p-q}}). \tag{2.30}$$

We now want to express the inequality in Lemma 2.5.2 only with norms of Hilbert scales $||.||_{-a}$ and $||.||_s$. Consequently, applying inequality 2.29 with $r = 0$, $q = -a$ and $s = s$, we have:

$$||\hat{\phi} - \phi_0|| \leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{a}{s+a}}.$$

Then:

$$\begin{aligned}||\hat{\phi} - \phi_0||_{-a}||\hat{\phi} - \phi_0||^2 &\leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{a+3s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{2a}{s+a}} \\ ||\hat{\phi} - \phi_0||_{-a}||\hat{\phi} - \phi_0|| &\leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{a+2s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{a}{s+a}} \\ |||\hat{\phi} - \phi_0||^3 &\leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{3s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{3a}{s+a}}.\end{aligned}$$

If we suppose that $a \leq s$ then it means that $||\hat{\phi} - \phi_0||_{-a}||\hat{\phi} - \phi_0||^2 = o(||\hat{\phi} - \phi_0||_{-a})||^2$ and that $||\hat{\phi} - \phi_0||^3 = o(||\hat{\phi} - \phi_0||_{-a}||\hat{\phi} - \phi_0||)$. Additionally if we apply 2.29 with $q = -a$ and $r = 2s - \beta$ we get:

$$||\hat{\phi} - \phi_0||_{2s-\beta} \leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{\beta-s}{a+s}} ||\hat{\phi} - \phi_0||_s^{\frac{2s-\beta+a}{a+s}}.$$

Finally:

$$\begin{aligned}||\hat{\phi}_n^\alpha - \phi_0||_{-a}^2 + \alpha_n||\hat{\phi}_n^\alpha - \phi_0||_s^2 \quad &\leq \quad \delta^2 + \delta||\hat{\phi} - \phi_0||_{-a} + \alpha(||\hat{\phi} - \phi_0||_{-a}^{\frac{\beta-s}{a+s}} ||\hat{\phi} - \phi_0||_s^{\frac{a+2s-\beta}{a+s}}) \\ &\quad +\gamma(\delta||\hat{\phi} - \phi_0||_{-a}^{\frac{s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{a}{s+a}} + ||\hat{\phi} - \phi_0||_{-a}^{\frac{a+2s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{a}{s+a}})\end{aligned} \tag{2.31}$$

Applying Property 2.30 four times to the terms of the right hand side of equation 2.31 we get:

$$||\hat{\phi}_n^\alpha - \phi_0||_{-a}^2 = \mathcal{O}\Big(\delta^2 + \alpha^{\frac{2(a+s)}{2a+3s-\beta}} ||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2(a+2s-\beta)}{2a+3s-\beta}} + (\gamma\delta)^{\frac{2(a+s)}{2a+s}} ||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a}{2a+s}} + \gamma^{\frac{2(a+s)}{a}} ||\hat{\phi}_n^\alpha - \phi_0||_s^2)\Big).$$

Using that for any $u, v \geq 0$: $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$ this can be rewritten:

$$||\hat{\phi}_n^\alpha - \phi_0||_{-a} = \mathcal{O}\Big(\delta + \alpha^{\frac{a+s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{2a+3s-\beta}} + (\gamma\delta)^{\frac{a+s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{2a+s}} + \gamma^{\frac{a+s}{a}}||\hat{\phi}_n^\alpha - \phi_0||_s\Big). \quad (2.32)$$

Then we wish to replace the order obtained in 2.32 for $||\hat{\phi}_n^\alpha - \phi_0||_{-a}$ in the right hand side of Equation 2.31. We use that for any ξ, $x = O(u+v)$ implies that $x^\xi = \mathcal{O}(u^\xi + v^\xi)$. If we use Assumption 2.5.2-(viii) we get that:

$$\begin{aligned}
||\hat{\phi}_n^\alpha - \phi_0||_{-a}^2 + \alpha||\hat{\phi}_n^\alpha - \phi_0||_s^2 \quad = \quad & \mathcal{O}\Big(\delta^2 + \delta\alpha^{\frac{a+s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{2a+3s-\beta}} \\
& +\alpha[\delta^{\frac{\beta-s}{a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{a+s}} + \alpha^{\frac{\beta-s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2(\beta-2s-a)}{\beta-2a-3s}}] \\
& +\Big\{\gamma^{\frac{a+s}{2a+s}}\delta^{\frac{3a+2s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{2a+s}} + \alpha(\gamma\delta)^{\frac{\beta-s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2s+2a-\beta}{2a+s}} \\
& +(\gamma^{\frac{a+s}{a}}\delta + \alpha\gamma^{\frac{\beta-s}{a}})||\hat{\phi}_n^\alpha - \phi_0||_s + \gamma\delta^{\frac{a+2s}{a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{a+s}} \\
& \gamma\delta\alpha^{\frac{s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a+2s-\beta}{2a+3s-\beta}} + (\gamma\delta)^{\frac{2a+2s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a}{2a+s}} \\
& \gamma\alpha^{\frac{a+2s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{4s+3a-2\beta}{2a+3s-\beta}} + \gamma^{\frac{3a+3s}{2a+s}}\alpha^{\frac{a+2s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{3a}{2a+s}}\Big\}\Big)
\end{aligned}$$

and finally that:

$$\begin{aligned}
||\hat{\phi}_n^\alpha - \phi_0||_s^2 \quad = \quad & \mathcal{O}\Big(\delta^2\alpha^{-1} + \delta\alpha^{\frac{\beta-a-2s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{2a+3s-\beta}} \\
& +[\delta^{\frac{\beta-s}{a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{a+s}} + \alpha^{\frac{\beta-s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2(\beta-2s-a)}{\beta-2a-3s}}] \\
& +\Big\{\gamma^{\frac{a+s}{2a+s}}\delta^{\frac{3a+2s}{2a+s}}\alpha^{-1}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{2a+s}} + (\gamma\delta)^{\frac{\beta-s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2s+2a-\beta}{2a+s}} \\
& +(\gamma^{\frac{a+s}{a}}\delta\alpha^{-1} + \gamma^{\frac{\beta-s}{a}})||\hat{\phi}_n^\alpha - \phi_0||_s + \alpha^{-1}\gamma\delta^{\frac{a+2s}{a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{a+s}} \\
& \gamma\delta\alpha^{\frac{\beta-2a-2s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a+2s-\beta}{2a+3s-\beta}} + \alpha^{-1}(\gamma\delta)^{\frac{2a+2s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a}{2a+s}} \\
& \gamma\alpha^{\frac{\beta-a-s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{4s+3a-2\beta}{2a+3s-\beta}} + \gamma^{\frac{3a+3s}{2a+s}}\alpha^{\frac{s-a}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{3a}{2a+s}}\Big\}\Big).
\end{aligned} \quad (2.33)$$

The first part of the right hand side of Equation 2.33 is identical to the case of standard nonlinear inverse problem, however a rather complicated term depending on γ appears after that. When we don't know anything about $\gamma = \gamma_n$ it's difficult to simplify the former expression. However, we can apply the property 2.30 to each term in this expression. The first part will be left unchanged and is similar to the case of standard nonlinear inverse problems. In the second part, delimited by $\{\dots\}$, when applying this property, the powers $\frac{a}{a+s}$, $\frac{a}{2a+s}$, $\frac{2a}{2a+s}$ of $||\hat{\phi}_n^\alpha - \phi_0||_s$ leads to terms that are $o(\delta^2/\alpha)$ using Assumption 2.5.2-(viii). The term in $||\hat{\phi}_n^\alpha - \phi_0||_s$ leads to a term that is only a $\mathcal{O}(\gamma^{\frac{2(\beta-s)}{a}})$. Finally, we obtain (after having taken the root of every expression):

$$\begin{aligned}||\hat{\phi}_n^\alpha - \phi_0||_s &= \mathcal{O}(\delta\alpha^{-\frac{1}{2}} + \delta^{\frac{2a+3s-u}{3a+4s-u}}\alpha^{\frac{u-a-2s}{3a+4s-u}} + \delta^{\frac{u-s}{a+u}} + \alpha^{\frac{u-s}{2a+2s}} \\ &+(\gamma\delta)^{\frac{(\beta-s)}{2a+\beta}} + \gamma^{\frac{(\beta-s)}{a}} \\ &+(\gamma\delta)^{\frac{(2a+3s-\beta)}{2a+4s-\beta}}\alpha^{\frac{(\beta-2a-2s)}{2a+4s-\beta}} \\ &+\gamma^{\frac{(2a+3s-\beta)}{a+2s}}\alpha^{\frac{(\beta-a-s)}{a+2s}} + \gamma^{\frac{3(a+s)}{a+2s}}\alpha^{\frac{(s-a)}{a+2s}}).\end{aligned} \tag{2.34}$$

The choice of α must be made by examining the first part of the expression. Then, the good choice for α is:

$$\alpha \sim \delta^{\frac{2a+2s}{a+\beta}}.$$

If we replace this quantity in the first line of Equation 2.34, this gives a speed of convergence of order:

$$O(\delta^{\frac{\beta-s}{a+\beta}}).$$

We now want to derive conditions under which the remaining terms in γ, with α equal to this value, are negligible compared to this speed of convergence. Concerning the second term we see that for this, $\gamma^{\frac{\beta-s}{a}}$ must be a $o(\delta^{\frac{\beta-s}{a+\beta}})$. As $\beta \geq s$, this is ensured by the condition:

$$\gamma^{\frac{a+\beta}{a}}\delta^{-1} \to 0. \tag{2.35}$$

The first term implies that we must have: $(\gamma\delta)^{\frac{\beta-s}{2a+\beta}} = o(\delta^{\frac{\beta-s}{a+\beta}})$. As we have:

$$\frac{\gamma\delta^{\frac{\beta-s}{2a+\beta}}}{\delta^{\frac{\beta-s}{a+\beta}}} = (\gamma\delta^{\frac{-a}{a+\beta}})^{\frac{\beta-s}{2a+\beta}}$$

condition 2.35 is still sufficient as $\beta - s$ and $2a + \beta$ are positive.

The expression of the third term is the following:

$$\gamma^{\frac{2a+3s-\beta}{2a+4s-\beta}}\delta^{\frac{5as+4s^2-5\beta s+2a^2-3q\beta+\beta^2}{(a+\beta)(\beta-4s-2a)}}.$$

To assess that it's a $o(\delta^{\frac{\beta-s}{a+\beta}})$, we estimate:

$$\frac{\gamma^{\frac{2a+3s-\beta}{2a+4s-\beta}}\delta^{\frac{5as+4s^2-5\beta s+2a^2-3q\beta+\beta^2}{(a+\beta)(\beta-4s-2a)}}}{\delta^{\frac{\beta-s}{a+\beta}}} = (\gamma\delta^{-\frac{a}{a+\beta}})^{\frac{2a+3s-\beta}{2a+4s-\beta}}$$

that ensures the desired convergence since $2a + 3s - \beta > 0, s > 0$.

Studying the fourth term, we must have $\gamma^{\frac{3s-\beta+2a}{a+2s}}\delta^{\frac{-2(s+a)(-\beta+a+s)}{(a+\beta)(a+2s)}} = o(\delta^{\frac{\beta-s}{a+\beta}})$. That is to say:

$$\frac{\gamma^{\frac{3s-\beta+2a}{a+2s}}\delta^{\frac{-2(s+a)(-\beta+a+s)}{(a+\beta)(a+2s)}}}{\delta^{\frac{\beta-s}{a+\beta}}} = (\gamma\delta^{\frac{-a}{a+\beta}})^{\frac{2a+3s-\beta}{a+2s}}$$

which converges to 0 with the same condition and the fact that again $2a + 3s - \beta$ and $a + 2s$ are assumed to be positive.

For the last term, we seek again if $\gamma^{\frac{3(a+s)}{a+2s}}\delta^{\frac{2(a+s)(s-a)}{(a+\beta)(a+2s)}} = o(\delta^{\frac{\beta-s}{a+\beta}})$. Consequently we remark that:

$$\frac{\gamma^{\frac{3(a+s)}{a+2s}}\delta^{\frac{2(a+s)(s-a)}{(a+\beta)(a+2s)}}}{\delta^{\frac{\beta-s}{a+\beta}}} = (\gamma^{\frac{a+\beta}{a}}\delta^{\frac{2(a+s)(s-a)-(\beta-s)(a+2s)}{3a(a+s)}})^{\frac{a(a+2s)}{3(a+\beta)(a+s)}}$$

It appears that $\frac{[2(a+s)(s-a)-(\beta-s)(a+2s)]}{3a(a+s)} \geq -1$ since $2s+a$ is supposed greater than β. Then the condition $\gamma^{\frac{a+\beta}{a}}\delta^{-1} \to 0$ is stronger and this is verified. Finally, we get the desired speed of convergence and the result of the lemma is verified.

□

2.7.7 Proof of Lemma 2.5.4

We study the minimization problem stated by Equation 2.20. All the norms and the scalar products will be relative to L^2-spaces. We will note $\hat{\phi}$, α and δ instead of $\hat{\phi}_n^\alpha$, α_n and δ_n and take $\phi^* = 0$ for sake of simplicity. As in the demonstration of Lemma 2.5.3 it is possible to show that:

$$||\hat{T}(\hat{\phi})||^2 + \alpha_n||\hat{\phi}-\phi_0||^2 \leq \delta^2 + 2\alpha < \hat{\phi}-\phi_0, \phi_0 > .$$

We try to re-express the scalar product in the former expression using the fact that $\phi_0 = {T'_{\phi_0}}^*.w$ and that $\hat{T}(\hat{\phi}) = \hat{T}(\phi_0) + \hat{T}'_{\phi_0}(\hat{\phi}-\phi_0) + \hat{r}$, we have clearly that:

$$\begin{aligned} < \hat{\phi}-\phi_0, \phi_0 > &= < w, T'_{\phi_0}(\hat{\phi}-\phi_0) > \\ &= < w, \hat{T}'_{\phi_0}(\hat{\phi}-\phi_0) > + < w, (T'_{\phi_0} - \hat{T}'_{\phi_0})(\hat{\phi}-\phi_0) > \\ &= < w, (T'_{\phi_0} - \hat{T}'_{\phi_0})(\hat{\phi}-\phi_0) > + < w, \hat{T}(\hat{\phi}) > - < w, \hat{T}(\phi_0) + \hat{r} > . \end{aligned}$$

We have using the Cauchy-Schwartz inequality the three following inequalities:

- $< w, (T'_{\phi_0} - \hat{T}'_{\phi_0})(\hat{\phi}-\phi_0) > \quad \leq \quad ||w||\ ||(T'_{\phi_0} - \hat{T}'_{\phi_0})(\hat{\phi}-\phi_0)|| \quad \leq \quad \gamma_n\ ||w||\ ||\hat{\phi}-\phi_0||$
- $< w, \hat{T}(\hat{\phi}) > \quad \leq \quad ||w||\ ||\hat{T}(\hat{\phi})||$
- $- < w, \hat{T}(\phi_0) + \hat{r} > \quad \leq \quad ||w||\ ||\hat{T}(\phi_0) + \hat{r}|| \quad \leq \quad ||w||(\delta + ||\hat{r}||) \quad \leq \quad ||w||(\delta + C||\hat{\phi}-\phi_0||^2)$.

Consequently, we have that:

$$||\hat{T}(\hat{\phi})||^2 + \alpha_n||\hat{\phi}-\phi_0||^2 \quad \leq \quad \delta^2 + 2\alpha||w||(\delta + C||\hat{\phi}-\phi_0||^2 + ||\hat{T}(\hat{\phi})|| + \gamma_n||\hat{\phi}-\phi_0||).$$

This rewrites easily:

$$(||\hat{T}(\hat{\phi})|| - \alpha||w||)^2 + \alpha(1-2||w||C)\ ||\hat{\phi}-\phi_0||^2 - 2\alpha\ ||w||\ \gamma_n||\hat{\phi}-\phi_0|| \leq (\delta + \alpha||w||)^2.$$

We can moreover express :

$$\alpha(1-2||w||C)||\hat{\phi}-\phi_0||^2 - 2\alpha||w||\gamma_n||\hat{\phi}-\phi_0|| = \Big(\sqrt{\alpha(1-2||w||C)}||\hat{\phi}-\phi_0|| - \frac{\sqrt{\alpha}||w||\gamma_n}{\sqrt{1-2||w||C}}\Big)^2 - \frac{\alpha||w||^2\gamma_n^2}{1-2||w||C}.$$

This gives :

$$(||\hat{T}(\hat{\phi})|| - \alpha||w||)^2 \leq (\delta + \alpha||w||)^2 + \frac{\alpha||w||^2\gamma_n^2}{1-2||w||C}.$$

As $a^2 \leq b^2 + c^2$ implies that $a \leq b + c$ when a, b, c are positive, we conclude that:

$$||\hat{T}(\hat{\phi})|| \leq \delta + 2\alpha||w|| + ||w||\gamma_n\sqrt{\frac{\alpha}{1-2||w||C}}.$$

Replacing in the former expression it's easy to obtain that:

$$||\hat{\phi} - \phi_0|| \leq \frac{\delta + \alpha||w||}{\sqrt{\alpha(1-2||w||C)}} + 2\frac{\gamma||w||}{1-2||w||C}.$$

As in Engl et al. (1996) the first term drives the choice of the α parameter: $\alpha_n \sim \delta_n$ implying that $||\hat{\phi}_n^\alpha - \phi_0|| = O(\sqrt{\delta_n})$. The second term remains of the same order as soon as $\gamma_n = o(\sqrt{\delta_n})$ or equivalently that $\gamma_n^2\delta^{-1}$ tends to 0. Finally, we can conclude that $\hat{T}(\hat{\phi}) = O(\delta_n)$.

□

2.7.8 Proof of Theorem 2.5.1

First, we can apply theorem 2.4 of Kaltenbacher et al. (2008) with $y = 0$ and we have that the non-linear Landweber iteration applied with the true operator converges to a solution of $T\phi = 0$. We could hope to use theorem 2.6 of Kaltenbacher et al. (2008) but our framework is different. For each estimated operator $\hat{T}_n$ the iteration is modified.

In the following, ϕ_0 will be the initial guess. For sake of simplicity we note $\hat{T}_n = \hat{T}$, $\hat{T}'_n = \hat{T}'$, $\hat{\phi}_k^\delta = \hat{\phi}_k$. $\hat{\phi}_k$ and ϕ_k are respectively the k-th iterate of the Landweber iteration for the problem with unknown and exact operator. ϕ^* will be the limit of the Landweber iteration with exact operator (i.e. exact solution).

$\triangleright$ **Step A** - Let $k_n = k_*(\delta_n, \hat{T}_n)$ the stopping index determined with the discrepancy principle applied to the nonlinear Landweber iteration with unknown operator. We assume that there exists k, an accumulation point of $\{k_n\}$. As $\{k_n\} \in \mathbb{N}^{\mathbb{N}}$, k may be assumed to be fixed, $k = k_n, \forall n \in \mathbb{N}$ (if not, we can always consider a subsequence of $\{k_n\}$). By definition we have:

$$|| - \hat{T}\hat{\phi}_k|| \leq \tau\delta_n$$

as k is the accumulation point of the k_n. k is not compulsory an accumulation point for the similar sequence for the exact operator. However we will show that ϕ_k is also a solution of the problem with an exact operator.

$\triangleright$ **A1** - We want to prove that $\hat{\phi}_k = \hat{\phi}_k^{\delta_n} \rightarrow \phi_k$. We remark in particular that:

- (2.5.8)-3 ensures that $\hat{\phi}_j$ stays bounded and that $||\hat{\phi}_j|| < 2\rho$ for $j \in [1; k]$.

- As $||T'_\phi|| \leq 1$ for each $\phi \in B(0; 2\rho)$ and $||T'_\phi - \hat{T}'_\phi|| \leq \gamma_n \rightarrow 0$ for $\phi \in B(0; 2\rho)$ then there exists M' such that for each $n \in \mathbb{N}$:

$$||\hat{T}'_\phi|| \leq M' \quad \forall\phi \in B(0; 2\rho).$$

Let's recall that for $j \in [1; k-1]$:

$$\phi_{j+1} = \phi_j + T'^*_{\phi_j}(-T\phi_j)$$
$$\hat{\phi}_{j+1} = \hat{\phi}_j + \hat{T}'^*_{\hat{\phi}_j}(-\hat{T}\hat{\phi}_j)$$

with $\phi_0 = \hat{\phi}_0$ an initial guess, $\phi_0 \in B(0; 2\rho)$. We show the following lemma:

Lemma 2.7.1. $\forall i \in [1; k], \exists A_i, B_i \in \mathbb{R}^+$ *such that:*

- A_i *and* B_i *do not depend on* n*;*
- $||\phi_i - \hat{\phi}_i|| \leq A_i \gamma_n + B_i \delta_n$.

Proof 1. We show this result recursively:

<u>**i=1 :**</u>

$$\begin{aligned} \phi_1 - \hat{\phi}_1 &= \phi_0 - \phi_0 + T'^*_{\phi_0}(-T\phi_0) - \hat{T}'^*_{\phi_0}(-\hat{T}\phi_0) \\ &= T'^*_{\phi_0}(-T\phi_0) - \hat{T}'^*_{\phi_0}(-\hat{T}\phi_0). \end{aligned}$$

Thus:

$$\begin{aligned} ||\phi_1 - \hat{\phi}_1|| &\leq ||T'^*_{\phi_0}(-T\phi_0) - \hat{T}'^*_{\phi_0}(-T\phi_0)|| + ||\hat{T}'^*_{\phi_0}(T\phi_0 - \hat{T}\phi_0)|| \\ &\leq \gamma_n ||T\phi_0|| + \delta_n ||\phi_0|| M' \end{aligned}$$

thus the property holds with $A_1 = ||T\phi_0||$ and $B_1 = ||\phi_0||M'$.

Let's suppose that **the property holds for a given** $j \in [1; k-1]$.

At <u>**step** $j+1$</u> we have:

$$\phi_{j+1} - \hat{\phi}_{j+1} = \phi_j - \hat{\phi}_j + [T'^*_{\phi_j}(-T\phi_j) - \hat{T}'^*_{\hat{\phi}_j}(-\hat{T}\hat{\phi}_j)]$$

Thus:

$$||\phi_{j+1} - \hat{\phi}_{j+1}|| \leq \underbrace{||\phi_j - \hat{\phi}_j||}_{I'} + \underbrace{||T'^*_{\phi_j}(-T\phi_j) - \hat{T}'^*_{\hat{\phi}_j}(-\hat{T}\hat{\phi}_j)||}_{I''}.$$

We have by assumption: $I' \leq A_j \gamma_n + B_j \delta_n$. Moreover:

$$\begin{aligned} I'' \leq\ & \underbrace{||T'^*_{\phi_j}(-T\phi_j) - \hat{T}'^*_{\phi_j}(-T\phi_j)||}_{I''_1} + \underbrace{||\hat{T}'^*_{\phi_j}(-T\phi_j) - \hat{T}'^*_{\hat{\phi}_j}(-T\phi_j)||}_{I''_2} \\ & \underbrace{||\hat{T}'^*_{\hat{\phi}_j}(-T\phi_j) - \hat{T}'^*_{\hat{\phi}_j}(-T\hat{\phi}_j)||}_{I''_3} + \underbrace{||\hat{T}'^*_{\hat{\phi}_j}(-T\hat{\phi}_j) - \hat{T}'^*_{\hat{\phi}_j}(-\hat{T}\hat{\phi}_j)||}_{I''_4} \end{aligned}$$

and it is easy to see that:

- $I''_1 \leq ||T\phi_j||\ ||T'^*_{\phi_j} - \hat{T}'^*_{\phi_j}|| \leq 2\rho||T||\gamma_n$;

- $I''_3 \leq ||T||\ ||\phi_j - \hat{\phi}_j||\ ||\hat{T}'^*_{\hat{\phi}_j}|| \leq ||T||M'(A_j\gamma_n + B_j\delta_n)$;

- $I''_4 \leq ||T - \hat{T}||\ ||\hat{\phi}_j||\ ||\hat{T}'^*_{\hat{\phi}_j}|| \leq 2M'\rho\delta_n$.

To obtain a bound on I''_2 we will have to observe that for $\phi, \tilde{\phi} \in B(0; 2\rho)$:

$$\begin{aligned}
||\hat{T}'^*_{\phi} - \hat{T}'^*_{\tilde{\phi}}|| &\leq \underbrace{||\hat{T}'^*_{\phi} - T'^*_{\phi}||}_{\leq||\phi||\gamma_n} + \underbrace{||T'^*_{\phi} - T'^*_{\tilde{\phi}}||}_{\leq||\phi-\tilde{\phi}||} + \underbrace{||T'^*_{\tilde{\phi}} - \hat{T}'^*_{\tilde{\phi}}||}_{\leq\gamma_n||\tilde{\phi}||} \\
&\leq 2\rho\gamma_n + ||\phi - \tilde{\phi}|| + 2\rho\gamma_n \\
&= 4\rho\gamma_n + ||\phi - \tilde{\phi}||.
\end{aligned}$$

Consequently:

$$\begin{aligned}
I''_2 &\leq ||T\hat{\phi}_j||\ ||\hat{T}'^*_{\phi_j} - \hat{T}'^*_{\hat{\phi}_j}|| \\
&\leq ||T||\ ||\hat{\phi}_j||(4\rho\gamma_n + ||\phi_j - \hat{\phi}_j||) \\
&\leq 2||T||\rho(4\rho\gamma_n + A_j\gamma_n + B_j\delta_n).
\end{aligned}$$

Finally:

$$\begin{aligned}
I'' &\leq 2\rho||T||\gamma_n + 2||T||\rho(4\rho\gamma_n + A_j\gamma_n + B_j\delta_n) + ||T||M'(A_j\gamma_n + B_j\delta_n) + 2M'\rho\delta_n \\
&= \delta_n(2||T||\rho B_j + 2M'\rho + ||T||M'B_j) + \gamma_n(2\rho||T|| + 8||T||\rho^2 + 2||T||A_j + A_j||T||M')
\end{aligned}$$

and the property is verified at step $j+1$ with :

- $A_{j+1} = A_j(1 + 2\rho||T|| + ||T||M') + 2\rho||T|| + 8\rho^2||T||$
- $B_{j+1} = B_j(1 + 2\rho||T|| + ||T||M') + 2\rho M'$

that do not depend on n. Then we obtain the proof of the lemma.

□

Consequently:

$$||\phi_k - \hat{\phi}_k|| \leq A_k\gamma_n + B_k\delta_n$$

and $\hat{\phi}_k = \hat{\phi}_k^{\delta_n} \to \phi_k$ when $n \to +\infty$.

▷ **A2 -** At this point, we have that for an accumulation point k of $\{k_n\}$, $\hat{\phi}_k^{\delta_n}$, the $k-th$ iterate of the Landweber iteration (with estimated operator) converges with $n \to \infty$ towards the $k-th$ iterate of the Landweber iteration with exact operator. We still have to show that this function is also solution of the initial problem. First we can notice that:

$$\begin{aligned}
||T\phi_k - \hat{T}\hat{\phi}_k|| &\leq ||T\phi_k - T\hat{\phi}_k|| + ||T\hat{\phi}_k + \hat{T}\hat{\phi}_k|| \\
&\leq ||T||\ ||\phi_k - \hat{\phi}_k|| + ||T - \hat{T}||\ ||\hat{\phi}_k|| \to 0
\end{aligned}$$

as $||T|| < +\infty$ (using Assumption 2.5.7), $\hat{\phi}_k \to \phi_k$, $||T-\hat{T}|| \leq \delta_n \to 0$ and $||\hat{\phi}_k|| < +\infty$. Consequently:

$$||T\phi_k|| \leq ||T\phi_k - \hat{T}\hat{\phi}_k|| + ||\hat{T}\hat{\phi}_k|| \to 0$$

as

$$||T\phi_k - \hat{T}\hat{\phi}_k|| \to 0$$

and

$$||T\phi_k - \hat{T}|| \leq \tau\delta_n \to 0.$$

The main point is that as k is fixed. ϕ_k depends continuously on the (exact) data, and and as $n \to \infty$, we get that $||T\phi_k|| = 0$. So ϕ_k is also a solution for the exact problem and $\hat{\phi}_k \to \phi_k$.

$\triangleright$ **Step B** - We still have to consider the case where $\{k_n\}$ has no accumulation point: as $\{k_n\} \in \mathbb{N}^{\mathbb{N}}$, $k_n \to \infty$ with $n \to +\infty$.

$\triangleright$ **B1 -** Suppose for the moment that we have the following lemma that will be shown later.

Lemma 2.7.2. *If $\hat{\phi}_k \in B(\phi_*;\rho)$ and $||\hat{T}\hat{\phi}_k|| > \tau\delta_n$ with $\tau > 2\frac{1+\eta}{1-2\eta} > 2$ then $\hat{\phi}_{k+1}$ is a better approximation of ϕ_* than $\hat{\phi}_k$.*

We admit it for the moment. Let $\epsilon > 0$, we wish that it exists $\bar{k}$ such as $\forall k_n \geq \bar{k}$, $||\hat{\phi}_{k_n} - \phi_*|| \leq \epsilon$. Let $k \in \mathbb{N}$. For $k_n \geq k$, we have by definition of k_n:

$$||\hat{\phi}_{k_n} - \phi_*|| \leq ||\hat{\phi}_k - \phi_*|| \leq ||\hat{\phi}_k - \phi_k|| + ||\phi_k - \phi_*||.$$

As ϕ_* is the limit of $(\phi_j)_{j\in\mathbb{N}}$, there exists $\bar{k} = k(\epsilon)$ such that:

$$||\phi_k - \phi^*|| \leq \frac{\epsilon}{2}.$$

Let $\bar{n} = \bar{n}(\epsilon, \bar{k})$ such that $A_k\gamma_n + B_k\delta_n \leq \frac{\epsilon}{2}$, $\forall n \geq \bar{n}$. Then for n sufficiently large, we have both $k_n > \bar{k}$ (with $\bar{k} \neq k(n)$) and $n \geq \bar{n}(\epsilon, \bar{k})$ implying $||\hat{\phi}_{k_n} - \phi^*|| \leq \epsilon$.

What's important is that at one moment, with ϵ fixed, we have fixed a $\bar{k}$ that do not depend on n such that A_k and B_k are fixed.

$\triangleright$ **B2 -** Finally we have to show Lemma 2.7.2. Let's assume that $\hat{\phi}_k \in B(\phi_*, \rho)$ then clearly ϕ_* and $\hat{\phi}_k \in B(0; 2\rho)$. We have moreover $(\hat{T}'_\phi)^* = (\hat{T'^*_\phi})$ as we work under suitable probability densities π and τ (see Florens et al. (2009))

$$
\begin{aligned}
||\hat\phi_{k+1}-\phi_*||^2-||\hat\phi_k-\phi_*||^2 &= 2<\hat\phi_{k+1}-\hat\phi_k;\hat\phi_k-\phi_*>+||\hat\phi_{k+1}-\hat\phi_k||^2\\
&= 2<\hat T'^*_{\hat\phi_k}(-\hat T\hat\phi_k);\hat\phi_k-\phi_*>+||\hat T'^*(-\hat T\hat\phi_k)||^2\\
&= 2<(-\hat T\hat\phi_k);\hat T'_{\hat\phi_k}(\hat\phi_k-\phi_*)>+\underbrace{||\hat T'^*(-\hat T\hat\phi_k)||^2}_{\le||\hat T\hat\phi_k||^2}\\
&\le 2<-\hat T\hat\phi_k;-\hat T\hat\phi_k+\hat T'_{\hat\phi_k}(\phi_*-\hat\phi_k)>-2||\hat T\hat\phi_k||^2+||\hat T\hat\phi_k||^2\\
&= 2<-\hat T\hat\phi_k;-\hat T\hat\phi_k-\hat T'_{\hat\phi_k}(\phi_*-\hat\phi_k)>-||\hat T\hat\phi_k||^2\\
&\le 2||\hat T\hat\phi_k||||-\hat T\hat\phi_k-\hat T'_{\hat\phi_k}(\phi^*-\hat\phi_k)||-||\hat T\hat\phi_k||^2\\
&= ||\hat T\hat\phi_k||(2||\hat T\hat\phi_k-\hat T'_{\hat\phi_k}(\phi_*-\hat\phi_k)||-||\hat T\hat\phi_k||).
\end{aligned}
$$

We have that $T\phi^*=0$, so:

$$
\begin{aligned}
||\hat T\hat\phi_k-\hat T'_{\hat\phi_k}(\phi_*-\hat\phi_k)|| &= ||\hat T\phi_*-\hat T\hat\phi_k-\hat T'_{\hat\phi_k}(\phi_*-\hat\phi_k)+\underbrace{T\phi^*}_{=0}-\hat T\phi^*||\\
&\le \underbrace{||T\phi^*-\hat T\phi^*||}_{\le\delta_n}+\underbrace{||\hat T\phi_*-\hat T\hat\phi_k-\hat T'_{\hat\phi_k}(\phi_*-\hat\phi_k)||}_{\le\eta||\hat T\phi_*-\hat T\hat\phi_k||}\\
&\le ||\hat T\hat\phi_k||(2\delta_n-2\eta\underbrace{||\hat T\phi_*-\hat T\hat\phi_k||}_{\le||\hat T\hat\phi_*||+||\hat T\hat\phi_k||}||\hat T\hat\phi_k||)\\
&\le ||\hat T\hat\phi_k||(2\delta_n+2\eta\delta_n-||\hat T\hat\phi_k||)+||\hat T\hat\phi_k||\\
&\quad = ||\hat T\hat\phi_k||(2\delta_n(1+2\eta)-||\hat T\hat\phi_k||(1-2\eta))<0 \quad \text{by assumption}
\end{aligned}
$$

and then $\hat\phi_{k+1}$ is a better approximation of ϕ_* than $\hat\phi_k$.

This achieves to prove the convergence of the method.

$\square$

Chapter 3

Nonparametric Analysis of Hedge Funds Lifetimes

joint work with Jean-Pierre Florens and Serge Darolles[1], Working Paper. To be submitted.

Abstract

Most of Hedge Funds databases are now keeping history of dead funds in order to control for biases in empirical analysis. It is then possible to use these data for the analysis of Hedge Funds lifetimes and survivorship. This chapter proposes two nonparametric specifications of duration models. First, the single risk model is an alternative to parametric duration models used in the literature. Second, the competing risks model consider the two reasons explaining why Hedge Funds stop reporting. We apply the two models to Hedge Funds data and compare our results to the literature. In particular, we show that a cohort effect must be considered. Moreover, the reason of the exit is a crucial information for the analysis of funds' survival as for a large part of disappearing funds, their exit cannot be explained by a low performance or a low level of assets.

[1]This chapter is adapted from Darolles, S., Florens, J.P., Simon, G. (2010), *Nonparametric Analysis of Hedge Funds Lifetimes*

3.1 Introduction

Hedge Funds' reporting is related to the objectives of their managers. All the information was formerly provided on alive funds only, but the industry of Hedge Funds, contrary to mutual funds, exhibits a high level of attrition. The appearance of those dead funds bases follows a stream of academic papers underlining that unavailable data on dead funds led to numerous statistical biases. With those data, it is thus possible to analyze in a static fashion, the differences between living and dead funds by conditioning on the status (alive or dead) of the fund.

However, a dynamic approach is a more ambitious and flexible framework to model the probability to die (across time) conditionally on some (potentially dynamic) fund characteristics. And in this framework, censorship is not always well taken into account in empirical studies. For example, Pojarliev and Levich (2010) presents a statistical study on funds categorized along a posterior, observable status at the date of study, depending on the fact that the funds are alive or dead. One cannot draw two distinct studies on funds depending on their current status since information is carried by both alive and dead funds. Similarly, conditioning by the age of the fund may be not sufficient, as the whole set of variables (returns, assets, and age of the fund) are random processes that are mutually linked. From a mathematical perspective, alive funds are individuals for which the death has not yet been observed (censored funds) but that still depend on the same framework of analysis. The dynamic approach is a growing field of research, but related papers mainly often use parametric specification and focus on a single risk framework. However, failure is not the only reason why Hedge Funds stop reporting to databases. Starting and stopping to report depend on proper and historic characteristics of the fund. If the fund performs well, the amount of assets under management (henceforth AUM) may reach a critical size, above which arbitrages may not be profitable. The manager can decide that the fund does not need new clients, and stops the publication of its performance. This may explain that in practice some funds exhibit just before their exit, a good performance and a high level of AUM. However, the exit from the database may also simply mean the end of the life of the fund (liquidation, default, etc.), or the willing of its manager to hide bad performances before liquidation.

The aim of this chapter is to discuss the specification choices made in this literature on Hedge Funds lifetimes. First, we propose to adopt a nonparametric framework, rather than a parametric one, to avoid mis-specification biases and get robust estimation of hazard intensities of default. Second, we take into account the reasons for the exit. In this context, we take as a standard the use of covariates, including dynamic ones, as performance and AUM for instance. It is straightforward that the lifetime duration have to be cautiously defined. We must admit that we cannot take into account a self-selection bias since our data are only available thanks to voluntary publication. We do not observe funds that have never decided to appear in a database. Those funds may have been liquidated just after their creation, or may be for instance family funds that do not need to publish their performances. However, it is coherent to think that any fund appearing in a database has needed once, to collect new investors. In the opposite case, database publication would be useless. We are consequently focusing on funds needing to report performance and to collect investors.

In Section 3.2, we present features of Hedge Funds lifetimes, including advantages and drawbacks of using Hedge Funds databases. We precisely define how lifetimes are calculated and present common biases studied by the literature in Section 3.3. Then, we present the estimation procedure, particularly focusing on the nonparametric framework and the use of covariates, including competing risks model. Section 3.4 presents our results, where we underline that nonparametric estimation avoids the problems of mis-specification when studying hazard intensities. We present also a strong influence of the inception date of the fund as a covariate and accounts for the importance of identifying the cause of the exit of the fund.

3.2 Data and definition of lifetimes

Each Hedge Fund defines its own rule of reporting, and potentially publishes its performances in a database. This induces several classical biases: and Hedge Funds lifetimes and survivorship are difficult to study precisely[2]. This section reviews the previous contributions made on the subject.

3.2.1 In the literature

All academic studies agree on some fundamental points. Hedge Funds show a high degree of attrition within each year[3]. Fung and Hsieh (1997b) study the evolution of this rate between 1990 and 1996 and find values around 19%, much higher than for mutual funds. Brown et al. (1999) find a value around 14% each year during the period 1987-1996, and Amin and Kat (2002) find values of the same order for period 1994-2001. However, these results depend on the database and on the period: crises impact deeply death and birth processes of Hedge Funds. Moreover, attrition and instantaneous probability of failure are two different concepts.

General considerations on the analysis of Hedge Funds survival and its link with style and characteristics may be found in Liang (2000) and Barès et al. (2007). Other contributions estimate survival probabilities as a function of characteristics, past performance and risk statistics. Hendricks et al. (1997) assume that performance could improve the survival probability, and also propose the idea that survivorship bias may induce some patterns in Hedge Funds survival analysis. Brown et al. (2004) identifies a strong relation between bad performance and consecutive disappearance from the database. Gregoriou (2002) provides a technical study of Hedge Funds survival based on covariates taken among lagged performance, fund strategy, leverage, or characteristics (fees, redemption, etc.). This work has been extended by models using default intensity like Grecu et al. (2007), pointing out that for dying funds, performance is generally poor at the end of the fund's life. However, when controlling for covariates, this has to be done on the whole life of the fund as in Ang and Bollen (2010) who use a dynamic Cox model with dynamic covariates. All these works use a single risk approach. They focus on the exit of the fund from the database, not on the cause of the exit.

In fact, there are two potential reasons to explain the exit from a database. But studying the exit from a database is useful only if we can separate between those reasons. This is a matter of identification, since if the reasons for the exit are fundamentally different, the blind mixing of the two risks leads to false interpretation, implying an over-estimation of the rate of attrition of the industry. In Rouah (2005) Hedge Funds lifetimes are studied with time-dependent covariates, along with the cause of exit under the assumption of independent risks. The paper underlines that avoiding to separate liquidation from other kind of withdrawal leads to severe biases (especially over-estimation of the survivorship bias). Amin and Kat (2002) explains that it may be due, for instance, to a lack of size

[2]We give in Section 3.A.4 the overview of the academic contributions on Hedge Funds durations.

[3]This degree of attrition can be easily estimated by counting the proportion of funds that are alive or dead at the end of each year.

or performance. In some databases, the main reasons explaining the death of a fund are sometimes detailed. It may have been liquidated, merged, its assets transferred in an other fund; the minimum capacity of the fund may have been reached after several redemptions during a market turmoil, or the main investor may have left the fund; the company may have been closed, the manager may have left, or the team-management has decided to concentrate on other strategies.

Appropriate covariates have to be found in order to analyze Hedge Fund lifetimes. For this, AUM is generally assumed to be a good indicator of funds' status. Assets and performance are generally supposed to decrease the funds' default intensity. But the other potential reason to explain the withdrawal of a fund from a database is that the fund has reached a maximum capacity and then that it is closed to investment. In some successful Hedge Funds, the manager may think that the fund has reached a sufficient size in order to keep its ability to generate performance. This is linked with characteristics that are relative to the specificity of each fund and the decision to stop to collect money is related to the fund's objectives and possibilities. Amin and Kat (2002) explicitly don't afford much attention on this point, as Grecu et al. (2007) who explain that the first reason (dying funds) is the most frequent and that the lack of data concerning successful funds stopping to collect is not a problem in practice. Conversely, Ackermann et al. (1999), Gregoriou and Rouah (2002) or Rouah (2005) stress that this kind of exit must be considered: if not the attrition rates are too high when this separation is not made. The survivorship bias is also affected.

3.2.2 TASS database

A large number of Hedge Funds databases are available, with their own specificities and characteristics. We use here the TASS database[4] for at least two reasons. Information concerns both alive and dead funds. A large number of fields is available, which allows to control for biases and to improve the study. We first describe the TASS database of 2009, June the 25th. The database is made of two group of funds, depending whether they are Managed Futures/CTA funds, or not. For each category, there is a file concerning alive funds, and a "graveyard" containing dead funds. For a given file, it consists in a list of fund shares, each share being identified by an unique number. Each share corresponds to a given currency. In addition to CTA and Managed Futures, the represented strategies are among the main Hedge Fund strategies such as single strategies (Long/Short Equity, Event Driven, Global Macro, Fixed Income Arbitrage, Multi-Strategy, Convertible Arbitrage, Dedicated Short Bias, Emerging Markets, Equity Market Neutral, Emerging Markets, Options Strategy) and fund of funds. Information on the fund status is published concerning status, dates, liquidity and performance. For example, the domicile country or state, public opening, leverage, management and incentive fees, presence of a highwater mark, and dead reason (when the fund is dead) are displayed. Moreover, dates when the fund has been added and/or removed from the database; dates of inception, start and end of performance are indicated. Concerning liquidity and investment constraints, subscription/redemption frequency, redemption notice period, lock-up and lock-up period may also be indicated. Finally, turning to historical performances, monthly returns (published or estimated by TASS), NAV (see p.170), and AUM are sometimes given (in domicile currency).

An open discussion concerns the field called *drop reason* about which academic contributors are still puzzled. As Baquero et al. (2005) or Rouah (2005), Boyson (2002) focuses first on explicitly liquidated

[4]As Boyson (2002) (who use a database from 2002), Amin and Kat (2002), Getmansky et al. (2004b), Grecu et al. (2007) (database of 2004) or Liang and Park (2008) (database of 2004) for instance.

funds only. Getmansky et al. (2004b) also insist on the importance of this field. Kundro and Feffer (2003) go further as they distinguish fund's failure (external reasons force the fund manager to stop) from fund liquidation (the collect of fees is not sufficient, the manager ceases activity before re-launching a new fund to reset the highwater mark to a new level). This suggests to use this drop reason with caution and with additional filters on the data.

3.2.3 Hedge Funds lifetimes

3.2.3.1 Lifetimes definition

In summary, the life of a fund in a database may be decomposed in the following way. First the fund is created (inception date). After that, the fund may possibly enter the database (date added) and reports performance. Consequently, the first date of report may be anterior to the date of entry in the database, but not to the inception date. When the fund disappears from the alive database, it enters the graveyard with a dead reason, and the last performance date is equivalent to the death date of the fund. The death of the fund may be explained by the fact that the fund is closed to new investment, dormant, liquidated, no longer reporting, liquidated, merged into an other fund, that the program is closed or that it is not possible to contact the fund. If the fund stops reporting, two potential causes may explain this event. First, the fund is poorly performing and is liquidated, merged, or closed down, and ceases activity. But if the fund performs well, the manager may decide to stop from collecting new investors, and does not report performance any more. If the fund seeks new investors, it continues to report and is neither closed or dead: its performance is still observable in the database. This implies to define in a clear fashion the entry and the exit from the database.

If each information in a database was perfect, any study on Hedge Funds lifetimes would be easy. The first reporting date would be equal both to the entry date in the database, and to the inception date. For funds still alive, report would be updated until the current date. Each field would be cautiously reported. Between entry and exit of the database, both performance and AUM would be given at each date. Consequently, lifetime durations of Hedge Funds would be easy to analyze with all covariates available in the database (fund characteristics, performance, AUM). However, databases show some practical limits that constrain our study.

3.2.3.2 Initial date

It is well known that Hedge Funds databases suffer from numerous drawbacks, the first one being the backfilling bias. A fund can be added to a database a long time after its inception and backfills its history with past performance[5]. When a fund decides to backfill its track return, the manager has the choice of the first date, with the possibility to skip first returns if this performance is bad. In some cases, funds report performances anterior to their inception date: these are synthetic (not real but *a posteriori* re-estimated) performances. Funds with first report before inception are clearly backfillers and must be dropped from the sample. This assesses the choice of the inception as the initial date, as it is commonly made in the literature.

Assumption A.1 The initial date corresponds to the inception date, and only funds that display explicitly an inception date are considered.

[5]This occurs frequently since funds can be older than the database: young and bad-performing funds that have not been able to survive, do not appear in the database.

3.2.3.3 Date of exit

A second drawback, named *reporting bias*, is in the potential delay in the report of performances. For instance, funds classified as alive in June 2009 (and thus with data at the end of May 2009) may not have published their performances at this date. The fund may still be in activity with figures that are not yet available (because of liquidity, mark-to-market reasons, or other causes). We then consider data at a former date than the last date available of our database and use this to potentially upgrade the graveyard. The date of study remains May 2009 (last data). But funds in the alive database are dropped to the graveyard if their status explicitly indicates it[6] (liquidated, no longer reporting, or unable to contact it) or if it's not the case, that the last report is anterior to four months (no data at the end of January).

Assumption A.2 A fund is considered as dead if it is in the graveyard, including funds with too much delay or explicit dead reason. Final date for dead funds is the date of last performance report, and the fund is uncensored. Final date for alive funds is the date of study, and the fund is considered as censored.

As a consequence to assumptions A.1 and A.2, we calculate the lifetime of the fund as the time between the initial and the final date. We must add that we have to assume throughout this chapter that censorship and durations are independent. If it was not the case we could not identify the law of the observed durations. This seems to be a plausible hypothesis in our framework since censor is made by the current observation date. Note that taking censorship into account is crucial in survival analysis and must always be done (contrarily to the approach of Grecu et al. (2007) that only make inference on uncensored funds).

3.3 Model specification and estimation

This section discusses the specification and the estimation of the single and the competing risks models.

3.3.1 Model specification

Single risk models are first considered. The competing risks approach can then be seen as a multivariate extension of this framework.

3.3.1.1 Single risk model

Our aim is therefore to estimate the hazard function $\lambda(t)$ for the survival of Hedge Funds, as its interpretation is intuitive and its shape carries valuable information. Mainly, since hazard rate at t is homogenous to the intensity for each fund to exit the database at date t, the higher the hazard rate, the more likely the fund is to stop reporting performance. In general, this rate is not constant and young funds have a greater instantaneous probability to exit the database (see e.g. Brown et al.

[6] As the database is updated by TASS at given dates, it may appear that between two of them, some dead funds have not joined yet the graveyard.

(2001), Amin and Kat (2002)). But the precise form of this shape is still subject to questions. The parametric framework constrains the analysis by imposing a predetermined form, and misses some information at specific moments of the life of the fund. Monotonic decreasing forms assume that fund managers take benefit from their experience and that aged funds are less likely to die. Inverted U-shape assumes that funds without experience take more risk and die; old funds cease to collect assets, explaining this U-shape. U-shape patterns is the most common choice (Gregoriou (2002), Grecu et al. (2007)). The intensity increases, reaches a maximum value and then decreases for aged funds.

An improvement of the previous approach is to specify this intensity as a (deterministic) function of time and also of (potentially dynamic) covariates. In Cox proportional hazard models, dependence towards covariates x is introduced via a function ψ such as the density of the duration is supposed to be $f(t\psi(x))\psi(x)$. In this framework, $\lambda(t)$ for each individual is assumed to be proportional to the product of a *baseline* hazard function $\lambda_0(t)$ and a function $\psi(x,\beta)$ of covariates $x: \lambda(t) = \lambda_0(t)\psi(x,\beta)$. The model remains semi-parametric as soon as no assumption is made on $\lambda_0(t)$, and β is a parameter to be estimated. The ψ function is usually $\psi(x,\beta) = exp(\beta' x)$, which allows to estimate separately the effect of the covariate, and the baseline intensity. Other forms are possible such as $\psi(x,\beta) = 1 + \beta' x$ or $\psi(x,\beta) = log(1 + e^{(\beta' x)})$.

3.3.1.2 Competing risks model

When several failure types exist the related framework is referred to as *competing risks* models. For each individual, the risks are of different natures, and several mechanisms are in *competition* to explain the causes of failures. Suppose that failure may be caused by m distinct types of risks: for each individual this implies m failure times, which are called *latent* or *potential* times. For each risk i, the potential time of failure will be T_i, and the observed time of failure is $T = min(T_1, ..., T_m)$. But the specification in those models have to be cautiously discussed. Three different questions arise (see Kalbfleisch and Prentice (2002)): first, one may study the specific behavior and the probability of occurrence of each failure type; second, the objective may be to examine, under a predefined set of hypotheses, the interdependence between those failure types; finally, an interesting point may be the effect of the removal of one risk on the occurrence of the others.

Turning to Hedge Funds lifetimes, the motivation for such models is straightforward. If the fund defaults because of bad performances, this failure type will be labelled T_-. On the contrary, if the fund is a good performer, the manager decides to stop reporting because he does not need any new investor, this will be labelled $T+$ since it implies a withdrawal from the database. If those two risks respectively lead to latent times of exit from the database T_- and T_+, the observed time of failure is $T = min(T_-, T_+)$.

Basically, identification is only possible if in addition of the time of failure, the *cause* of the failure (namely the nature of the risk) is observed. Then, the set of observations has to be of the form (T, j) where T is the failure time and $j \in [1; m]$ is the cause of the failure. However, even corresponding time-independent covariates are insufficient to fully explain the interrelation between risks and to ensure identification (see Tsiatis (1975)). The problem is that given a competing-risks model with dependent durations, it is possible to find a different model, with independent durations, that is however observationally equivalent.

Given time-dependent covariates $x(t)$, an identifiable quantity is however the *type-specific hazard* which is defined as :

$$\lambda_j(t, x(t)) = lim_{\Delta t \to 0} \frac{\mathbb{P}[t \leq T \leq t + \Delta t, J = j | T \geq t]}{\Delta t}$$

where T is the observed failure time and j the cause of the failure. In our study $m = 2$, and two causes cannot occur simultaneously. Some other functions may be introduced (*overall failure rate*, *overall survivor*, *cumulative incidence function*, etc.) but even a function of the form : $S_j(t, x) = exp(-\int_0^t \lambda_j(u, x(u))du))$ cannot be interpreted as a survivor function, when no additional assumption on the interdependence of the competing risks is introduced.

We assume that the two risks of type 1 and 2 are independent and specify a semi-parametric form for the j-th intensity:

$$\lambda_j(t, x(t)) = \lambda_0^{(j)}(t) exp(\beta_j' x(t))$$

where $x(t)$ is a set of covariates whose underlying left-continuous, with right limits, and $j \in [1; m]$. $\lambda_0^{(j)}(t)$ is left unspecified and will be nonparametrically estimated.

3.3.1.3 Estimation

We provide in Appendix A.3 parametric and nonparametric estimation procedures in the single risk framework without covariates. We first focus here on the estimation of a dynamic Cox model using covariates, with a nonparametric baseline intensity. Then, we discuss the estimation of competing-risks models.

3.4 Empirical study

In addition to the results on the whole database, we also present results by separating funds along with their declared strategies. All the results are compiled in Appendix 3.A.

3.4.1 Descriptive statistics

First we merge the CTA file in an overall database. Then we obtain a database made of 6121 funds and a graveyard of 8102 funds. The description of the database by strategy is available in Table 3.1. The histogram of raw durations is given in Figure 3.1. We also present in Table 3.2, separating along main strategies, some descriptive statistics: mean duration, empirical mode and empirical quantiles. The category Long-Short Equity Hedge represents the main part of single funds (1521 funds out of 3963). We aggregate CTA and Managed Futures categories, which form one of the main categories after Multi-Strategy, Event Driven, Global Macro, Equity Market Neutral, Fixed Income Arbitrage, and Emerging Markets.

The mean durations for single funds and fund of funds are quite close, around 60 months (5 years). It's straightforward that single funds durations behavior is mainly driven by Long-Short Equity Hedge. This quantity is however coherent as for each strategy, the mean duration is comprised between 50 and 70 months. The strategies with the shortest mean durations are Global Macro, Equity Market Neutral, and Multi-Strategy, this being also verified for the upper 95% empirical quantile. Conversely, Event Driven and CTA-Managed Futures are strategies with the longest mean duration. It may be observed that in the case of CTA, there are much more dead funds, then less censored individuals, which may explain this fact. The more recent is a strategy, the younger the funds and the shortest the mean duration.

3.4.2 Single risk analysis

3.4.2.1 Nonparametric intensities

In this section, we compare parametric and nonparametric estimations of the hazard intensity for Hedge Funds lifetimes. First, our aim is to check whether the specifications commonly used in the literature are justified. Second, we try to evaluate whether the nonparametric estimation can improve the understanding of a hedge fund lifetime. Third, we want to adapt our conclusions depending on the several strategies. Concerning the parametric fitting, we use a log-logistic law. This distribution is commonly chosen (see e.g. Gregoriou (2002) or Grecu et al. (2007)). Results[7] are given in Table 3.3. We recall that the density of a log-logistic distribution of parameters (μ, σ) is given by:

$$f_{\mu,\sigma}(x) = \frac{e^{\frac{ln(x)-\mu}{\sigma}}}{\sigma x(1 + e^{\frac{ln(x)-\mu}{\sigma}})^2}$$

where μ is the location and σ the scale parameter. A convenient expression is obtained when $\rho = e^{-\mu}$, $\kappa = 1/\sigma$:

$$f(t) = \frac{\kappa \rho^\kappa t^{\kappa-1}}{(1 + (\rho t)^\kappa)^2} \qquad S(t) = \frac{1}{1 + (\rho t)^\kappa} \qquad \lambda(t) = \frac{\kappa \rho^\kappa t^{\kappa-1}}{(1 + (\rho t)^\kappa)}$$

.

The obtained modes for the hazard rates (that is the time at which the highest intensity of default is reached), are around 40 months, which is coherent with the values obtained in the literature (see Grecu et al. (2007)), yet a bit inferior. They are also coherent with the order of magnitude of the empirical modes[8].

The true interest of such a model is to compare these results with nonparametric estimation of the hazard intensity. For this, we must first build the confidence bounds of the parametric curve. We note $\theta = (\mu, \sigma)$ the parameters of the log-logistic distribution and $\hat{\theta}$ their maximum likelihood estimates. If the hazard function is given by $t \mapsto \lambda(t, \theta)$ the confidence interval for $\lambda(t)$ at level α is given by:

$$\lambda(t, \hat{\theta}) \pm \frac{q_{1-\alpha/2}}{\sqrt{N}} \sqrt{\frac{\partial \lambda}{\partial \theta} \times V \times \frac{\partial \lambda}{\partial \theta'}}$$

where N is the number of observations, $q_{1-\alpha/2}$ is the quantile of a standard gaussian law, V is the estimated variance of the parameters. This expression is obtained with the application of a usual δ-method. The results are presented from Figures 3.2 to 3.12. For each strategy, it appears that the choice of the log-logistic function is coherent with the general shape of the nonparametric intensity obtained most smoothed Kaplan-Meier estimator. These estimators evokes an inverted U-shape pattern. The intensity is increasing in the first years, reach a peak and then decreases. This is at least valid for the earlier years of existence of the fund. It is not coherent to draw this intensity for very high

[7]The same study has been done for discrete durations. The results are not presented here, but they are very similar and do not modify deeply the conclusions.

[8]However, as the histogram of durations is often sparse or not well designed, the empirical mode is not very informative.

durations (over 10 or 12 years) as there are then not enough funds in the sample for the estimation to be reliable. For each strategy, we add the evolution of the proportion of funds still in the sample across time, along with a threshold (set to 10%) to ensure that the intensities are still representative.

At a finer scale, several differences appear. First, the log-logistic specification is more appropriate for fund of funds than for single funds. For single strategies, the log-logistic specification is particularly adapted to Event Driven and Fixed Income Arbitrage funds. For Event-Driven funds, the nonparametric law has nearly exactly the same shape and the same mode than the parametric one. Long-Short Equity Hedge, Equity Market Neutral, and CTA intensities have roughly a nonparametric intensity that is of the same shape than in the parametric case, even if the intensity is not in the confidence interval during some months, before becoming close to the parametric specification after same time. For Multi-Strategy funds, the parametric specification is plausible but the nonparametric intensity decrease is stronger after 60 months than suggested by the parametric specification. The situation is more critic for Emerging Markets and Global Macro funds where the intensity is quite monotonous, increases, crosses the parametric intensity and exhibits an irregular pattern. Consequently for those two categories, the mode of the default intensity suggested by the nonparametric estimation is higher than the one obtained with the parametric estimation. Excepted Multi-Strategy and Event Driven funds, this conclusion is also valid for most strategies. Generally speaking, we cannot confirm the finding of Gregoriou (2002) that default intensity decreases after having reached a peak : except for Multi-Strategy, the behavior of the intensities after having reached a local mode is quite steady or non-monotonic.

In conclusion, the parametric specification of the log-logistic distribution is coherent with the general shape suggested by the nonparametric estimation. If the choice may be justified for several strategies, the parametric distribution systematically under-estimates the modes (up to two years) and the long-term hazard intensities. On the contrary, the real initial hazard intensity is strictly different from zero, a fact which is missed by the parametric analysis. Then, the nonparametric analysis capture specific features of the intensities and avoid specification problems (short-term under-estimation, and long-term over-estimation).

3.4.2.2 Including covariates

The former analysis is descriptive and useful, yet limited. We then introduce covariates. The intensity writes : $\lambda(t) = \lambda_0(t)\phi(\underline{x}_t)$, $\underline{x}_t$ being the past and current values of appropriately chosen covariates x. When the remaining $\lambda_0(t)$ is constant equal to λ, all information is captured by the model. A "noisy" structure remains and $\lambda_0(t) = \lambda$ is the intensity of a homogenous Poisson process.

Performance and AUM appear as two natural covariates for the study, but other variables have been considered. Amin and Kat (2002) examine the fund leverage, even if Barry (2002) minimizes the impact of this variable; Grecu et al. (2007) proposes to incorporate the Sharpe Ratio and the volatility of the fund return. Baquero et al. (2005) is concerned with the investment style and Gregoriou (2002) tests the influence of management fees, performance fees, minimum purchase, mean returns, or redemption periods. This is also in line with Avellaneda and Besson (2007) that develop the concept of skill-capacity i.e. the maximum amount of money a manager can expect without decreasing its arbitrages and consequently its performance. As in Couderc et al. (2008), one may also think to include business or economic variables or indexes. In our study, we mainly consider performance and AUM.

But an additional variable is considered: the distance between the date of inception and a static date[9] to deal with potential cohort effects. A mistake may be found in parametric studies related to parametric or semi-parametric estimation of default intensity with covariates. When assuming an intensity $\lambda(t) = \lambda_0(t)\psi(Z)$ where Z are (potentially time-dependent) covariates, it is pointless to include time as an explanatory variable in Z, as all the time dependence will be captured by $\lambda_0(t)$.

As a fund cannot enter our analysis as soon as at a given date, some data remain undisclosed, we recall that we use the method described in Appendix 3.A.5 to avoid to drop too much funds in our study. Moreover, we exclude backfillers, i.e. funds that present data before their inception date. The chosen form of the hazard intensity is:

$$\lambda(t) = \lambda_0(t)exp(\beta' x(t))$$

where $\lambda_0(t)$ is left unspecified. As the missing data problem shrinks the number of available funds for the study, we will mainly focus on Single and Fund of Funds. Moreover, only funds in the same currencies (returns, performance, assets) can be part of the study. We decide here to consider all the funds in US Dollars. This is coherent as the funds in the database are shares, which are often shares from the same fund in different currencies. Selecting USD funds is a way of selecting some representative shares of funds: this is also done by Liang and Park (2008). The number of funds available for the study is given in Table 3.4. The coefficients of several "nested" models are given in Table 3.5 for single funds, and Table 3.6 for funds of funds.

Among the chosen variables we consider the monthly returns (*Returns*), the level of equity[10] (*Equity*), the logarithm of the level of AUM[11] ($ln(AUM_t)$), the monthly log-return of AUM ($\Delta\, ln(AUM_t/AUM_{t-1})$) and the difference in days between the inception date of the fund and[12] a static date, taken here as the 1st of January 1990 (*Date Ref*). Results are provided in Table 3.5 for single funds, and in Table 3.6 for funds of funds.

For single funds and fund of funds, we see that for models with one or two variables, and without *Date Ref*, all the coefficients are significant. All these coefficients are negative, suggesting that a higher level of the covariate decreases the intensity of default. This is intuitive as high levels of performance, return and AUM, may be indicators of good health of the fund. However, when we include the *Date Ref* variable, this variable captures most of the dynamic of the intensity. Both for single and fund of funds, return and log-return of AUM are no more significant. For single funds, Equity and AUM are plausible explicative variables, whereas for fund of funds, only the AUM variable appears to affect the intensity. In conclusion, the only variables of concern are AUM, Equity and the "reference date". This last variable accounts for a cohort effect, suggesting that all other parameters being equal, younger funds are more at risk than old ones.

A parametric specification in this context of the baseline intensities (non-monotonic and not always with an inverted U-shape) would be difficult since it would be too constraining and would lead us

[9]We take here arbitrarily this date as the 1st January 1990 but any static reference date could be used. However, a reference date too far from the sample of the inception dates could lead to potential scale effects that shrink the mean level of the baseline intensity.

[10]If R_t denotes the monthly return of a fund, we define the level E_T of equity at time T as $E_T = \Pi_{t=1}^{T}(1 + R_t)$.

[11]In millions of USD.

[12]We specifically here focus on a limited set of variables. Nonlinear functions of the return or financial indicators as in Couderc et al. (2008) have also been included but appeared to be not really informative.

to miss the specific features of the baseline (especially its flatness). The obtained nonparametric intensities are given in Figure 3.16 and 3.16.

3.4.3 Competing risks analysis

In the single risk framework, no distinction is made on the nature of the exit. When a fund is still reporting, one may be interested in the potential issues he is going to face in a near future. The probability of exit from the database is interesting in itself, but what is more pertinent, is to forecast or to assess the reason for it. Bad performers are liquidated, whereas good performers may be still in activity after their exit from the database.

3.4.3.1 Competing risks model

A simple simulation exercise can be made to illustrate the importance of using competing risks model. Draw a sequence of N couples of parameters $(\mu_i, \sigma_i)_{i \in [0;N]} \in [-r;+r] \times [s_m; s_M]$ with $r, s_m, s_M > 0$. Then, make for each i a random simulation path of independent monthly gaussian returns R_t with mean μ_i and variance σ_i. For each i, simulate two independent durations T_i^+ and T_i^-, with the respective intensity processes: $\lambda_+(t) = \lambda_0(t)exp(\beta \times R_t)$ and $\lambda_-(t) = \lambda_0(t)exp(-\beta \times R_t)$. The easiest way to run the simulation is to draw for each individual i two independent realization E_i^+ and E_i^- of $Exp(1)$ variables. Then T^+ and T^- are obtained as :

$$\int_0^{T_i^+} \lambda_0(t)exp(\beta R_t^i)dt = E_i^+ \quad \text{and} \quad \int_0^{T_i^-} \lambda_0(t)exp(-\beta R_t^i)dt = E_i^-.$$

The simulations are made with the same positive value for β but with a positive sign for exit T^+ and a negative sign for T^-. λ_0 is chosen identical for both risk (log-logistic for instance).
If we leave out in our analysis the true generating process of the durations, and that we only consider the times $T_i = min(T_i^+, T_i^-)$ (regardless of the cause, T_+ or T_-), with the return as a dynamic covariate and the exact model $\lambda(t) = \lambda_0(t)exp(\gamma R_t)$, we find a value of γ which is close to zero, significatively different from β. As two populations with opposite risks are aggregated in similar proportions, both effects compensate the other in the estimation. Then, leaving aside the cause of the failure leads to estimated coefficients that are severely biased.

3.4.3.2 Estimation

To estimate a competing risks model, we have to identify the cause of the failure of the fund. This cause is not always observed and we have to set a procedure to decide whether the fund exits the database because of bad or good performance. The decision is made along two criterions: the death reason of the fund in the database and/or its level of AUM at the time of exit.

First, all the funds that classified as "closed to new investment" in May 2009 are considered as exiting the database because of good performance (exit of type T_+). For the other uncensored funds, we consider that the funds have an exit of type T_-, when their AUM at the time of exit is below 200 millions of US dollars or that this value is less than 75% of the maximum value reached by the total AUM. In the opposite case (high AUM, above 200 millions, and more than 75% of the maximum value) the exit is considered as of type T_+ and that the fund is exiting from the database because it has reached a sufficient capacity[13]. Those two values are chosen to represent a selective, upper quantile of the distribution of AUM. It may be observed empirically that the distribution of the variable made by the ratio of the last AUM on the maximum AUM of the fund is quite uniform except in one, where it reaches a peak. This suggests that a huge proportion of funds exit the database without experiencing a large decollect of assets. The number of resulting funds are given in Table 3.7. The number of funds for the T_+ label are not very sensitive at this level to changes in the quantitative criterions on the AUM selected here.

The result of the estimation are given in Table 3.8. The effect of the reference date is quite homogenous for single and fund of funds. However its effect is modulated for single funds as the effect is more important for the T_+ exit. For single funds and T_+ exit, we find that both Equity and AUM have a positive influence in the increase of the positive default intensity. This is coherent since better performance and more AUM increases the possibility of the fund to be closed to new investment, and then to stop reporting. Concerning the T_- exit, we recover the former interpretation concerning AUM, but equity seems to be not significant in this case. For fund of funds, we recover for the negative exit the former interpretation for AUM, but which is not significant for the positive exit. The resulting baseline intensities are given in Figure 3.16 and 3.17.

3.5 Conclusion

We define in this chapter a precise framework to study Hedge Funds lifetimes. The more convenient way to study durations is to define them as the difference between the last date of performance report (after consolidation of the database) and the inception date (when available) of each fund. Inclusion of censorship is crucial. Results in the single risk framework are greatly improved by nonparametric estimation since even if the choice of the log-logistic function is justified for some Hedge Funds strategies, the parametric hazard intensity are systematically under-estimating the probability of exit in the early times, and behavior on the long-term depends on the strategy, with modes of distributions that are too low, up to two years. The use of covariates is crucial, and the inclusion of the absolute date of inception of the fund allows to account for a cohort effect capturing most of the information.

Finally, competing risks model are a necessary improvement since a large part of the exiting funds with disclosed data have an exit which may not be explained by poor performance. Not taking the nature of this exit into account would lead to a severe bias in the estimation. Further research is needed on the subject, on the identifiability conditions of the competing risks (by including for instance a rule taking into account the highwater mark) and on the potential heterogeneity arising through the dynamic nature of the link between performance, AUM, liquidity and survival.

[13]This method is close to the one of Liang and Park (2008) who underlines that AUM and performance are better indicators of failure than status. They identify real failure when three criterions are met: the fund has stopped reporting, performance is deteriorating on the last six months, and AUM is decreasing during the last year. They also use highwater mark to improve the identification of true liquidation.

3.A Tables and Figures

3.A.1 Descriptive Statistics

Strategy	**Alive**	**Dead**	**Total**
All Funds	6121	8102	14223
Fund of Funds	2158	1653	3811
Single Funds	3963	6449	10412
Long/Short Equity Hedge	1529	2024	3553
Equity Market Neutral	212	415	627
Event Driven	281	489	770
Dedicated Short Bias	16	33	49
Fixed Income Arbitrage	159	310	469
Convertible Arbitrage	64	198	262
Emerging Markets	307	323	630
Multi-Strategies	580	517	1097
Options Strategies	9	2	11
CTA-Managed Futures	497	1784	2281
Global Macro	252	344	596
Other-Undefined	57	10	67

Table 3.1: Number of funds for single risk estimation

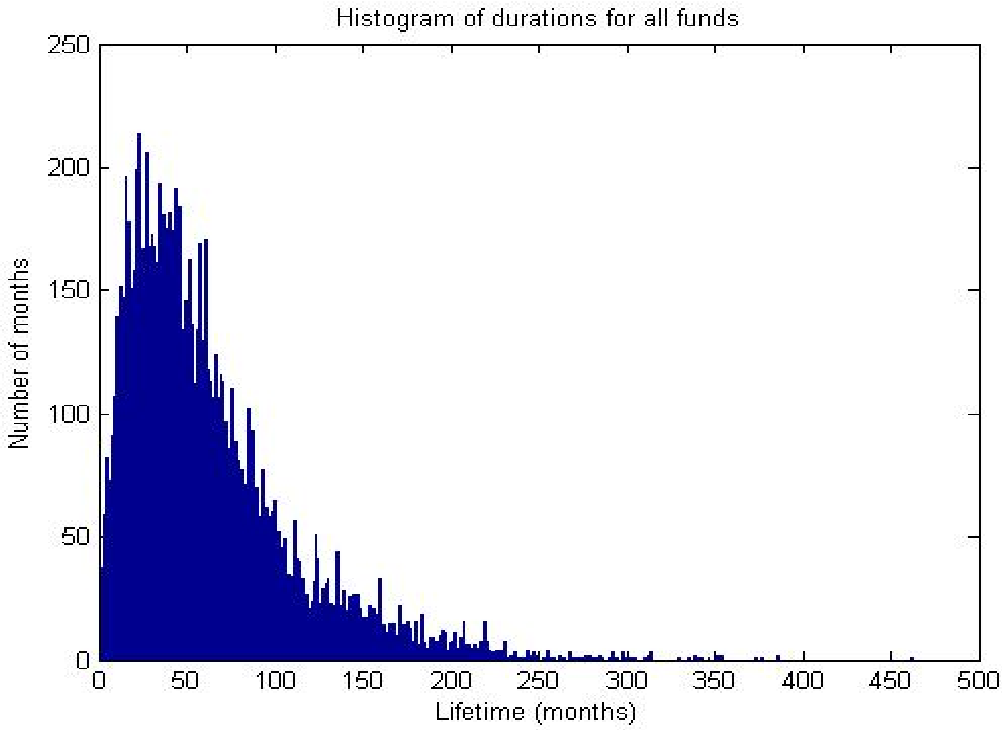

Figure 3.1: Empirical distribution of durations (in months).

Strategy	**Mean**	**Qu.**(5%)	**Qu.**(95%)	**Mode**	**Alive**	**Dead**
All Funds	62, 0	11	158	23	6121	8102
Fund of Funds	61, 0	12	149	23	2158	1653
Single Funds	62, 4	11	160	35	3963	6449
Long/Short Equity Hedge	62, 1	12	154	35	1529	2024
Equity Market Neutral	54, 8	12	134	24	212	415
Event Driven	70, 4	12	184	61	281	489
Fixed Income Arbitrage	59, 1	12	144	42	159	310
Emerging Markets	61, 2	12	160	17	307	323
Multi-Strategies	53, 3	8	145	22	580	517
CTA-Managed Futures	69, 3	10	189	29	497	1784
Global Macro	51, 8	7	135	44	252	344

Table 3.2: Durations (in months) statistics and corresponding number of funds.

3.A.2 Single risk estimation : empirical results

3.A.2.1 Parametric estimation

Strategy	μ	σ	**Mode**	ρ	κ
All Funds	$4,2762$	$0,5638$	$35,0$	$0,0139$	$1,7737$
	$[4,2581;4,2944]$	$[0,5538;0,5739]$			
Fund of Funds	$4,4859$	$0,5637$	$43,2$	$0,0113$	$1,7740$
	$[4,4469;4,5250]$	$[0,5419;0,5864]$			
Single Funds	$4,2073$	$0,5615$	$32,9$	$0,0149$	$1,7808$
	$[4,1866;4,2279]$	$[0,5504;0,5729]$			
Long/Short Equity Hedge	$4,2733$	$0,5669$	$34,6$	$0,0139$	$1,7639$
	$[4,2369;4,3097]$	$[0,5470;0,5876]$			
Equity Market Neutral	$4,0348$	$0,5136$	$31,6$	$0,0177$	$1,9470$
	$[3,9598;4,1099]$	$[0,4746;0,5558]$			
Event Driven	$4,3104$	$0,5382$	$39,0$	$0,0134$	$1,8579$
	$[4,2385;4,3824]$	$[0,5004;0,5790]$			
Fixed Income Arbitrage	$4,1287$	$0,4818$	$37,4$	$0,0161$	$2,0754$
	$[4,0465;4,2109]$	$[0,4399;0,5278]$			
Emerging Markets	$4,3617$	$0,5499$	$39,7$	$0,0128$	$1,8184$
	$[4,2727;4,4508]$	$[0,5042;0,5998]$			
Multi-Strategies	$4,2603$	$0,5979$	$31,0$	$0,0141$	$1,6725$
	$[4,1852;4,3353]$	$[0,5574;0,6413]$			
CTA-Managed Futures	$4,0821$	$0,5834$	$27,2$	$0,0169$	$1,7140$
	$[4,0390;4,1251]$	$[0,5615;0,6062]$			
Global Macro	$4,0984$	$0,5575$	$29,9$	$0,0166$	$1,7938$
	$[4,0103;4,1865]$	$[0,5115;0,6077]$			

Table 3.3: Single risk parametric estimation for a log-logistic specification.

3.A.2.2 Comparison between parametric and nonparametric intensities

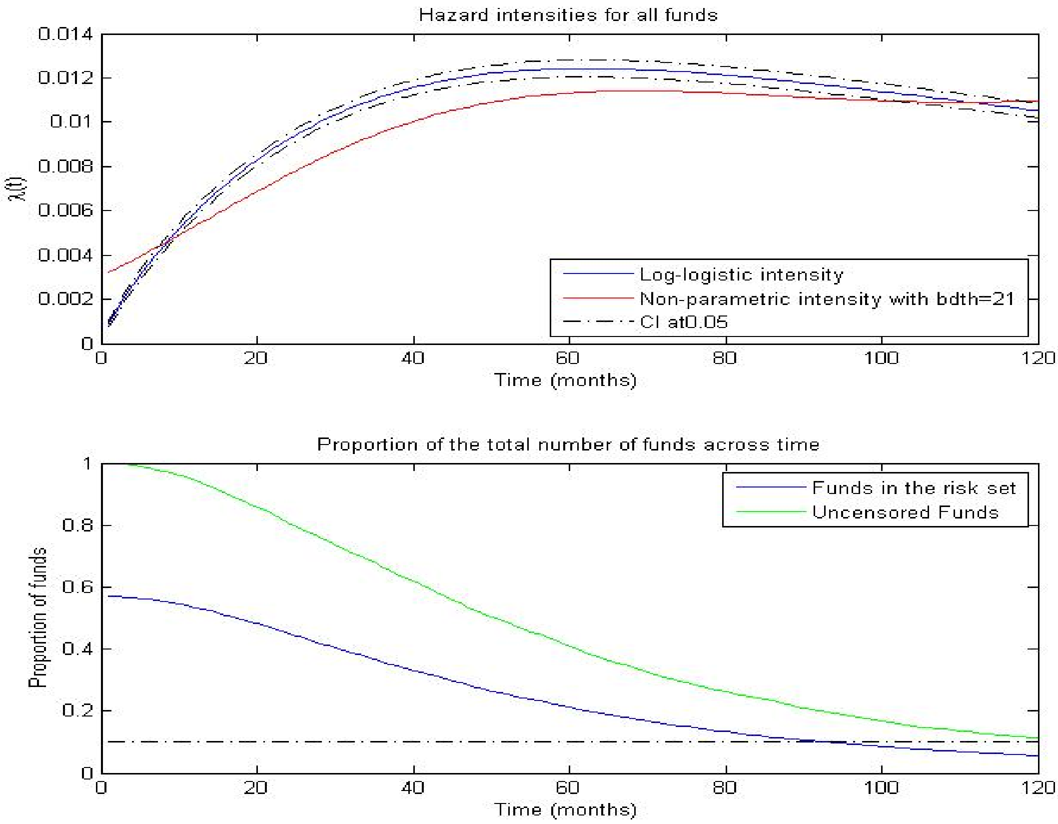

Figure 3.2: Parametric and nonparametric intensities for all funds.

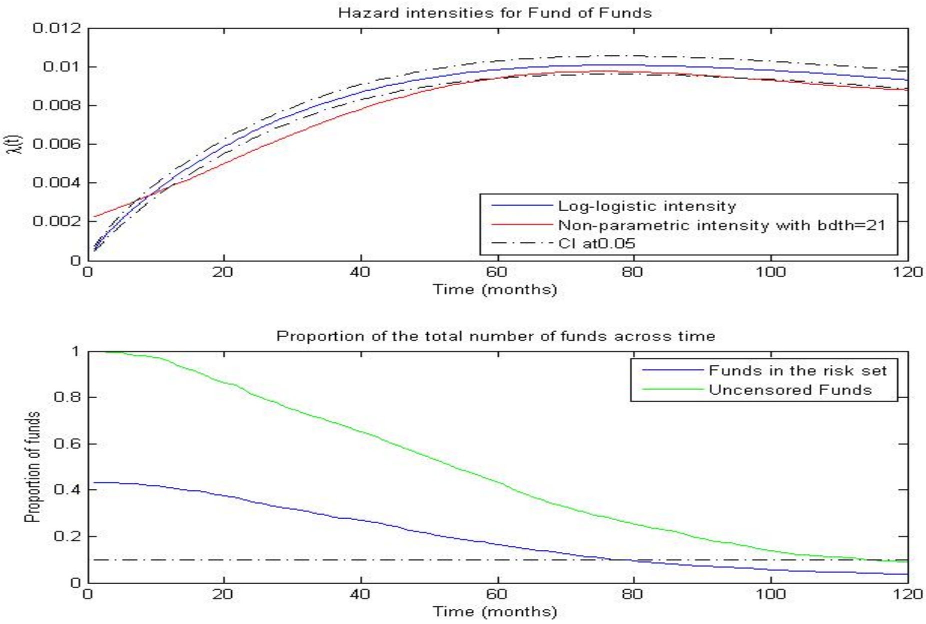

Figure 3.3: Parametric and nonparametric intensities for fund of funds.

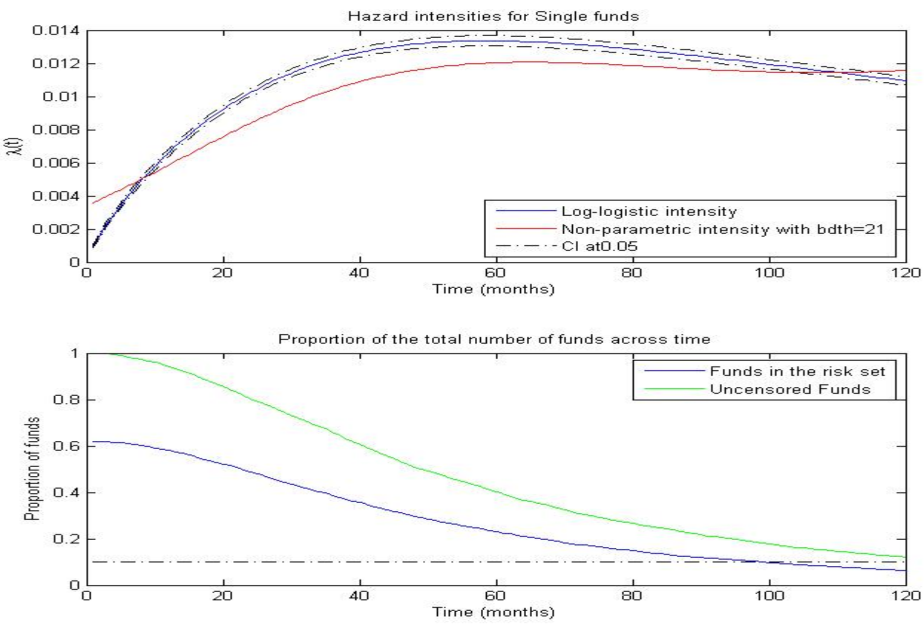

Figure 3.4: Parametric and nonparametric intensities for single funds.

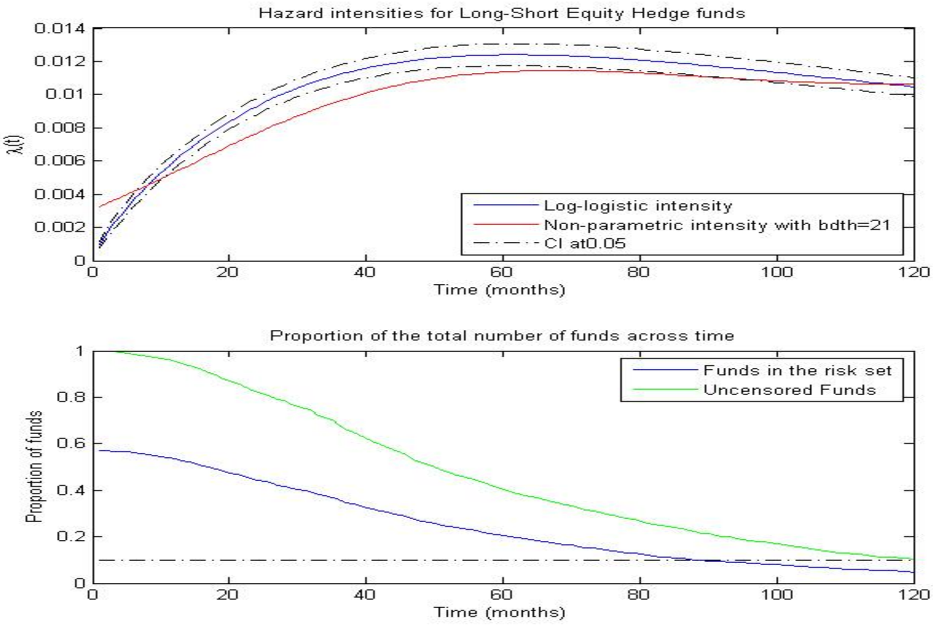

Figure 3.5: Parametric and nonparametric intensities for Long-Short Equity Hedge funds.

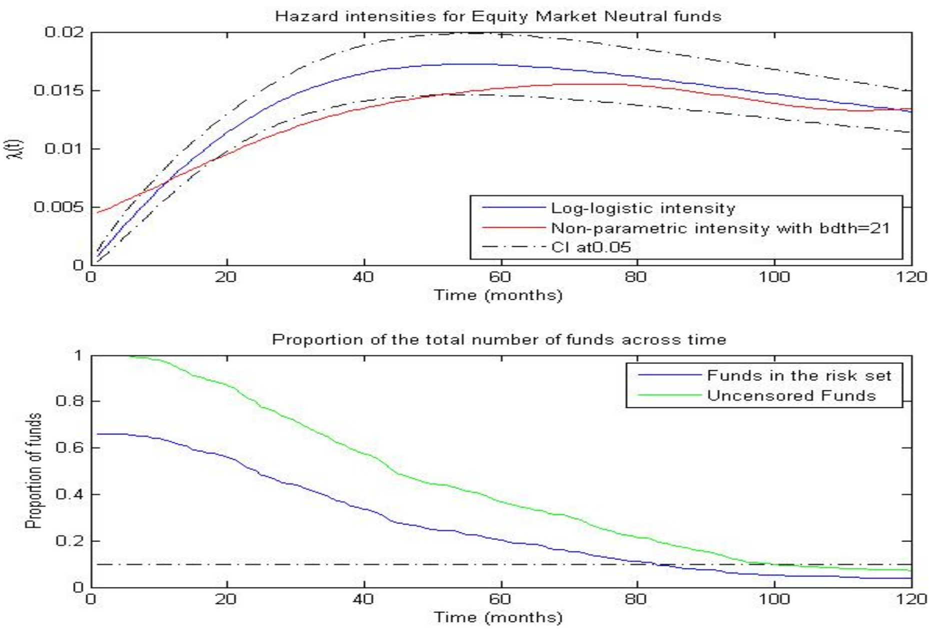

Figure 3.6: Parametric and nonparametric intensities for Equity Market Neutral funds.

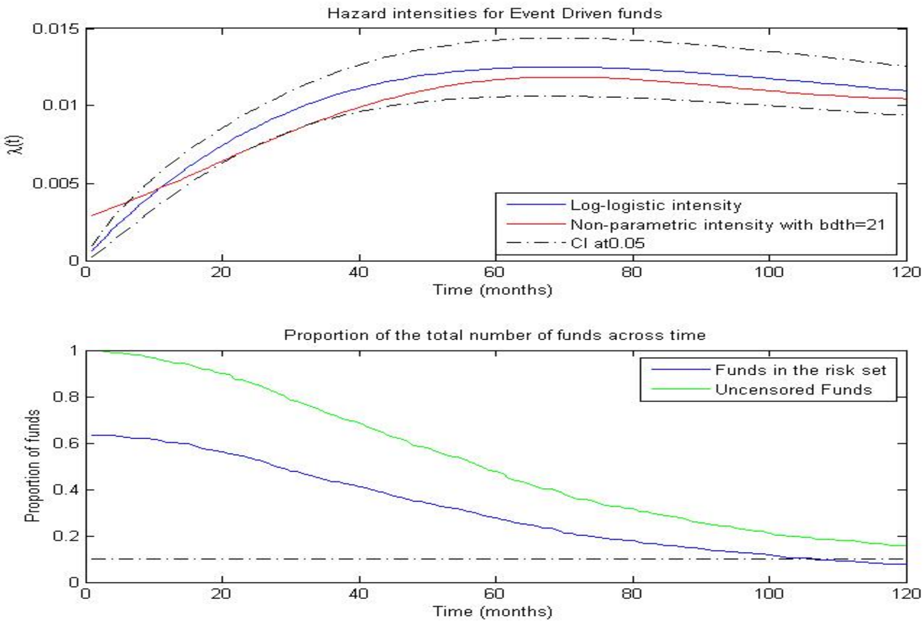

Figure 3.7: Parametric and nonparametric intensities for Event Driven funds.

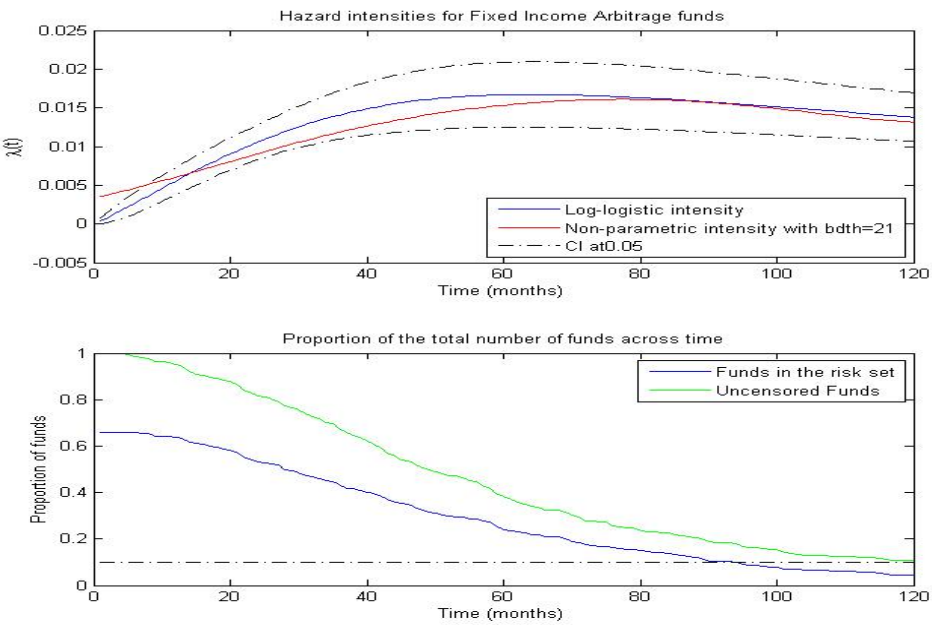

Figure 3.8: Parametric and nonparametric intensities for Fixed Income Arbitrage funds.

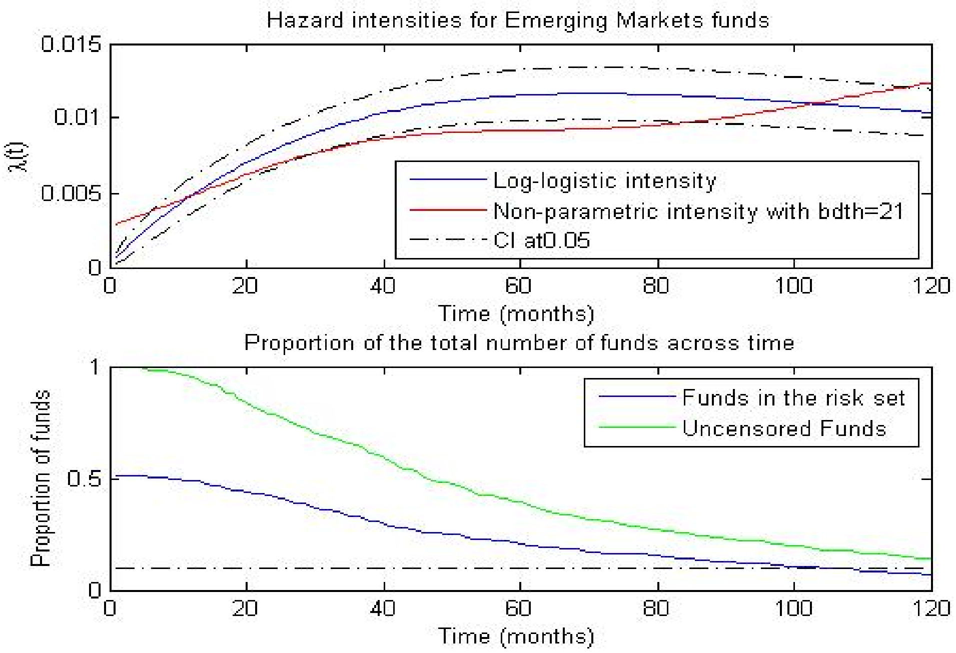

Figure 3.9: Parametric and nonparametric intensities for Emerging Markets funds.

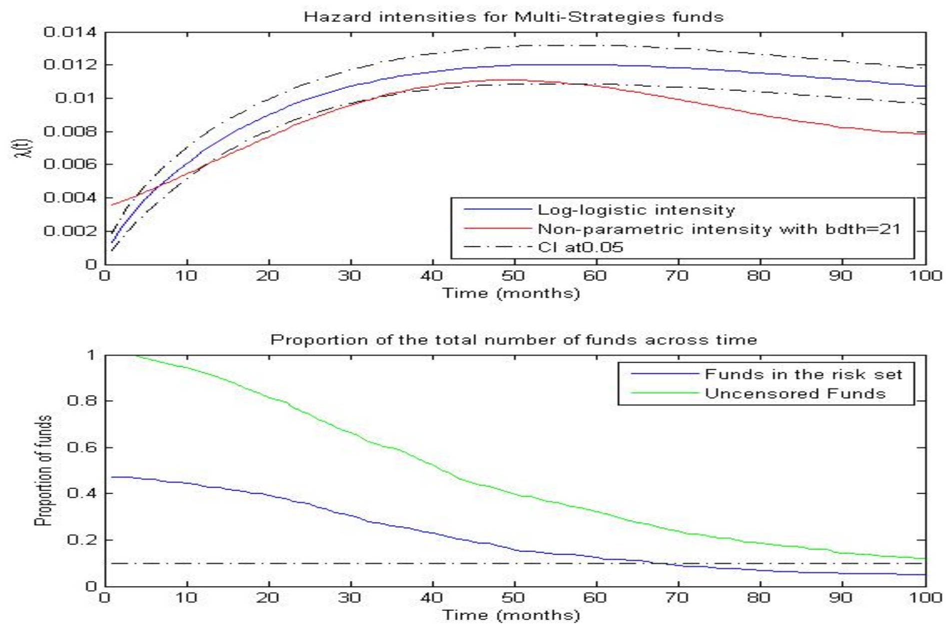

Figure 3.10: Parametric and nonparametric intensities for Multi-Stategies funds.

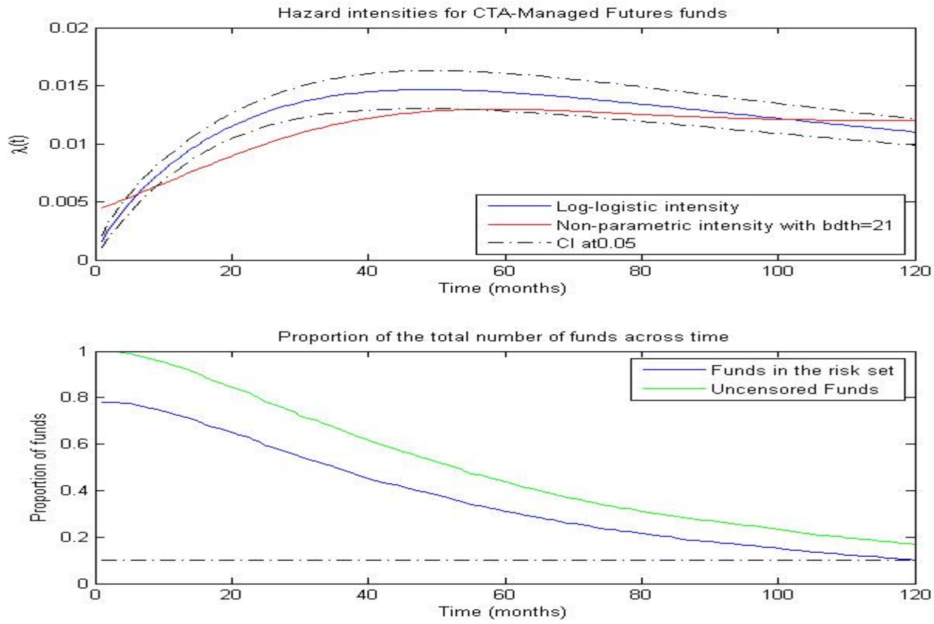

Figure 3.11: Parametric and nonparametric intensities for CTA-Managed Futures funds.

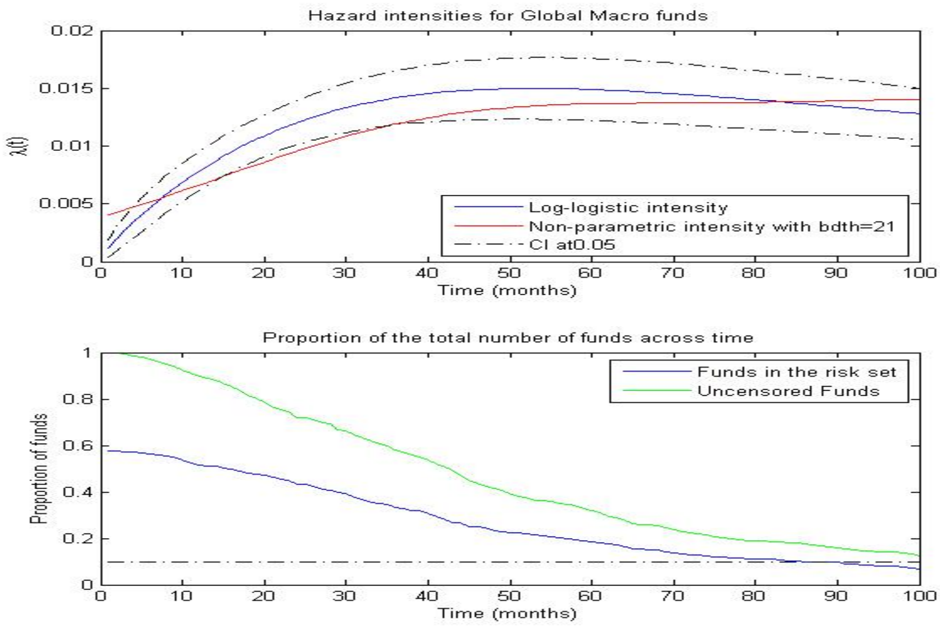

Figure 3.12: Parametric and nonparametric intensities for Global Macro funds.

3.A.2.3 Single risk estimation with covariates

Strategy	**Alive**	**Dead**	**Total**
Single Funds	1075	2487	3562
Fund of Funds	281	415	696

Table 3.4: Number of funds for a single risk analysis with dynamic covariates.

Var. 1	**Var. 2**	**Var. 3**	β_1	β_2	β_3
Equity	-	-	$-0,216$	-	-
			$[-0,259;-0,174]$		
ln(AUM)	-	-	$-0,239$	-	-
			$[-0,297;-0,182]$		
Equity	*ln(AUM)*	-	$-0,144$	$-0,218$	
			$[-0,176;-0,112]$	$[-0,279;-0,157]$	
Equity	*ln(AUM)*	*Date Ref*	$-0,0946$	$-0,2863$	$4,20.10^{-4}$
			$[-0,123;-0,066]$	$[-0,351;-0,221]$	$[3,14.10^{-4};5,26.10^{-4}]$
Returns	Δ *ln(AUM)*	-	$-3,61$	$-0,716$	-
			$[-5,12;-2,10]$	$[-1,06;-0,367]$	-
Returns	Δ *ln(AUM)*	*Date Ref*	$1,41.10^{-4}$	$-2,03.10^{-4}$	$3,65.10^{-4}$
			$[-2,23;13,6]$	$[-0,841;0,841]$	$[2,63.10^{-4};4,65.10^{-4}]$

Table 3.5: Single risk estimation for single funds with dynamic covariates.

Var. 1	**Var. 2**	**Var. 3**	β_1	β_2	β_3
Equity	-	-	-0.035	-	-
			$[-0,057;-0,014]$		
ln(AUM)	-	-	$-0,202$	-	-
			$[-0,263;-0,141]$		
Equity	*ln(AUM)*	-	$-0,015$	$-0,199$	-
			$[-0,026;-0,004]$	$[-0,263;-0,135]$	
Equity	*ln(AUM)*	*Date Ref*	$0,0031$	$-0,2948$	$4,69.10^{-4}$
			$[-0,003;0,009]$	$[-0,363;-0,227]$	$[3,53.10^{-4};5,84.10^{-4}]$
Returns	Δ *ln(AUM)*	-	$-6,38$	$-0,925$	-
			$[-9,26;-3,49]$	$[-1,23;-0,618]$	-
Returns	Δ *ln(AUM)*	*Date Ref*	$-2,18.10^{-4}$	$-2,06.10^{-4}$	$3,93.10^{-4}$
			$[-4,13;4.13]$	$[-0,770;0,770]$	$[2,77.10^{-4};5,10.10^{-4}]$

Table 3.6: Single risk estimation for fund of funds with dynamic covariates.

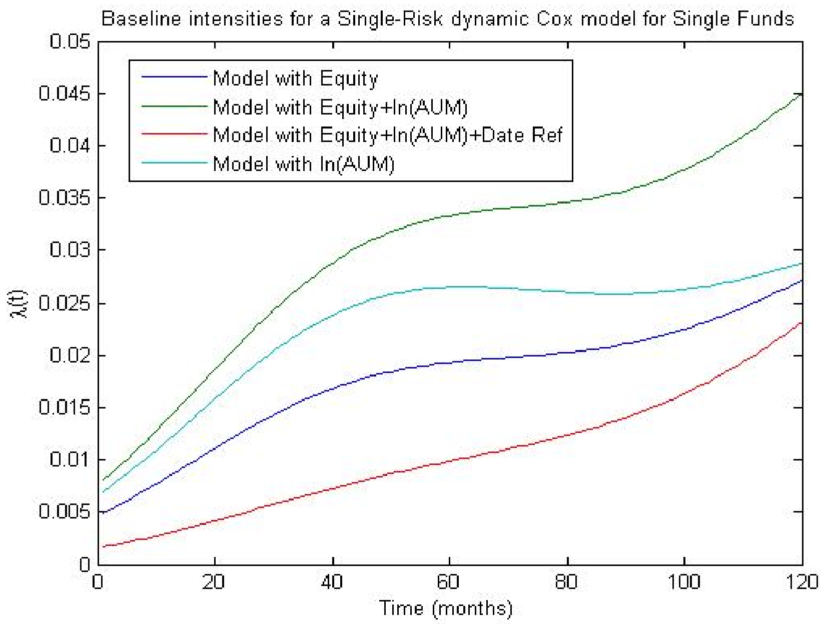

Figure 3.13: Nonparametric baseline intensities of nested dynamic Cox models for single risk estimation for single funds.

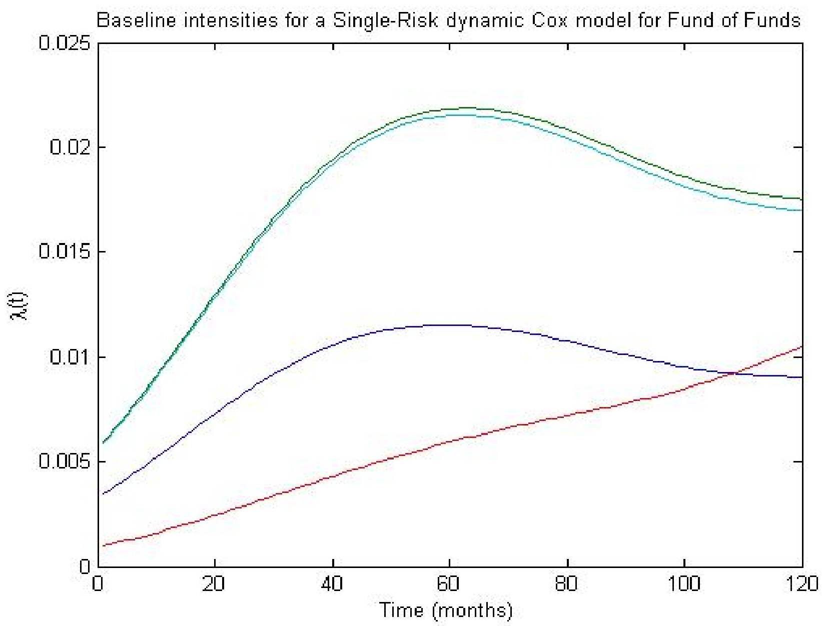

Figure 3.14: Nonparametric baseline intensities of nested dynamic Cox models for single risk estimation for fund of funds.

3.A.3 Competing risks

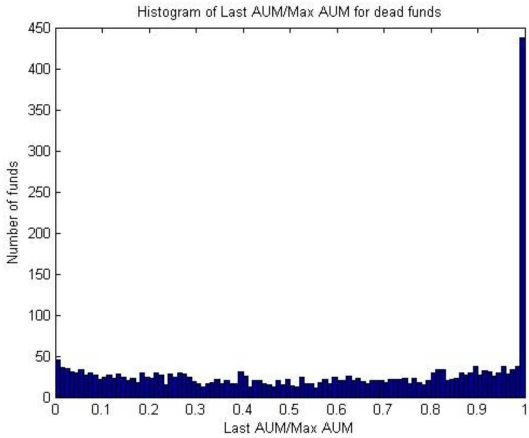

Figure 3.15: Histogram of Last AUM/Max AUM for all dead funds

Strategy	Censored	T_+	T_-
Single Funds	1075	1018	1469
Fund of Funds	281	215	200

Table 3.7: Resulting number of funds for each category of risk.

Single Funds					
T_+					
Var.1	**Var.2**	**Var.3**	β_1	β_2	β_3
Equity	*Ln(AUM)*	*Date Ref*	$0,0592$	$0,186$	$5,42.10^{-4}$
T_-					
Var.1	**Var.2**	**Var.3**	β_1	β_2	β_3
Equity	*Ln(AUM)*	*Date Ref*	$-3,61.10^{-3}$	$-0,364$	$3,36.10^{-4}$
Fund of Funds					
T_+					
Var.1	**Var.2**	**Var.3**	β_1	β_2	β_3
Equity	*Ln(AUM)*	*Date Ref*	$-0,115$	$3,00.10^{-2}$	$4,85.10^{-4}$
T_-					
Var.1	**Var.2**	**Var.3**	β_1	β_2	β_3
Equity	*Ln(AUM)*	*Date Ref*	$0,0175$	$-0,568$	$4,50.10^{-4}$

Table 3.8: Parameter estimates for a competing risks model with covariates.

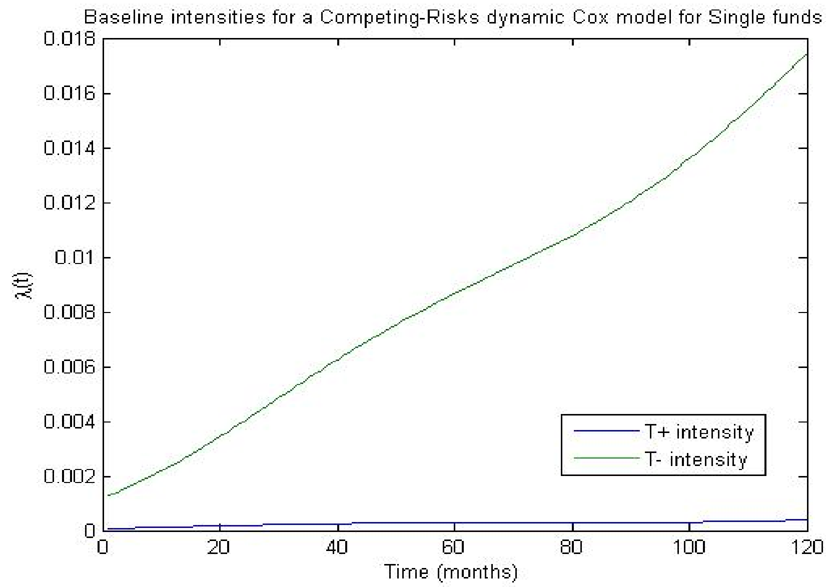

Figure 3.16: Nonparametric baseline intensities of competing risks for single funds for a dynamic Cox model (covariates : Equity, ln(AUM), Date Ref).

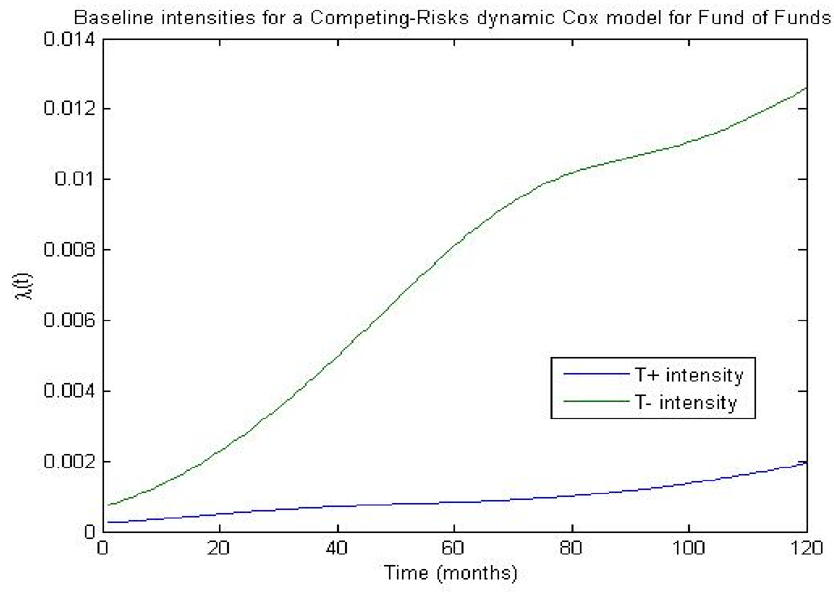

Figure 3.17: Nonparametric baseline intensities of competing risks for fund of funds for a dynamic Cox model (covariates : Equity, ln(AUM), Date Ref).

3.A.4 Academic Contributions

We sum-up here the approaches of the main academic contributions on Hedge Fund lifetimes, depending on the chosen model (parametric or nonparametric approach, use of covariates, dynamic or static covariates, etc.). The legend is the following :

- NDM : No Duration Model
- PWC : Parametric without Covariates
- NPWC : Nonparametric Without Covariates
- PSC : Parametric with Static Covariates
- PDC : Parametric with Dynamic Covariates
- SPSC : Semi-Parametric with Static Covariates
- SPDC : Semi-Parametric with Dynamic Covariates
- CR : Competing Risks
- DCM : Discrete Choice Model

	NDM	PWC	NPWC	PSC	PDC	SPSC	SPDC	CR	DCM
Amin and Kat (2002)	$\surd$								
Lunde et al. (1999)							$\surd$		$\surd$
Baquero et al. (2005)	$\surd$								$\surd$
Boyson (2002)							$\surd$		
Fung and Hsieh (1997b)	$\surd$								
Ang and Bollen (2010)					$\surd$		$\surd$		
Barès et al. (2007)			$\surd$						
Brown et al. (2001)							$\surd$		
Gregoriou (2002)			$\surd$			$\surd$			
Grecu et al. (2007)		$\surd$				$\surd$			
Rouah (2005)							$\surd$	$\surd$	
Liang and Park (2008)							$\surd$		

3.A.5 Empirical method for missing data

Working with dynamic covariates, as soon as a fund does not provide data for a given month, it cannot enter the estimation, as the partial likelihood procedure needs at any date the value of the covariate for all fund in the risk set. Undisclosures may be caused by the fact that funds may forget to report at some dates, due to exogenous reasons. More precisely, and contrary to performance which is quite well reported, AUM may be often missing. Our correction method will then only be focused on treating missing AUM.

We correct data in two cases. First, when AUM is explicitly provided at the beginning and at the end of the lifetime of the fund. Then it happens that AUM is missing for isolated dates. If missing data represent less than 33% of the total length of the track, they are replaced by the first next available AUM. If AUM is missing at the end of the track, the data are forward-filled and replaced by the last available AUM.

An important lack of data may also occur at the beginning of the life of the fund. As in this period of the life of the fund AUM is particularly volatile, we only keep funds with missing data inferior to six months.

Chapter 4

Does Lockup Increase Hedge Funds Lifetimes?

Abstract

Hedge fund liquidity for the investor is often modelled using share restrictions such as lockup (impediment on the withdrawal of initial capital), redemption notice and redemption frequency (time of request and time to wait before an investor gets her money back). Those variables are frequently used along with share restrictions or managerial incentives in econometric studies in conditional duration models. Their potential endogeneity is not tackled if not explicitly discarded without minimal justifications. It is shown in this chapter that (i) the lockup variable is endogenous, that (ii) using redemptions variables as instruments allows to estimate a causal effect for setting up a lockup on the survival of the fund (larger than the duration of the lockup itself) and that (iii) this effect is in line with a "rebound" *effect for funds complying into the lockup settlement. Such a study (identification of the endogenous nature of the lockup and quantification of its effect), up to our knowledge, has never been published.*

4.1 Introduction

The liquidity of a Hedge Fund depends on the contractual features that specify if and how the assets of an investor can be withdrawn from the fund. This reduced liquidity on the investor's side is claimed to help managers in their investment procedure: reducing the liquidity of capital withdrawal is assumed to allow less liquid investments, and thus to access asset classes with a higher return. This liquidity is mainly characterized by share restrictions such as lockup and redemption constraints. A lockup specifies a period during which it is impossible for the investor to redeem its initial shares. After this period, the investor must notify her redemption sufficiently in advance (*redemption notice*) before the redemption is processed. The processing itself is possible at a given frequency (*redemption frequency*) that may be days, weeks, or months. Additionally with *gates*, the manager may have the possibility to discretionary limit the redemption to a certain proportion (or to block the process in an extreme situation).

As discussed in Chapter 3, Hedge Funds report voluntarily in databases and two reasons may explain that a fund exits a database. In this chapter we will be interested in the exit caused by poor performances. In order to prevent huge downfalls in the amount of asset under management, the manager has the possibility to set a lockup on her fund. We argue here that the nature of this lockup is endogenous: the manager seeks an *ex-ante* protection to prevent the fund from massive capital withdrawals and then from liquidation. In practice, the manager decides herself whether she sets a lockup or not.

This depends on her level of own-confidence, to prevent her fund from failure and thus to increase its survival probability. The decision of setting a lockup is not neutral, as it partly reveals manager's own confidence. It is also a restriction for the investors, resulting in a reduced liquidity. Empirically, it is however often viewed as an advantage by the investor, as it prevents other investors to get their money back and then protects the fund from failure in bad times. Then, setting a lockup is not compulsory considered as a drawback from a commercial perspective.

Information being limited concerning Hedge Funds, the true ability of the manager cannot be observed. The lock-up may then be viewed as a treatment into which the manager self-selects. The question that arises naturally is: what is the effective causal effect of this binary treatment variable on the duration of a fund in a database? Such a question is, up to our knowledge, not asked in the literature. Our approach is therefore different from most academic papers that are often more concerned by the inclusion of managerial incentives, viewed as exogenous proxies for manager's ability in conditional survival models.

Section 4.2 states the problem and links it to the common stylized facts of Hedge Funds analysis. We also define precisely the objects of our study and review the related literature. Section 4.3 explores causal effects and presents how to deal with endogeneity in this precise case. Section 4.4 tests this framework on Hedge Funds data to estimate the causal effect in terms of lifetimes of a lockup. We find promising and coherent result accounting for a treatment effect of lockup settlement.

4.2 Data and definition

4.2.1 Hedge Funds lifetimes

We use here the same database as the one presented in Chapter 3, that is the TASS database of June 2009, with data at the end of May 2009, with the same features. The lifetime of a fund is again computed as the difference in months between the death or current date and the initial date as in Section 3.2, with the same assumptions. Censorship and durations are again assumed independent. If not, we could not identify the law of the observed durations. This seems a coherent hypothesis in our context, since censorship occurs through the current observation date.

4.2.2 Share restrictions

Lockup and redemption periods are share restrictions. They are sometimes aggregated in some studies (Agarwal et al. (2009)), but we will explicitly separate them in our analysis (as in Ang and Bollen (2010) who distinguish lockup from notice periods). An investment then obeys to the following process. After an initial investment, in presence of a lockup, money is blocked for a specified amount of time (typically one year). During this period, the investor can't go out of the fund, regardless of the performances of the fund. This feature is negotiated at inception of the contract between the investor and the fund manager. This is particularly adapted to young funds with short-track records and uncertainty on the manager's skills.

After this period (if there is any), when the investor decides to redeem her shares, she has to do with a *redemption notice* which is the contractually set period of time before the fund processes to redemption. This does not mean that the investor receives her money after that: this is the case when the *redemption frequency* is attained. It may be a time or a date (quarterly, annually) at which redemption is effective. For example, if redemption is annually in December, and processed, say, in January, the investor has to wait nearly a year before getting her money back.

Intuitively, the longer these periods, the greater the illiquidity of the fund. In practice, this illiquidity is associated to a higher flexibility for the manager and with superior performance (Agarwal et al. (2009)). As it allows the manager to make illiquid bets on a longer term, those instruments are assumed to provide higher returns. Investors can then take benefit from those share restrictions. Aragon (2007) estimates that the reward for such restrictions is around $4-7\%$ a year, and this is the price for the so-called illiquidity premium. The positive relation between lockup and hedge fund returns has been also underlined by Ackermann et al. (1998) and Liang (1999). Ding et al. (2009) also tries to evaluate the empirical links between returns and capital flows taking into account the redemption restrictions.

However, one may think that such restrictions have also a cost. This cost may be interpreted as exit fees. Such a cost has been estimated to be around 1.5% of the initial investment by Ang and Bollen (2010) for a two-years lock-up and 3-months notice period. This cost is then inferior to the average benefit of restrictions. However, this changes dramatically when the manager can proceed to *gates*, that is to say the possibility to stop totally or partially any AUM withdrawal (in this case the cost can go up to 10 or 15 % of the initial investment).

Our aim is then to consider lockup as a static, endogenous variable, in the framework of duration models. Lerner and Schoar (2009) have already tried to explore whether a lockup is part of an optimal incentive contract for fund managers. Agarwal et al. (2009) argues that managerial incentives, share restrictions, and fund features chosen by the manager, are probably exogenous and discards endogeneity of those variables in the analysis of performance. This could be subject of discussion as even managerial incentives could be viewed as being endogenous. Skilled managers with a long-term view will set high incentive fees as a price to the access to their fund; short-term, low-skilled managers may also be tempted to set high fees as their time-horizon (and their anticipated survival) is short, then they could try to accumulate as much money as possible in a short time. We focus however on the lockup and follow then the opposite approach, considering this variable as being deeply endogenous.

We are then close to the approach of Hombert and Thesmar (2009) which shows that arbitrageurs with stable funding (proxied by share restrictions for Hedge Funds) can encounter some low performance periods without fearing fund liquidation, and will even outperform other funds in the future. The funds performing poorly on the short-term can still maintain their (potentially profitable) long-term investments. This is an argument for a *treatment effect* of lockup that protects initial investment and decreases the probability to fail. However, such an endogeneization of the lockup variable is only possible when there is no overwriting of this variable in databases through time (as it is static and that data providers do not keep data of historic share restrictions). Liang (1999) and Aragon (2007) have shown that the empirical portion of funds modifying a posteriori restriction characteristics is really reduced (from 1% to 10%) but the main modifications on the lockup variable are made on its duration, not on its nature (0 versus 1).

4.3 Duration models and causal effects

We will work in a single risk framework and the only risk we are concerned with is the risk of exit from the database because of failure (including merging, liquidation, etc.). The mathematics needed for duration analysis have been used in Chapter 3 and are exposed in Appendix A.3. But such models are conditional and only describe the instantaneous probability for funds to fail. To hold they assume that the covariates are exogenous, which is a drawback in our context. Other models are then needed to provide a causal interpretation of the effect of those variables.

4.3.1 Causal effects using instrumental variables

The treatment effect literature aims at estimating the causal effect of a treatment variable D (here lockup) on an outcome of interest Y (duration of the lifetime of the fund). The scope of such studies ranges from medical treatment to social policies evaluation. What sophisticates the analysis is of course when the assignment of the treatment variable is correlated to the outcome and when the agent under treatment may self-select in the treatment program, potentially anticipating the outcome. We face this situation in our framework since anticipating a short lifetime Y of her fund, the manager self-decides to set a lock-up D to protect the fund from massive capital withdrawals. We will note $D = 0$ if there's no lock-up, $D = 1$ if there is one. Of course, we ***cannot*** identify any treatment or lockup effect by estimating a quantity such as $\mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0]$ since the assignment of D strongly depends on anticipated Y.

4.3.1.1 Counterfactual outcomes and treatment effects

We suppose that for a given individual i, variable Y has two potential outcomes Y_i^1 and Y_i^0 under the two potential values of the variable D. As we can observe only a given outcome at a time ($D = 0$ or $D = 1$) the two are considered as being counterfactual. We can only observe $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$. The individual-level causal effect is then equal to $\delta_i = Y_i^1 - Y_i^0$ but cannot be observed as either one of the $Y_i^{1,0}$ does not exist. We must switch to the population level and estimate $\mathbb{E}[\delta] = \mathbb{E}[Y^1] - \mathbb{E}[Y^0]$. The literature commonly writes :

$$Y = \mu_0 + D(\mu_1 - \mu_0) + \{v_0 + D(v_1 - v_0)\}$$

with $\mathbb{E}[\delta] = \mu_1 - \mu_0$, $\mu_i = \mathbb{E}[Y^i]$ and $v_i = Y^i - \mathbb{E}[Y^i]$.

The seminal idea of Heckman and Robb (1986) was to introduce observable variables to explain the treatment selection D and break the dependence between D and the error term. However, when individuals anticipate the outcome Y, they self-select into D, and we face *selection on the unobservables.* This is the main reason that explain that the usual regression:

$$Y = \alpha + \delta D + \epsilon$$

will lead to severe biases, as D is known to be correlated to ϵ since D is self-selected. In this respect, several treatment effects have been defined. First, the *Average Treatment Effect* (henceforth ATE) is defined as $ATE = \mathbb{E}[\delta] = \mathbb{E}[Y^1] - \mathbb{E}[Y^0]$. Second, the *Average Treatment Effect on the Treated* (henceforth ATT) is defined through $ATT = \mathbb{E}[\delta|D = 1]$ and is the mean effect on the population that has chosen to follow the treatment. Conversely the *Average Treatment Effect on the Controls* (ATC) writes $ATC = \mathbb{E}[\delta|D = 0]$. ATE is the hypothetical difference of the two outcomes if one could observe a random individual (in the total population) in the two universes (treatment and no treatment). This is the more general effect that we can conceive, and as other effects will be developed, they will correspond to more and more "local" views. ATT is the same quantity but for a population of individuals that are in the group with $D = 1$.

The first approach in order to evaluate the causal effect of the causal variable D on the outcome of interest Y, is to condition on some variables in order to block all the potential causal effects of other perturbating variables. This includes techniques known as matching, weighting, regression or stratification. The second approach is to use an instrumental variable (IV). As explained before the IV W is chosen to be related with D, with an effect on Y but only through D. The causal effect of D on Y is evaluated through the relative response in Y related to changes in D following an exogenous variation of W. The traditional IV estimator is for an instrument in the scalar case W:

$$\hat{\delta}_{IV} = Cov(W, Y)/Cov(W, D).$$

4.3.1.2 The framework of Angrist and Imbens

The appropriate framework for a causal analysis in our context is developed in Angrist and Imbens (1994). They show how under suitable assumptions, the IV estimator may be interpreted as a causal effect, but **only** for a limited subset of individuals, called the *compliers* which constitutes however an unindentifiable subpopulation. In this situation, the causal effect estimated through the IV estimator is the LATE (Local Average Treatment Effect, called local as it only applies for a subgroup of the population). The LATE may be considered as the average treatment effect for the individuals where treatment incitation is driven by monotonous shifts in an exogenous instrument which satisfies some specific conditions which rely on the potential outcomes that are only indexed by the treatment. The first technical assumption is of course that the instrument W is a valid one :

Assumption A0 : $Y = \beta D + U$ with $cov(D, W) \neq 0$ and $cov(D, U) = 0$.

Assumption A1 : W is a random variable such that for all value w of the support of W, and all individual i, the vector $(Y_i(0), Y_i(1), D_i(w))$ is jointly independent of W_i. Moreover, it is required that $\mathbb{E}[D_i|W_i = w]$ is a non-degenerate function of w.

We will be able to check the second part of Assumption A0, but it will not be possible to test its first part as the random assignment of Z_i will not ensure this (and this part of the assumption is not itself sufficient to estimate any causal effect).

Assumption A2 : For all individual i and w_1, w_2 in the support of W, we have either $D_i(w_1) \leq D_i(w_2)$ or $D_i(w_1) \geq D_i(w_2)$.

Assumption A2 ensures that the instrument influences the selection in a monotonic way. It allows that the treatment effect of those who move in the treatment is not reduced by an opposite effect induced by a potential existence of a subpopulation that moves out of the treatment, following a shift in W. Unfortunately, this assumption is again untestable. This model has been also developed by Angrist et al. (1996) (who split and refine the assumptions of Angrist and Imbens (1994)): and in the case of a binary instrument they describe the population in four subgroups depending on the individual reactions to shifts in the instruments. In our framework, their assumptions translate as follows :

Assumption A'1 : The potential lifetimes $Y_i^{0,1}$ for each fund i are unrelated to the treatment status of the other funds (SUTVA - Stable Unit Assumption).

Assumption A'2 : The assignment of the instrument W is random (Random Assignment).

Assumption A'3 : Whatever the value of the instrument, the outcome first depends on D i.e. for $w \neq w'$: $Y(D, w) = Y(D, w')$ (Exclusion Restriction).

Assumption A'4 : The average causal effect of D on Y is non-zero.

Assumption A'5 : For each individual, D is monotonically increasing as a function of W (Monotonicity).

We give here a more simple version of the original set of assumptions which are technically detailed in Angrist et al. (1996). Under $A0; A'1 - A'5$ the LATE computed as the IV estimator has a causal interpretation and the population may be described in the following way. The "never takers" never self-select into the treatment program whatever the value of the instrument, on the contrary to the "always-takers" that always take the treatment. The "defiers" are reluctant to take the treatment and switch to the non-treatment situation with W increasing. The last category is made of the "compliers" who switch to treatment participation with an increasing value of the instrument. Assumption $A'5$

excludes “defiers” and the two first categories are characterized by a causal effect which is equal to zero. $A'4$ ensures that the overall population has a non-zero proportion of compliers. The only limitation is that the individuals that are members of this category are not identifiable. Then the LATE is only the treatment effect for compliers but not for the whole population, nor for an observed part of it.

With a binary instrument, the LATE and the ATT match only if we assume that the causal effect is the same for every individual. When this is not the case, we will not be able to identify the ATT unless we have a continuous instrument W such as $\mathbb{E}[D = 1|W]$ covers all the values between 0 and 1. For this see Heckman and Vytlacil (2005) which completely generalizes and unifies the notions of treatment effects. We will check in Section 4.4.3 the validity of our chosen instrument and of the previous assumptions.

4.3.2 Estimation with a counting process approach

We adapt then the approach developed in Chapter 2 to estimate a causal effect in the framework of duration models. $Y = \tau$ is estimated as a function $Y = \phi(U, D)$ with the help of an instrument but using an intensity approach. ϕ is expressed as the solution of Equation 2.12; the causal interpretation comes in a second step.

4.3.2.1 Integral equation for ϕ

Equation 2.12 may be re-expressed in this case through Equations 2.10 or 2.14. The integral equation of interest is then:

$$\boxed{\int_d \int_{t=0}^{\phi(u,d)} \lambda_\tau(t|D = d, W) g(d|\tau \geq t, W) dddt = u.} \tag{4.1}$$

$\lambda_\tau(t|Z = z, W)$ is the conditional intensity of the process relatively to the $Z = D$ and W variables, and $g(z|\tau \geq t, W)$ is the density of the treatment variable $Z = D$ conditionally to W and to the fact that the event of interest has not yet occurred (the fund is still alive and reporting in the database). where $\lambda_\tau(t|D = d, W)$ is the conditional intensity of the process relatively to the D and W variables, and $g(d|\tau \geq t, W)$ is the density of the treatment variable D conditionally to W and to the fact that the event of interest has not yet occurred (the fund is still alive and reporting in the database).

4.3.2.2 Application to causal effect estimation

In our case, $\phi(u, d) = \alpha(u) + \beta(u)d$ as the D variable will be the lockup (binary, 0/1). Then the integral over the support of Z must be understood as a finite sum from 0 to 1. $\lambda_\tau(t|D = d, W)$ may be estimated with a Cox model and then accounts for censorship:

$$\lambda_\tau(t|D = d, W) = \lambda_0(t) exp(\beta_D D + beta_W W).$$

We assume that $g(d|T \geq t, W) = g(d|W)$. Such an assumption will be plausible as soon as W and D are static variables, fixed at inception of the fund. We must notice that at a given date, a database is a picture taken at this date of the state of the industry. We cannot observe if the lockup status of the fund has been modified across time. No data on the evolution of this variable is available, but as expressed before p.108, this in practice variable is not much subject to evolution.

For a fixed $W = w$, $g(d|W)$ is a two-valued vector estimated as :

$$\hat{g}(D|W = w) = (\hat{g}(0|W = w), \hat{g}(1|W = w)) = \Big(\frac{\sum_{i=1}^{N} \mathbf{1}_{(D_i=0, W_i=w)}}{\sum_{i=1}^{N} \mathbf{1}_{(W_i=w)}}, \frac{\sum_{i=1}^{N} \mathbf{1}_{(D_i=1, W_i=w)}}{\sum_{i=1}^{N} \mathbf{1}_{(W_i=w)}} \Big).$$

For a fixed u, the left hand side of Equation (4.1) becomes:

$$\Lambda_0(\alpha(u)) exp(\hat{\beta}_W W)\hat{g}(0|W) + \Lambda_0(\alpha(u) + \beta(u)) exp(\hat{\beta}_D + \hat{\beta}_W W)\hat{g}(1|W)$$

where $\Lambda_0(\tau) = \int_0^{\tau} \lambda_0(t)dt$.

For a given u we estimate $(\alpha(u), \beta(u)) \in \mathbb{R}^2$ as:

$$argmin_{(\alpha(u),\beta(u))} \sum_{i=1}^{N} \Big[(\hat{\Lambda}_0(\alpha(u)) exp(\hat{\beta}_W w_i)\hat{g}(0|w_i) + \hat{\Lambda}_0(\alpha(u) + \beta(u)) exp(\hat{\beta}_D + \hat{\beta}_W w_i)\hat{g}(1|w_i) - u)^2 \Big].$$

Then LATE is simply:

$$\delta_\phi = \mathbb{E}[\phi(U, 1) - \phi(U, 0)] = \mathbb{E}[\beta(u)] \tag{4.2}$$

which will be easily estimated as we have assumed that $U \sim Exp(1)$.

4.4 Empirical Study

4.4.1 Database and Instruments

A full description of the current database is already provided in Chapter 3. Funds with a lockup represent 17.49% of the database. This figure is of 20.02% for single funds and 10.55% for fund of funds. Corresponding figures are displayed in Table 4.1. The mean lockup duration is around one year: 12.29 months for single funds, 11.91 months for funds of funds, and 12.23 months for all funds. However, the present study cannot be performed on this whole dataset. First, we will only consider funds that disclose the several quantities: inception date, currency code, performance start/end date, redemption frequency and lockup period. In addition, we also present results also by separating funds along with their declared investment style and consider three categories: single funds, funds of funds and all funds. Funds are considered as dead for the database of 05/31/2009 if:

- funds are in the graveyard;
- funds are "*alive*" but do not have reported any figure since 01/31/2009; or funds are "*alive*" but do not have published at the 05/31/2009 but are categorized explicitly as "*no longer reporting*", "*liquidated*", or "*unable to contact fund*".

We explicitly exclude backfillers and focus on funds in USD currency. This will allow us to get similar funds (without redundant shares in other currencies, potential copies of the main share) with homogenous characteristics. Our dataset is reduced as we only consider funds that have no missing data in their returns and AUM[1]. For our analysis, we will take as instrument the redemption frequency[2].

[1]A correction is applied as in Chapter 3 since the AUM track is backfilled in case of missing data. We apply the method if less than the six first months months of data, and less than 1/3 of the total lifetime of the fund are missing. Else the fund is discarded.

[2]This variable is estimated in **days** as the maximal amount of time to wait before in-process redemption is effective. For instance, if the redemption is annually, then this variable will be equal to 365 days.

We must only consider as dead and uncensored, funds leaving the database because of bad performance. Exiting good-performers are then taken as censored for a competing reason. As we wish to model lockup as a treatment against liquidation, we have to distinguish among funds those whose future survival is in danger. As we cannot always identify the reason of the exit, we must set a discretionary criterion, based here on the status of the fund in the database and/or its level of AUM at the time of exit. For all the other uncensored (i.e. dead) funds, we consider that the funds exit because of bad performance, when their AUM at the time of exit is below 200 millions of US dollars or that this value is less than 75% of the maximum value reached by the total AUM. In the opposite case (high AUM above 200 millions, and more than 75% of the maximum value), we consider that the exit of the fund is not caused by bad performance, but because it has reached a sufficient capacity. This is motivated by the observation that the distribution of the variable of last reported AUM divided by maximum AUM is flat with a pronounced peak at one, as suggested by Figure 3.15. Those funds are then set as censored for an other reason, but are of course still included in the risk set.

Our sample is made of 4131 funds. For this sample, the proportion of funds (see Table 4.2) with a lockup is slightly higher than for the global population but yet of the same order: 33.62% for single funds, 26.84% for fund of funds and 32.51% for all funds. The mean lockup duration is again around 12 months (respectively 12.64, 12.28 and 12.59 months). Such figures are very similar to those obtained in Aragon (2007) or Hombert and Thesmar (2009). A description of the lockup periods is given in Table 4.3: lockup of 6 and 12 months are standard and constitutes the major part of the funds with a lockup.

4.4.2 Estimation of causal effect

4.4.2.1 "Naive" estimation

Difference of mean durations

If we compute the difference between the mean duration of the funds with a lockup versus funds without lockup, we estimate a quantity which has *nothing to do* with any causal effect. We call it here $\delta_{NAIVE} = \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0]$. We however compute this quantity as a reference to illustrate the benefits of the treatment approach.

For single funds, we get $\mathbb{E}[Y|D=1] = 72.37$ months, $\mathbb{E}[Y|D=0] = 81.93$ months and $\delta_{NAIVE} = -9.56$ months. For funds of funds $\mathbb{E}[Y|D=1] = 74.71$ months and $\mathbb{E}[Y|D=0] = 81.65$ months, consequently: $\delta_{NAIVE} = -6.94$ months. We obtain for the whole sample $\mathbb{E}[Y|D=1] = 72.69$ months and $\mathbb{E}[Y|D=0] = 81.88$ months with a resulting $\delta_{NAIVE} = -9.19$ months.

Relying on this to build a causal analysis would lead to conclude that lockup is ineffective in extending a fund's life. This is because the counterfactual alternative is not observed for each individual: the two populations (funds with and without lockup) are heterogenous in their nature and in their anticipations. The selection bias is then so strong that the conditional observation (negative difference in duration between the two populations) is the opposite of what a treatment effect should be.

Using a hazard intensity

In fact, using a more sophisticated model including a hazard intensity would lead to the same conclusion, as it remains a conditional models with exogeneity assumptions on the variables. For instance, we could set a Cox model with $\lambda(t) = \lambda_0(t)exp(\beta_D D)$ and λ_0 following a specified density. Doing this we would have then the duration τ which writes:

$$\tau = \phi_{Cox}(U, D) = \Lambda_0^{-1}\Big(\frac{U}{exp(\beta_D D)}\Big)$$

where Λ_0 is the cumulative log-logistic hazard. The pseudo-causal effect would be:

$$\delta_{Cox} = \mathbb{E}_U[\Lambda_0^{-1}\Big(\frac{U}{exp(\beta_D)}\Big) - \Lambda_0^{-1}(U)].$$

With a log-logistic hazard we would get for instance:

$$\Lambda_0(t) = log(1 + (\rho t)^{\kappa}) \quad \longrightarrow \quad \Lambda_0^{-1}(u) = \frac{1}{\rho}\Big(exp(u) - 1\Big)^{\frac{1}{\kappa}}.$$

With $U \sim Exp(1)$, we get the final expression:

$$\delta_{Cox} = \int_0^{\infty} \frac{e^{-u}}{\rho}\Big[\Big(\frac{e^{\beta_D}}{e^{-u}} - 1\Big)^{\frac{1}{\kappa}} - \Big(\frac{1}{e^{-u}} - 1\Big)^{\frac{1}{\kappa}}\Big]du.$$

We see directly that this would lead to similar results than in the previous paragraph. Empirically, conditionally on the lockup variable, the default intensities are higher for funds with a lockup. As we work with a static covariate D with a conditional model, and that the mean duration conditional on D is shifted downwards with $D = 1$, we would get a positive coefficient β_D. The wrong reading would be that setting up a lockup decreases the lifetime of a fund. Such a causal interpretation is not the good one. Lockup **does not** cause the shift in intensity, it only underlines that managers of funds with anticipations, endogenously choose to set a lock-up on the fund. Therefore, such previous procedures do not take into account the potential endogeneity of covariates, and are improper in producing causal statements.

4.4.2.2 Estimation with IV (2SLS)

We follow here the approach described in Section 4.3.1.2. We choose as the outcome the duration of the fund in the database (variable Y), as explicative variable the binary lockup D (0/1). Our chosen instrument W is the redemption frequency. The causal effect δ_{LATE} will be given as the IV estimator in this context, as the result of the following two-stage least squares procedure:

$$\begin{cases} Y &= \gamma_0 + \delta_{LATE} D + U \\ D &= \gamma_1 + \gamma_2 W + V \end{cases}$$

Results are given in Table 4.4. This means that the LATE (the average treatment effect on the compliers) is equal to 44.53 months for single funds, 66.02 months for fund of funds and 48.91 months for all funds. Those quantities are far different from the previous δ_{NAIVE} and all are statistically significant. We will explore in Section 4.4.3 whether we can consider this value as an average treatment on the treated. Unfortunately, this method has a drawback as it cannot take account of censorship and considers durations as random variables and not as the mark for the jump for an underlying stochastic process. This is the precise aim of the next section, in order to precise those results.

4.4.2.3 Estimation with a counting process approach

We use here the approach described in Section 4.3.2 which both considers the endogeneity and the censorship problem. A practical limitation is that after a given threshold for u, the estimation of $\beta(u)$

has no sense[3]. We must set a maximum value M_u for u, that we choose here as being slightly superior to the (perimetrically estimated) cumulated hazard $\hat{\Lambda}_0(M_\tau)$ where M_τ is the maximal duration of all funds of the sample (388 months). Then we only compute $\beta(u)$ for $u \in [0;3]$. We then obtain the estimation of $\beta(u)$ in Figure 4.1 for single and funds of funds, and Figure 4.1 for all funds. The estimated causal effect $\delta_\phi = \mathbb{E}[\phi(1,U) - \phi(0,U)]$ is computed through:

$$\delta_\phi = \frac{\int_0^{M_u} \beta(u)e^{-u}}{\int_0^{M_u} e^{-u} du}.$$

as we cannot observe $\beta(u)$ for an infinite set of u. Then δ_ϕ may be seen as a weighted average of computed $\beta(u)$ values for plausible values of u. Finally, we obtain that 39.79 months for single funds, 44.52 months for funds of funds, and $\delta_\phi = 40.38$ months for all funds. Those figures are close to the ones obtained with the 2SLS model, but are slightly inferior and less dispersed. We however recover that the LATE of lockup for Hedge Funds is around 40 months.

4.4.3 Comments

We have seen in Section 4.3.1.2 that several assumptions have to be checked in order to provide a causal interpretation for the estimators using instrumental variables. Even if main assumptions cannot be tested, we provide for some of them a discussion and test empirically the others.

Our chosen instruments W (redemption frequency) seem to satisfy the conditions required for instrumental variables as they are correlated with the explicative variable (lockup D) and not with the second-step estimated residual $\hat{U}$. The relative quantities are given in Table 4.5. Consequently, assumption $A0$ seems to be plausible as the instrument is well correlated with the explanatory variable but not with the estimated residual[4].

Turning to the second part of assumption $A1$, we provide in Table 4.6 a crude estimation of the function $w \mapsto \mathbb{E}[D|W = w]$ (since W has a discrete support). We also provide card$\{w|W = w\}$ for each w in the support of W to assess the relevance of those quantities. $w \mapsto \mathbb{E}[D|W = w]$ is a non-trivial function. Even if $A'5$ cannot be tested, $w \mapsto \mathbb{E}[D|W = w]$ is however approximately monotonously increasing in w which is a positive (yet not sufficient) argument to assume monotonicity (if we leave aside the size problem of some samples that are too small to allow us to conclude).

Concerning other untestable assumptions such as $A'1$, $A'3$ and $A'4$, we can only make some comments. $A'1$ seems to be quite realistic as it assumes that the lifetime of a given fund is not influenced by the presence of a lockup on *other* funds. This sounds not odd as even if the Hedge Fund industry is a competitive universe, we may think that links in survivance or disappearance between funds may be driven by systematic crisis, frauds, or by economic/global factors, rather than by individual status or features. Then, setting a lockup on a fund is likely to have few influence on other funds. $A'3$ is again plausible as it is more credible to think that the decision of capital withdrawal for an investor relies more on the time during which her money is blocked, rather than by the minimal time to wait before

[3]For u being large, the upper bound of the inner integral has to be large itself. But beyond the empirical support of the durations of the sample, $\lambda(|Z, W)$ becomes null and this the parametric estimation of the upper bound of the integral makes no sense.

[4]Further work could be to adapt a validity test of the instruments. A Sargan test could be implemented with the 2SLS model but no test for our procedure is available at this point.

getting it back when possible. An other argument for this is that the magnitude of redemption notice durations are often days or months, whereas lockup durations are often a year or two.

4.5 Conclusion

When studying hedge funds, most of the models developed by the literature assume exogenous variables, in particular for the analysis of their lifetimes and survival. In this chapter, we provide strong evidence for the contrary, i.e. that some of them (essentially share restrictions) must be conceived as being endogenous.

Furthermore, lockup must be considered as a treatment for funds whose manager, anticipating a short lifetime, self-selects this variable. The estimated treatment effect is a Local Average Treatment Effect (LATE). It can be interpreted under mild assumptions as an Average Treatment on the Treated since the expectation of the treatment variable conditional to the instrument, covers the range of probabilities towards its lower bound.

A simple 2SLS method with redemption frequency taken as an instrument gives satisfying results, and helps in discarding any unappropriate causal statement coming from conditional hazard intensity modeling. We find then an ideal framework to illustrate the developments of Chapter 2, which allows to include censorship. We obtained more homogenous and more robust results than with classical 2SLS. The short conclusion is that a lockup extends the life of a compliant fund up to 40 months. The magnitude of this effect is clearly different from the usual lockup period (12 months) and is in favor of a “rebound” effect for funds setting a lockup, thus being in line with the approach of Hombert and Thesmar (2009) who find that funds with a stable funding can experiment a greater mean-reversion in its returns. Up to our knowledge, such a conclusion has never been reached before in the field of Hedge Funds survival analysis.

4.6 Tables and Figures

4.6.1 Descriptive Statistics

Category	**Alive, with lck.**	**Alive, without lck.**	**Dead, with lck.**	**Dead, without lck.**
All	943	5178	1544	6558
Single	780	3183	1305	5144
Fund of Funds	163	1995	239	1414

Table 4.1: Number of funds of the whole database with or without lockup.

Category	**Alive, with lck.**	**Alive, without lck.**	**Dead, with lck.**	**Dead, without lck.**
All	910	1646	433	1142
Single	766	1299	395	993
Fund of Funds	144	347	38	149

Table 4.2: Number of funds of the sample under study with or without lockup.

Period (months)	Number of funds
1	13
2	2
3	45
4	5
5	1
6	113
9	5
11	1
12	1048
14	1
18	4
24	61
25	6
27	1
30	6
36	23
>36	7

Table 4.3: Lockup periods in months.

4.6.2 2SLS estimation of LATE

Category	LATE (months)	C.I. at 95%
All funds	48.91	[36.92; 60.92]
Single funds	44.53	[31.98; 57.09]
Funds of funds	66.02	[33.23; 98.80]

Table 4.4: Estimated 2SLS LATE with corrected 95% confidence interval.

4.6.3 Nonparametric estimation of $\beta(u)$

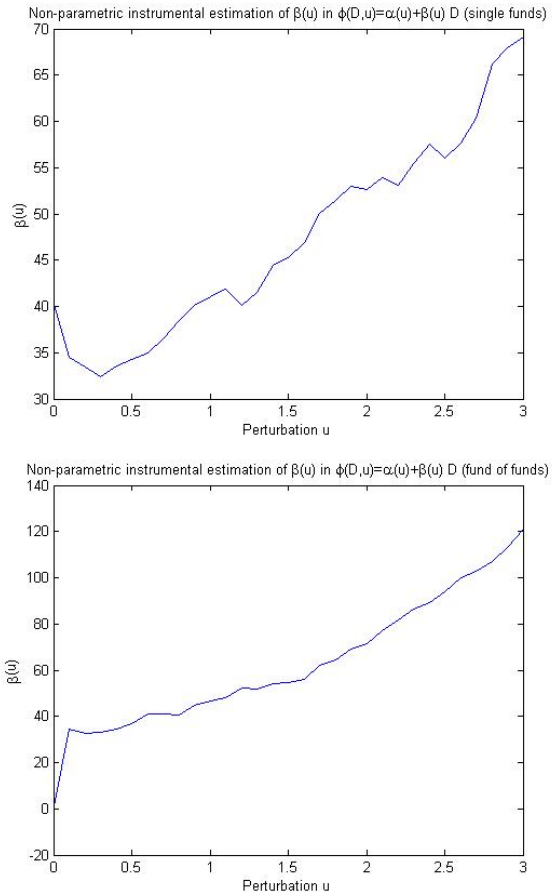

Figure 4.1: Nonparametric estimation of $\beta(u)$ (in months) for single and funds of funds.

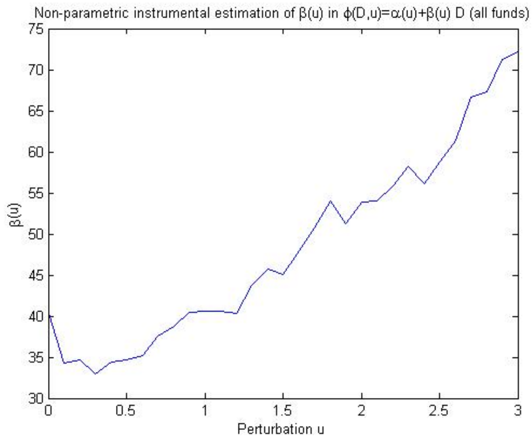

Figure 4.2: Nonparametric estimation of $\beta(u)$ (in months) for all funds.

4.6.4 Test of causal assumptions

Category	$corr(D, W)$	$corr(\hat{U}, W)$
All funds	31.92%	-2.59×10^{-15}
Single funds	32.49%	5.10×10^{-16}
Funds of funds	32.62%	2.24×10^{-16}

Table 4.5: Estimated correlations between variables and estimated residuals.

w (days)	$\text{card}\{w \mid W = w\}$	$\mathbb{E}[D \mid W = w]$
1	43	4.65 %
7	64	1.56 %
15	16	18.75 %
30	1961	15.04 %
120	1676	48.87 %
180	149	59.06 %
365	216	60.19 %
720	2	100 %
1095	4	75 %

Table 4.6: Description of W and of function $w \mapsto \mathbb{E}[D \mid W = w]$.

Chapter 5

Portfolio Allocation as an Inverse Problem

joint work with Anna Simoni[1].

Abstract

Finding appropriate methods to allocate investor's wealth in various securities is a topic of great importance, still subject of active academic research. For equity portfolio allocation, Markowitz defined a framework that using moments of the return distribution of the portfolio assets. This leads however to a numerical inconsistency when the sample covariance matrix of asset returns has to be inverted. This fact is well known and is mainly caused by the magnitude of its lowest eigenvalues. Arguments used in the empirical literature to compare allocation methods are based most-of-the-time on out-of-sample backtests rather than on mathematical arguments. Therefore, the aim of this chapter is (i) to write the allocation problem as an inverse one; (ii) to identify which allocation procedures are relevant from the inverse problem theory; (iii) to draw the connections between methods such as shrinkage and Black-Litterman and (iv) to examine whether the usually invoked Bayesian justification is relevant. Our focus is mainly theoretical but up to our knowledge, some of our findings have not been documented yet in the literature.

[1]This chapter is adapted from a preliminary work with Anna Simoni for forthcoming papers.

5.1 Introduction

5.1.1 Motivations

Portfolio allocation has always been a topic of great interest for practitioners as for academics. The objective is to find a technique that allows to get satisfying performances, while maintaining a low level of risk. This view is however too restrictive: some other criterions have to be taken into account : this sophisticates the analysis. Allocation is not confined to portfolio optimization as the necessity to find new financial indices has also arisen. The traditional financial theory has from long defined the efficient portfolio from being the portfolio weighted with the observed market-capitalizations. Consequently, main indexes on the markets are capital-weighted. Recent works preconize to use other approaches in order to build more efficient indices.

The advantage of capital-weighted indices is that they are simple, representative and easy to build, with a low turnover. However, they are (by definition) concentrated in the biggest stocks, which implies an insufficient diversification in terms of risk. Discussions on their efficiency have appeared and alternative ways to traditional index have been proposed (equally-weighted indices, fundamental indexation, etc.). We face then the usual dichotomy of portfolio allocation, balancing between performance and risk diversification.

In a nutshell, the aim of the portfolio manager (or alternatively the investor) is to rebalance positions to build optimal portfolios. But the traditional mean-variance optimization, derived from the theory of Markowitz, often leads to a numerical instability as the inversion of the sample covariance matrix makes any small change in the expected return vector dramatically amplified in the resulting vector of portfolio weights. Hence, the resulting portfolio allocation appears unstable through time. Let N denote the number of assets in the portfolio and T the number of observations. When N is larger than T, the sample covariance matrix is not even invertible. When N is less than T, but still very large, the sample covariance matrix is invertible but numerically ill-conditioned, as some of its eigenvalues are very small.

The Markowitz problem can be stated as an inverse one. It becomes *ill-posed* after that the true covariance matrix and expected return vector are replaced by their sample counterpart. A unique solution to the portfolio allocation problem still exists, but it is highly sensitive to estimation errors in the expected return vector. Recovering the optimal portfolio allocation is one of the many applied problems that require to inverse a sample covariance matrix with a dimension that is large compared to the sample size. The problem of regularizing the inverse of the sample covariance matrix is then more general than the example we consider in this chapter: at the best of our knowledge, many available regularization methods have not been applied to the Markowitz problem.

The aim of this chapter is to explicit the inverse nature of the allocation problem and also to identify which practical allocation procedures may be considered as providing a regularizing effect. Finally, we wish to draw connections between the chosen methods and examine whether they are relevant from a Bayesian perspective. Section 2 writes the portfolio allocation puzzle in mathematical terms, and introduces the problems arising with the sample covariance matrix inversion. We try to make a review of the most important allocation procedures used by practitioners, and to determine those which are relevant for our analysis. It appears particularly that two kind of procedures may be identified as having a regularizing action: the shrinkage method (Section 3) and the Black-Litterman approach (Section 4). We try to highlight the action of each method, to draw the respective mathematical links between them, and to interpret each of them from a deriving Bayesian setting. Section 5 concludes.

5.1.2 Notations

We consider a number N of risky assets, indexed by $i = 1, \ldots, n$. W_t denotes the wealth of the investor at time t, i.e. the amount of money to be invested at time t in the risk-free and in the risky assets. In the portfolio there is a quantity a_0 of risk-free asset; this asset pays the risk-free rate $r_{0,t}$ at time $t+1$ that is supposed to be known at time t. The components of the n-dimensional vector $a = (a_1, \ldots, a_n)'$ denote the quantities of risky assets in the portfolio. The price of the i-th risky asset at time t is denoted by p_t^i and the corresponding return, at time $t+1$, by r_{t+1}^i. Then $r_{t+1}^i = (p_{t+1}^i - p_t^i)/(p_t^i)$.

Moreover, we use the notation $p_t = (p_t^1, \ldots, p_t^n)'$ and $r_{t+1} = (r_{t+1}^1, \ldots, r_{t+1}^n)'$ for the vectors of prices and returns of risky assets. The capitalization of the portfolio in the i-th asset is $a_i^* = p_t^i a_i$ and the corresponding portfolio weight is denoted with $w_i = \frac{a_i^*}{\sum_{j=1}^n a_i^*}$.

The vector of price ratios $y_{t+1} = (y_{t+1}^1, \ldots, y_{t+1}^n)'$ (where $y_{t+1}^i = p_{t+1}^i / p_t^i$) is a random variable at time t with mean and variance $\tilde{\mu}$ and Ω, respectively. Hence, $\tilde{\mu}$ is an $n \times 1$ vector and Ω is an $n \times n$ matrix that are fixed moments, thus constant over time. Given that r_t is known at t, it turns out that Ω is the conditional covariance matrix of the asset excess returns. The vector of expected excess return μ is defined as $\mu = \tilde{\mu} - (1 + r_{0,t})e$, where e denotes the $n \times 1$ vector of ones. The market capitalization in asset i is given by a_M^i and the market weight of the i-th risky asset is given by w_M^i, with $w_M^i = \frac{a_M^i}{\sum_{j=1}^n a_M^j}$, $i = 1, \ldots, n$ and $w_M = (w_M^1, \ldots, w_M^n)$.

When the risk-free rate is the same for each date, we use the notation r_0. R^M and σ_M^2 are the expected return and the variance, respectively, of the market portfolio, i.e. of the chosen benchmark (which is usually a proxy for an aggregated market behavior and may be an index like S&P, CAC, DAX, FTSE, etc.). The *risk-aversion* of the market will be denoted by $A := \frac{R_B - r_0}{\sigma_B^2}$.

5.2 Portfolio allocation: instability, regularization

5.2.1 Mean-variance allocation

Starting with an universe of assets, an investor wants to select the optimal weights of each asset in a portfolio trying to control its expected performance and its expected level of risk. In this section, we briefly restate the initial modeling and its limitations.

5.2.1.1 The Markowitz framework

In the usual framework initially set up by Markowitz (1952), an investor has to maximize her expected level of utility, measured in terms of expected wealth at a given horizon, under an accepted maximum level of risk. The wealth of the investor at time $t+1$, as a function of the quantities invested in the risk-free and risky assets at the previous date, is $W_{t+1}(a_0, a) = W_t(1 + R_t) + a' diag(p_t)[y_{t+1} - (1 + r_{0,t})e]$. The budget constraint implies that $W_t(a_0, a) = a_0 + a' diag(p_t)$.

With a mean-variance optimization objective in mind, the investor's objective may resume (see Gourieroux and Jasiak (2001)) to the choice of a level of risk v and then to the constrained maximization problem:

$$\max_a \mathbb{E}_t[W_{t+1}(a_0, a)] \quad \text{such that} \quad V_t[W_{t+1}(a_0, a)] = v$$

where $\mathbb{E}_t$ and V_t denote respectively the conditional expectation and variance given the information set at t. The Lagrangian writes: $L_t = \mathbb{E}_t[W_{t+1}(a_0, a)] - \frac{\lambda}{2} V_t[W_{t+1}(a_0, a)]$. In order to solve this problem, we only need to know the first two conditional moments $\mathbb{E}_t[y_t] = \tilde{\mu}$ and $V_t[y_t] = \Omega$. The first order condition leads to: $\Omega diag(p_t) a = \frac{1}{\lambda}[\tilde{\mu} - (1 + r_{0,t})e]$ where λ may be determined through the constraint on the variance (i.e. the level of risk v). The inclusion of the budget constraint: $a_0 = W_t(a_0, a) - a^{*'} diag(p_t)$ defines the allocation in the risk-free asset. λ is then also a proxy for the risk-aversion of the investor (as when $\lambda \to \infty$, the norm of the vector of allocation in quantity vanishes). Since the prices p_t are observable, we will simplify the notation of the problem by writing $a^* := diag(p_t)a$. Alternatively, this can also be simplified by assuming that we normalize to unitary prices. Hence, the optimal portfolio allocation (in the Markowitz sense) is an allocation a^* such that:

$$\Omega a^* = \frac{1}{\lambda}\mu \tag{5.1}$$

where μ is homogenous to a vector of asset returns. Let's remark that the allocation a is in quantities. It is easy to work with weights since the efficient allocations are scaled by the risk-aversion parameter and are thus proportional[2].

5.2.1.2 The inversion problem

In practice, Ω and μ have to be estimated. Estimation itself is not a problem, since μ and Ω can be replaced by their empirical counterparts $\hat{\mu}$ and $\hat{\Omega}$. However, when N is high, $\hat{\Omega}$ may be close to singular with eigenvalues close to zero. This occurs especially when T is of the same order of N (T/N is greater than one, but close to). This fact has been well identified in the literature (see e.g. Muirhead (1987) or Pafka and Kondor (2004)) and the asymptotic theory with $N, T \to \infty$ has been set up by Ledoit and Wolf (2002).

Hence, in those cases, multiplying $\hat{\mu}$ by $\hat{\Omega}^{-1}$ leads to unstable solutions or extreme, over-weighted vectors. Merton (1980) already remarked the sensitivity of the resulting empirical portfolio to slight modifications of expected returns. Michaud (1989) underlined the problems arising when inverting the sample covariance matrix as errors in the return vector are amplified. Coordinates with abnormal weights are unfortunately those corresponding to the greatest estimation error.

This fact is a real limitation for practitioners. Proceeding to a rolling estimation of $\hat{\Omega}$ and $\hat{\mu}$ at two consecutive dates t and $t+1$, we can hand up with estimates $\hat{\mu}_t$ and $\hat{\mu}_{t+1}$ that are very close, then, even if $\hat{\Omega}_t$ and $\hat{\Omega}_{t+1}$ are also quite close, a_t^* and a_{t+1}^* may be far different due to the nearly singular nature of $\hat{\Omega}^{-1}$. This is a drawback for the investor since it implies a huge turnover of its portfolio, that is a deep rebalancing of its positions. This is not suitable since it induces practical constraints and

[2]Working with weights (i.e. $e'a = 1$) then corresponds to a particular case of risk-aversion (and level of risk as the two quantities are univocally related). In the following, we will work with portfolio weights (this being consequently equivalent.

transaction costs. Moreover the resulting portfolio track results may show an *a posteriori* variance that is too high. When the allocation weights in the portfolio are constrained to be positive, the resulting portfolios may be concentrated in a relatively small number of assets.

However, this must not be taken as an argument to fully reject the framework of Markowitz. Practicioners have set up some techniques to circumvent this limitation. One can try to reduce the dimensionality of the problem. This is the goal of the factor-model estimators that specify a given structure for $\hat{\Omega}$, in order to reduce the number of parameters to be estimated. The main contribution in this field has been made by Fan et al. (2008) who show that - surprisingly - this method works far better when using directly Ω^{-1} rather than Ω (which is of straightforward interest for the allocation problem). The obtained estimator is then always invertible and the asymptotic theory is derived, in the case where $N, T \to \infty$, with an increasing number of factors. Other techniques exist, using data on a more precise time-grid, including high-frequency methods. This increases the ratio T/N and aims at correcting the numerical instability by increasing the number of observations.

5.2.1.3 Numerical instability

We highlight in this section the instability of the empirical optimal portfolio from a spectral point of view. As expressed before, allocation a^* is very sensitive to slight changes in μ. Suppose that $\mu = (\mu_1, \ldots, \mu_n)'$ and that there exists $k \in [1; N]$ such that the k^{th}-component is modified : $\mu_k \mapsto \mu_k + d\mu_k$. What is the effect on a^*? We denote with $a^*_{(0)}$ the value, associated to the vector of estimated returns $\mu^{(0)} = \mu$, and with $a^*_{(1)}$ the new value associated to $\mu^{(1)}$ after modification on the k^{th} coordinate. We note: $\mu = \sum_{l=1}^{N} \mu_l E_l$, where E_l is the elementary $N \times 1$ matrix with 1 on its l^{th}-coordinate, 0 otherwise.

As Ω is symmetric and positive semi-definite, it can be diagonalized. Let $(\lambda_1, \ldots, \lambda_N)$ be its eigenvalues, sorted in descending order, and Φ be the matrix made of its eigenvectors $\Phi = (\phi_1, \ldots, \phi_N)$. Then, if $D = diag(\lambda_i)$, $\Omega = \Phi D \Phi'$. We have: $\Omega^{-1} = \Phi' diag(1/\lambda_1, \ldots, 1/\lambda_N)\Phi$ and

$$a^*_{(1)} - a^*_{(0)} = \Omega^{-1}(\mu^{(1)} - \mu^{(0)}) = \Omega^{-1}(d\mu_k E_{k1}).$$

Elementary calculus gives that the i^{th} component of a^* is modified such that:

$$(a^*_{(1)} - a^*_{(0)})_i = \sum_{l=1}^{N} d\mu_k \phi_{il} \phi_{kl} \lambda_l^{-1}.$$

Conversely, if the eigenvalue λ_l is estimated with $\hat{\lambda}_l$ and this estimation is modified from $\hat{\lambda}_l$ to $\hat{\lambda}_l - d\hat{\lambda}_l$, then we have:

$$(a^*_{(1)} - a^*_{(0)})_i = \sum_{k=1}^{N} \mu_k \hat{\phi}_{il} \hat{\phi}_{kl} \left(\frac{1}{\hat{\lambda}_l - d\hat{\lambda}_l} - \frac{1}{\hat{\lambda}_l} \right).$$

A first order development gives:

$$(a^*_{(1)} - a^*_{(0)})_i \simeq \sum_{k=1}^{N} \mu_k \hat{\phi}_{il} \hat{\phi}_{kl} (\hat{\lambda}_l)^{-2} d\hat{\lambda}_l.$$

A slight modification of the smallest eigenvalues is amplified of an order 2 with inversion. This shows that a modification on μ_k affects all the coordinates of a^*, and its effect is amplified by the quantity $\sum_{l=1}^{N} \phi_{il} \phi_{kl} \lambda_l^{-1}$.

5.2.2 Operator induced by the allocation problem

The Markowitz allocation problem $\Omega a^* = \mu$ is an inverse problem in finite dimension. The covariance operator Ω plays the role of the operator K:

$$K : \left\{ \begin{array}{ccc} \mathbb{R}^N & \rightarrow & \mathbb{R}^N \\ a & \mapsto & \Omega a. \end{array} \right.$$

$\mathbb{R}^N$, endowed with the usual scalar product, is a Hilbert space and Ω is a linear operator with finite dimensional range since $\mathcal{R}(\Omega) \subset \mathbb{R}^N$. Thus, Ω is a compact and continuous operator. Moreover, K is self-adjoint (for the usual scalar product in $\mathbb{R}^N$) since Ω is a symmetric matrix.

Is this problem well-posed? Technically, it depends on the characteristics and the degree of knowledge of Ω. If Ω is known, symmetric and has strictly positive eigenvalues, the answer is affirmative. The inverse exists and is continuous (since Ω can be diagonalized with strictly positive eigenvalues). When Ω is unknown it has to be replaced by a consistent estimator $\hat{\Omega}$. When T/N is close to one, $\hat{\Omega}$ becomes close to singular. Then, $\hat{\Omega}^{-1}$ is unbounded and the ill-posedness of the Markowitz inverse problem arises.

We will explore adaptations of regularization techniques such as Tikhonov regularization and spectral cut-off, respectively explained in Appendix A.2.3.3 and Appendix A.2.3.2. In numerical applications, where operator K is a matrix, Tikhonov regularization is well-known under the name of *ridge regression* and it takes the form $a_\alpha = (K'K + \alpha I)^{-1} K' \mu$, where $\alpha > 0$ is a parameter unlinked to the sample size. Ridge regression estimation was introduced by Hoerl et al. (1970) in the multiple linear regression framework. They show that when data on regressors are non-orthogonal, the OLS estimators are highly unreliable in the mean squared error sense and sensitive to small changes in data. The ridge estimator, although it is biased, allows to reduce the mean squared error and the length of the regression vector. The same results are obtained through a Tikhonov regularization in infinite dimension. The general form for the ridge regression substitutes αI with a diagonal matrix where the ridge parameter changes from a dimension to another one. The ridge regression technique has been extended to the general problem of matrix inversion by Vinod (1976). So, in the portfolio allocation the ridge technique gives: $a_* = (\Omega + \alpha I)^{-1} \mu$, i.e. the same as Tikhonov regularization. Hoerl et al. (1970) suggested to select the value for α through the so-called “ridge trace” which traces the values of the elements of the ridge estimator against α. It is possible to give a Bayesian interpretation to the ridge regression. The ridge estimate can be seen as the posterior mean based on giving the parameter to estimate a prior normal distribution with zero mean and variance-covariance matrix Ω. Indeed, Bayesian estimation automatically solves the problem of ill-conditioning of the regressor matrix through the incorporation in the estimator of the prior information (see the work of Zellner (1971)).

5.2.3 Overview of allocation techniques

We present here an overview of the allocation techniques used in practice, and try to distinguish in which way they are relevant for our analysis.

5.2.3.1 Redefinitions of the allocation problem

For **Equally-Weighted** portfolios, all the components of the N underlying assets of the portfolio are equal and assigned to be weighted as $1/N$, regardless of their statistical properties. This technique is effectively used in practice (see e.g. Windcliff and Boyle (2004)). Empirically, DeMiguel et al. (2009) working on several datasets (mainly equity data) found that naive equally-weighted portfolios are even more interesting in out-of-sample when compared to allocation strategies where parameters have to be estimated. The resulting portfolios are diversified in proportions and focus on performance, which is much better in out-of-sample (compared to GMV portfolios, see below). A financial justification (the portfolio is located on the efficient frontier) is obtained in the particular case where all assets have identical statistical moments (mean and variance) with identical correlations. However, the diversification argument fails in case of heterogenous statistical properties.

The **Global Minimum Variance** (henceforth GMV, or simply **Minimum-Variance** portfolios) only aims at minimizing the variance of the resulting portfolio, that is to say to minimize $a'\Omega a$. This portfolio has a classical financial interpretation, and is on the efficient frontier. It is simple to compute as it does not require to invert the covariance matrix. It provides out-of-sample satisfying Sharpe ratios, even if the induced risk diversification is not sufficient (portfolios are too concentrated and potentially exposed to high drawdowns).

The **Most Diversified Portfolio** (henceforth MDP) developed by Choueifaty and Coignard (2008) tries to reach maximum diversification by maximizing the diversification ratio which is the average of the volatilities of the assets divided by the portfolio volatility. This allocation technique is mostly considered by practitioners. Closely related, the **Maximum Sharpe Ratio** (MSR, see Martellini (2009)), aim at maximizing the Sharpe ratio of the resulting portfolio, that is for a vector of weights w the quantity $(\mu w - r_0)/\sqrt{(}w'\Omega w)$. As shown in Goldfarb and Iyengar (2003), MSR portfolios can be interpreted as specific minimal-variance portfolios with modified constraints on the return.

We remark also that the Markowitz allocation uses the two first moments of the return distribution. However, as portfolio allocation is also represented by an expected utility maximization problem, some approaches redefine the problem with a closer expression for the expected utility. The latter is represented through a Taylor series expansion, including higher moments of the distribution. This allows to deal with high-dimensional portfolios and with highly non-Gaussian return distribution (see for instance among others Jondeau and Rockinger (2006), Chabi-Yo et al. (2008) or Martellini (2009)).

Equally-weighted Risk Contributions (henceforth ERC) portfolios may be viewed as minimum-variance portfolios where a constraint is added on the risk-diversification, as all assets are weighted to have an equal contribution in risk. Those allocation procedures are simple to build and appear to be a good compromise empirically between $1/N$ and minimum-variance allocation (see Qian (2005), or Demey et al. (2010)). The philosophy of ERC allocation is to weight the assets such as each of them has the same contribution to the overall variance. For a set $w = (w_1, ..., w_N)$ of weights of N risky assets with covariance matrix Ω. The variance of the portfolio $\sigma(w)$ is $\sigma(w) = \sqrt{w'\Omega w}$. The standard-deviation is decomposed as follows:

$$\sigma(w) = \sum_{i=1}^{N} w_i \partial_{w_i} \sigma(w) = \sum_{i=1}^{N} w_i \frac{w_i \Omega_{ii} + \sum_{j \neq i} w_j \Omega_{ij}}{\sigma(w)}.$$

The resulting vectors of weights w_0 are such that:

$$\sum_{i=1}^{N} w_i = 1 \quad , \quad w_i \in [0;1] \quad , \quad w_i(\Omega w)_i = w_j(\Omega w)_j \quad , \quad \forall i \neq j.$$

At this point, those allocations procedures do not provide a mathematical treatment of the theoretical optimal Markowitz portfolio (optimal from the risk-performance perspective), but rather choose to focus on sparseness, or on nominal/risk diversification for instance. Obviously, they *do not* constitute regularization techniques as no problem is even stated for the $1/N$ allocation and MDP, MSR techniques can be viewed as reformulation of the Markowitz problem. However, those methods do not get rid off the instability problem of the variance matrix. Consequently, using one of these techniques does not prevent from using a preliminary cleaning algorithm for the covariance matrix used in the procedure.

5.2.3.2 The relevant techniques

We will study in this work two specific allocation methods from the regularization perspective: the shrinkage technique and the Black-Litterman approach. The shrinkage (or reduction) technique has been applied to control the form of the variance-covariance matrix with the aim of obtaining more stable portfolios. The idea of shrinkage procedure, reviewed in Section 5.3.1, is to shrink the sample covariance matrix towards a matrix with a better structure, called target matrix, so that the resulting estimator has a stable inverse. Alternatively, Black and Litterman (1992) developed a setting where investors introduce in the allocation procedure their views on the market in order to stabilize their portfolio. This method, claimed to be inspired by the Bayesian framework and is reviewed in Section 5.4.2.1.

5.2.3.3 Other methods

To be complete, we present other allocation techniques. Their link with regularization are not tackled here but this could be a perspective for future work. Jagannathan and Ma (2003) show that imposing a **no short-sell** constraint allows to decrease the out-of-sample risk, even if the constraints are wrong or unjustified (stocks with high covariances tend to receive huge negative weights). Broadie et al. (2009) add a l_1 penalty term to the optimization program up to a parameter to be calibrated. This may be considered as a version of lasso regression. In addition to the regularizing effect, this allows to take into account transaction costs, to control short-selling, and finally to obtain preferentially sparse portfolios (i.e. with few non-zero positions). As expressed before, the constraint on short-selling ensures the stabilization of the portfolio. However, this paper also underlines that such a constraint naturally implies the construction of sparse portfolios. Empirically, this method works well as the obtained portfolio are more robust, less affected by estimation error and noisy information. Finally, an other approach is to define a more structured model for the covariance matrix, as e.g. in Specht and Gohout (2003). A principal component analysis is made on the return series and the corresponding principal components are studied as univariate series through a given volatility model (GARCH for instance).

5.3 Linear methods

5.3.1 Shrinkage

5.3.1.1 Setting

Shrinkage (also called *reduction*) aims at correcting the sample variance-covariance matrix when this latter is ill-conditioned. A simple version of the shrinkage estimator of the covariance matrix is built as a convex linear combination of the empirical estimator with some suitably chosen target matrix that has a better structure. Initial work on shrinkage estimators for the mean of a multivariate normal distribution has been done by Stein (1986) who shows that if we are willing to give up a little in terms of bias, we could do better in terms of mean squared errors. Mathematically, the intensity of the reduction is a parameter[3] $\alpha \in [0;1]$ and represents the relative weight of a target matrix in the new covariance matrix. If $\hat{\Omega}$ is the sample variance-covariance matrix, and T_a denote the target matrix, then the new matrix Ω_s is computed as:

$$\Omega_s = \alpha T_a + (1-\alpha)\hat{\Omega}.$$

The mixing parameter α may be chosen in order to provide an arbitrage between the estimation error coming from $\hat{\Omega}$ and the specification error of T_a. Ledoit and Wolf (2004) gives a review of this method, including its application in practice. In their framework, α is chosen in a so-called "optimal way". The criterion is to maximize the expected accuracy of the “shrunken” estimator, and to minimize the distance, computed using the Frobenius norm, between the covariance matrix and its reduced estimator. When α is estimated, such procedures must not involve the inverse of the covariance matrix for obvious reasons. Two common choices for T_a are a constant correlation matrix or a matrix derived from a market model.

In the constant correlation matrix model, the terms of the matrix T_a only depends on the variance of the assets. Its diagonal is made of the individual variances v_i of the asset returns r_t^i, $i = 1, \ldots, N$. The non-diagonal terms T_{ij} are equal to $c\sqrt{v_i v_j}$ where c is computed from an overall quantity estimated upon returns. In the market model, one first writes from the CAPM the following expression for an asset i of returns r_t^i: $r_t^i = \alpha_i + \beta_i r_t^b + \eta_t^i$ where the r_t^b are the return at time t of a market benchmark. η_t^i is supposed to be an i.i.d. noise with variance σ_i. Thus, the target matrix T_a is set equal to $T_a = v_0^2 \beta\beta' + \Sigma$ where v_0 is the variance of r^b, $\beta = (\beta_i)_{i=1:N}$ and $\Sigma = diag(\sigma_1, \ldots, \sigma_N)$. Ledoit and Wolf (2003) precise that when $T_a = \mu I$, μ must be equal to the trace of the sample covariance matrix. This ensures that the initial allocation problem will be replaced by a problem with the same overall level of risk. Finally, shrinkage technique can be extended to other asset classes than equity and stocks.

5.3.1.2 Effect on the spectrum

Intuition

The effect of shrinkage on the spectrum of the sample matrix is more intuitive when $T_a = \mu I$, μ being the trace of the sample matrix: $\Omega_s = \alpha\mu I + (1-\alpha)\Omega$. Let λ_k be an eigenvalue of Ω and ϕ_k the associated eigenvector. Then ϕ_k is also an eigenvector of Ω_s for the eigenvalue $\alpha\mu + (1-\alpha)\lambda_k$ since:

[3] Despite its notation, this parameter α is not the same as the parameter which will be introduced below in the B-L approach.

$$
\begin{aligned}
\Omega_s \phi_k &= \alpha\mu\phi_k + (1-\alpha)\Omega\phi_k \\
&= \alpha\mu\phi_k + (1-\alpha)\lambda_k\phi_k \\
&= (\alpha\mu + (1-\alpha)\lambda_k)\phi_k.
\end{aligned}
$$

With α between 0 and 1, when λ_k is greater than μ, $\alpha\mu + (1-\alpha)\lambda_k$ is greater than μ but lower than λ_k. Conversely, when λ_k is lower than μ, this quantity is inferior to μ but greater than λ_k. The effect of shrinkage is then to *shrink* the range of the spectrum and to concentrate it around the trace. The major effect is on the lowest eigenvalues that are shifted upwards. This is a simple intuition that confirms the argument of Ledoit and Wolf (2003) explaining that when the target matrix is μI, the shrinkage reduces the dispersion of the sample eigenvalues towards their mean. α is then interpreted as "*a normalized measure of the error of the sample covariance matrix*". The paper also shows why the eigenvalues of the sample covariance matrix are more dispersed than the eigenvalues of the true matrix, and why it is preferable to use estimators with eigenvalues that are less dispersed that the eigenvalues of the true covariance matrix. This dispersion is also function of T/N.

In the general case, $\hat{\Omega}$ is replaced by $\Omega_s = \alpha T_a + (1-\alpha)\hat{\Omega}$. Let $\hat{\lambda}_k$ be the eigenvalues of $\hat{\Omega}$ and $\hat{\phi}_k$ be the corresponding eigenvector. Without a target matrix, the solution a_* of $\hat{\Omega}a_* = \mu$ is then obtained as:

$$
a_* = \sum_{k=1}^{N} \frac{1}{\hat{\lambda}_k} < \mu, \hat{\phi}_k > \hat{\phi}_k.
$$

With $\Omega_s = \alpha I + (1-\alpha)\hat{\Omega}$:

$$
a_*^s = \sum_{k=1}^{N} \frac{1}{\alpha + (1-\alpha)\hat{\lambda}_k} < \mu, \hat{\phi}_k > \hat{\phi}_k
$$

since Ω_s has eigenvalues equal to $\hat{\lambda}_k(1-\alpha) + \alpha$ with the same eigenvectors than $\hat{\Omega}$. With notations of Appendix A.2.3.1 we get a regularization scheme $q_I(\alpha, \hat{\lambda})$:

$$
q_I(\alpha, \hat{\lambda}) = \frac{\hat{\lambda}}{\alpha + (1-\alpha)\hat{\lambda}}
$$

which is a first order regularization scheme.

General case

We turn now to the general case. We consider a general target matrix T_a with eigenvalues l_k and corresponding eigenvectors φ_k and we state the following proposition (the proof is available in Section 5.6.1):

Proposition 5.3.1. Let T_a be a nonsingular square matrix and $\Omega_s = \alpha T_a + (1-\alpha)\hat{\Omega}$ be the shrinkage estimator of Ω. The approximate solution $a_*^s = \Omega_s^{-1}\hat{\mu}$ corresponds to a regularized solution obtained through a first order regularization scheme. Moreover, if $(\hat{\lambda}_j, \hat{\phi}_j)_j$ denote the eigensystem associated to $\hat{\Omega}$, then Ω_s has eigenvectors $(\hat{\phi}_j)_j$ and eigenvalues $((1-\alpha)\hat{\lambda}_j + \sum_{u=1}^N \alpha l_u < \varphi_u, \phi_j >^2)_j$ and a_*^s can be rewritten as:

$$a_*^s = \sum_{j=1}^N \frac{1}{(1-\alpha)\hat{\lambda}_j + \sum_{u=1}^N \alpha l_u < \varphi_u, \hat{\phi}_j >^2} < \hat{\mu}, \hat{\phi}_j > \hat{\phi}_j.$$

□

In particular, since $\sum_{u=1}^N l_u < \varphi_u, \hat{\phi}_j >^2 = ||T_a^{\frac{1}{2}} \hat{\phi}_j||^2$ when $Ker(\hat{\Omega}) = \{0\}$, we see that a_*^s takes the form in (A.11) so that the shrinkage technique corresponds to a generalized Tikhonov regularization scheme. The asymptotic theory is here quite particular but interesting. This aspect has been treated in Ledoit and Wolf (2002). It is implicit that when $N \to \infty$, we must have also $T \to \infty$ in order to have $T \geq N$. Intuitively, the degree of ill-posedness is decreasing in T (with N fixed) and increasing in N (with T fixed, $T \geq N$). This point has been underlined by Ledoit and Wolf (2003) : they found that the shrinkage intensity α has to increase with the estimation error made on the sample covariance matrix, and to decrease with the misspecification error due to the target matrix T_a. The estimation error disappears as T increases while the latter does not. Hence, the misspecification error is negligible when T is small, but its influence matters as T becomes large. This is coherent if we interpret α as a true regularization parameter. Ledoit and Wolf (2003) derive the behavior of the "optimal" α (optimal as defined in paragraph 5.3.1.1) that must be of order $O(1/T)$, with a specified constant of proportionality that can be estimated.

From a regularization perspective, the point of interest is to know whether $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$. In usual regularization schemes, α is increasing in the measurement error, given in our case by the estimation error made on $\hat{\Omega}$. In portfolio allocation, the sample has two dimensions: N and T, one linked to the misspecification error and the other one to the measurement error. Therefore, α must be decreasing in T. This is satisfied if α depends on T, but we have also to control that T remains greater than N as $N \to \infty$ (we can still consider α as a $O(1/T)$). This asymptotic framework is known as *general asymptotics* (Ledoit and Wolf (2003)). Moreover, we must keep in mind that for a given $j \in [1; N]$, $q(\alpha, \lambda_j)$ depends on the term $\sum_{u=1}^N l_u < \varphi_u, \phi_j >^2$, but all the terms change with N. A very simple hypothesis can help to solve this situation.

Proposition 5.3.2. If we assume that the spectrum of the **target** matrice is uniformly bounded from **above**(uniformly in N and T) by a constant M_0 (i.e. for every N, and sample of size T, $\forall u \in [1; N]$, $|l_u| \leq M_0$), then $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$ and shrinkage is a regularization technique.

□

The assumption that the spectrum of the **covariance** matrix is bounded from **below** may often be found in the literature. For obvious reasons, and as compact operators appear commonly in ill-posed inverse problems, this is typically a class of assumption that we **cannot** make as we aim at studying the opposite. With assumptions of propositions 5.3.1 and 5.3.2 we have that:

$$0 \leq \sum_{u=1}^N l_u < \varphi_u, \hat{\phi}_j >^2 \leq M_0 \sum_{u=1}^N < \varphi_u, \hat{\phi}_j >^2 = M_0$$

since $\sum_{u=1}^{N} < \varphi_u, \phi_j >^2 = ||\phi_j||^2 = 1$ as (ϕ_j) is an orthonormal family for each N. Then, we can control for the convergence of $q(.,.)$ independently of the order of the eigenvalue, as a general function of λ (and no more λ_j), as $\lambda \to q(\alpha, \lambda) = \lambda/((1-\alpha)\lambda + \alpha M)$ so that $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$.

Remark 1. Ledoit and Wolf (2003) attributes the fact that the sample eigenvalues are more dispersed than the true ones to the presence of error also in the sample eigenvectors, and that shrinkage is a correction for this effect. The former calculus provides a beginning of explanation to see how the eigenvectors are related to eigenvalues of the shrunken estimator.

Remark 2. Jagannathan and Ma (2003) impose a no short-sell constraint even if the constraints are wrong or unjustified. As underlined by Ledoit and Wolf (2004), this can be somewhat related to a shrinkage approach. Considering such an approach as a regularization under convex constraints (see Engl et al. (1996)) could be a question of future research.

5.3.2 Spectral cut-off

Techniques that explicitly manipulates the eigenvalues of the sample covariance matrix are empirically less used but may be found in the literature. We see how it adapts to the problem of portfolio allocation, and we explore how it can be linked to existing techniques.

5.3.2.1 Eigensystem interpretation

Eigenvectors ϕ_i are homogenous to portfolios. Matrix Φ, whose columns are the eigenvectors ϕ_i, is a basis of orthonormal portfolios resuming best the linear information available through the observation of asset returns. Eigenvalues are the specific variances of those portfolios. This framework is very close to Principal Component Analysis (PCA) where the covariance matrix of the variable of interest is diagonalized. The first eigenvectors (corresponding to the highest eigenvalues) explain best the information, in the sense that they have the greatest contribution to the overall variance. Hence, the last eigenvectors are the less informative. In the degenerate case, if $\Omega\phi_k = 0$, this means in financial terms that portfolio k has a variance that is close to zero and may appear as risk-free. Then, if the expected return is positive, even slightly, the allocation process understands it as an arbitrage opportunity and overweights the allocation in this portfolio, artificially considered by the algorithm as a "good opportunity". In terms of information, this means that ϕ_k is collinear to the other ones and then it can be expressed as a linear combination of other eigenvectors. Thus, eigenvectors corresponding to the smallest eigenvalues are portfolios that are redundant in term of information. Imposing constraints on the portfolio weights (like for instance, nonnegativity constraints and upper bounds) will prevent from getting resulting portfolios with arbitrarily huge weights.

5.3.2.2 Selection of the threshold

Intuition

Following this analysis, there is an intuitive procedure to ensure the stability of mean-variance portfolios, that naturally appear as deriving from the spectral cut-off. The aim is then to stay in the initial framework of Markowitz and to tackle the true source of portfolio instability. First, one has diagonalizes $\hat{\Omega}$ and get the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$ with corresponding eigenvectors $\hat{\phi}_1, \ldots, \hat{\phi}_N$. The next step is to choose an order k (with an auxiliary criterion) and to keep only the information relative to the k greatest eigenvalues.

A possible approach is to simply drop this information and to restart the usual Markowitz allocation with the set of assets $X = (X_1, \ldots, X_N)$ replaced by $\hat{\phi}_1.X, \ldots, \hat{\phi}_k.X$. This does not reduce the dimensionality of the problem, but reduces the sources of instability by working on a set of sufficiently informative and linearly independent portfolios. This method has at least two drawbacks. First, the information coming from the eigenvectors corresponding to the lowest eigenvalues is discarded. Second, as the overall level of variance of the allocation problem is equal to the trace of the covariance matrix, dropping some eigenvalues leads to a different problem, as it defines a new level of risk.

In, Potters et al. (2005), eigenvalues over a given threshold are kept unchanged; the smallest being modified such that they are set to a common constant value. This value is chosen in order to preserve the trace of the initial matrix. This allows to work with the same level of overall risk for the allocation. We must add that such methods are already implicitly behind factor models, as they seek to sum up the information between assets through market factors that appear to be sufficient[4].

Random Matrix Theory

The Random Matrix Theory is gaining in popularity in finance, as this theory may provide insights on the singular value decomposition of the sample correlation matrix. A detailed description would be far beyond the scope this thesis but a see for instance Mehta (2004) for a complete review. The question becomes whether the information coming from this matrix, and conveyed through its eigenvalues, can be considered significantly different from pure noise. Eigenvalues and eigenvectors that are significantly different from the "random" case are those which carry the most valuable information. Comparison of the empirical spectrum of the covariance matrix with the theoretical distribution of the eigenvalues of a random matrix helps in practice to calibrate thresholds for spectral cut-off.

Let's suppose that the ratio $r = T/N$ is fixed, greater than 1. When we assume that the coefficients of the matrix of assets price changes are independent, identically distributed, the distribution of the eigenvalues of the resulting correlation matrix is known when $N, T \to \infty$ and r stays fixed. This density takes sense only in an asymptotic framework and is known thanks to the work of Marcenko and Pastur (1967). This can be a useful guideline to evaluate the available information carried by a sample correlation matrix and a rather large number of studies have already been made on financial time-series (see also among other works Malevergne and Sornette (2004), Potters et al. (2005), Bouchaud et al. (2007)).

[4]As Malevergne and Sornette (2004) expresses it, the difficult part is a reasonable and robust choice of the market portfolio. The authors show that the existence and the appearance of factors is in fact the "*result from a collective effect of the assets*". Factors are then both a cause and a consequence of asset correlation. Consequently, the fact of being "data driven" allows to find the most informative portfolios and to allocate in a traditional fashion, dropping redundant information that generates instability

5.4 Bayesian methods

5.4.1 Introduction

5.4.1.1 Motivation

The source of ill-posedness lies in the spectral configuration of the variance matrix, but the manifestation of this instability occurs when modifications are made on the vector of returns. Replacing a sample mean of returns by a "smoother" version is a quite natural idea. Bayesian methods can evidently help in achieving this objective, first by providing alternatives to the common sample mean when beliefs on the vector of returns are smoothly updated. But the Bayesian approach can be extended to include the Markowitz optimization problem into a more global framework. As an illustration, the next paragraph states that one may also provide to the shrinkage estimator a Bayesian interpretation.

5.4.1.2 Empirical Bayes interpretation of the shrinkage estimator

The fact that a shrinkage estimator can be motivated by a Bayesian argument is well known (see Copas (1983)). In the literature, the form of several shrinkage estimators has been motivated through an empirical Bayes modeling as, for instance, the James-Stein estimator that corresponds to a prior mean of the parameter of interest set close to 0. Here, we propose to interpret the shrinkage estimator of the covariance matrix as an Empirical Bayes estimator. Our interpretation is motivated by Haff (1980).

Let $\hat{T}_a$ be a consistent estimator of the target matrix T_a. Let assume that the sampling and prior distributions are specified as

$$\hat{\Omega}|\Omega \sim \mathcal{W}_{N\times N}\Big(\frac{1}{1-\alpha}\Omega, n+1\Big) \quad \text{and} \quad \Omega^{-1} \sim \mathcal{W}_{N\times N}\Big(\frac{1}{\alpha}\hat{T}_a^{-1}, 2n+3\Big),$$

where $\mathcal{W}_{N\times N}$ denotes a Wishart distribution on the cone of the $N \times N$ matrices. The matrix $\hat{T}_a$ must be positive definite and it is replaced by its true value T_a if it is known. By writing down the kernels of the two Wishart distributions and by using the properties of the trace operator we have

$$\begin{aligned} pdf(\Omega|\hat{\Omega}) &\propto |\Omega|^{-\frac{N+1}{2}} \exp\Big\{-\frac{1}{2}tr((1-\alpha)\Omega^{-1}\hat{\Omega})\Big\} \times \\ &\quad |\Omega|^{-\frac{N+2}{2}} \exp\Big\{-\frac{1}{2}tr(\alpha\hat{T}_a\Omega^{-1})\Big\} \\ &\propto |\Omega|^{-\frac{(N+2)+N+1}{2}} \exp\Big\{-\frac{1}{2}tr([(1-\alpha)\hat{\Omega}+\alpha\hat{T}_a]\Omega^{-1})\Big\}. \end{aligned}$$

Hence, the posterior distribution of Ω is an *Inverse Wishart*

$$\Omega|\hat{\Omega} \sim \mathcal{W}^{-1}_{N\times N}((\alpha\hat{T}_a + (1-\alpha)\hat{\Omega}), N+2)$$

and it has mean $\mathbb{E}(\Omega|\hat{\Omega}) = \alpha\hat{T}_a + (1-\alpha)\hat{\Omega}$. We have proven that the shrinkage estimator Ω_s is the posterior mean of Ω given $\hat{\Omega}$.

In general, the parameter α is unknown, but its optimal value is usually specified in terms of quantities that can be estimated. We denote with $\hat{\alpha}$ the estimated optimal shrinkage intensity and we substitute this value in the sampling and prior distribution. Hence, the shrinkage estimator Ω_s is called *Empirical Bayes* since the prior distribution depends on the sampling information through $\hat{\alpha}$.

5.4.2 Black-Litterman

The Black-Litterman approach (B-L hereafter, Black and Litterman (1992) or Lee (2000)) is a technique proposed to integrate in the allocation process the forecasts of the investor, while solving the instability. Due to its simplicity this technique is now commonly used, and raise the attention of portfolio managers are interested in incorporating personal views in the model. Even if the model is not usually presented as a regularization technique, this is one of the fundamental goal that this model seeks in fact to achieve.

B-L approach primary focuses on μ. We will see that however, there is also an effective action on the covariance matrix. It consists in computing portfolio weights without having to estimate or to completely specify $\hat{\mu}$. The idea is that an investor can introduce some views that are subjective forecasts of deviations from market equilibrium. Hence, an investor invests first in the market portfolio; then by adding some views, he deviates in the direction of those views. If an asset is not concerned by any view, its corresponding allocation will be directly related to the equilibrium portfolio. Several extensions of the B-L model exist. All these are beyond the scope of this chapter as we initially wish to explore the regularization properties of the technique. However, we underline that the B-L setting extends to other markets and classes of assets than equity as returns may be replaced by risk factors. Other improvements may exist, where volatilities and correlations may also be stressed; even non-linear and very general views can be included. See Meucci (2008) for a general approach and review.

5.4.2.1 Setting

The main feature of the B-L model is that it allows to mix investors specific forecasts with equilibrium views. Those views may be general, derived from the CAPM theory, or deliberately chosen by the fund manager (for instance, according to an historical analysis). The CAPM expresses each stock return as the sum of the risk-free return r_0, an excess return with respect to this risk-free return, which is proportional to the benchmark excess return, and an idiosyncratic return with zero mean. Formally, for an asset i, the *expected return* is:

$$\mathbb{E}[r_t^i] = r_0 + \beta_i(R^M - r_0)$$

where β_i is the sensibility parameter, *i.e.* the correlation of asset i with the market portfolio. The CAPM theory derives the following expression for Π, the implied asset excess returns vector:

$$\Pi = A\hat{\Omega}w_M$$

that is computable since A and w_M can be observed (see Section 5.1.2) and $\hat{\Omega}$ is estimated (and need not to be inverted). The market capitalization weights w_M emulates the weights of the market-equilibrium portfolio. It is straightforward that w_M can be replaced by any w_0 by the manager if she thinks that her new w_0 is a more accurate proxy for the equilibrium weights of the market he is participating in. Expression using capitalizations is the most often used but may also be modified, as the most important is that Π is, in a way, a kernel of stability for the allocation algorithm.

The method starts by supposing that the assets extra return[5] $R := r_{t+1} - r_{0,t}e$ to be normally distributed as $R \sim N(\mu, \Omega)$ where μ cannot be observed and is itself random. The model specifies two distributions for μ, one suggested by the market equilibrium and another one suggested by investors' forecasts. The model aims at specifying in the same time a distribution for Π, conditional on μ. Within a Bayesian framework the posterior distribution of μ conditional on Π may be derived. Rather than using the sample mean of μ as an estimator, this latter is replaced by the mean (denoted as $\overline{ER}$ in the following) of the posterior distribution of μ conditional on Π. Of course, the expression of $\overline{ER}$ will incorporate $\hat{\Omega}$ in order to tackle the ill-posedness.

The expression of the forecasts specifies the distribution of l linear combinations of the elements of μ, $P\mu \sim N(V, F)$, where we adopt the following notations:

- P is a $l \times N$-dimensional matrix selecting the assets concerned with the views (or forecasts);
- F is a $l \times l$-dimensional diagonal covariance matrix of the errors on these views;
- V is a l-dimensional vector of the absolute expression of those views.

Π is assumed to be normally distributed as $N(\mu, \alpha\hat{\Omega})$. The scalar parameter α has to be calibrated (empirically, α is assumed to lie between 0 and 1, see Lee (2000)). Under those assumptions, μ is normal with mean $\overline{ER}$ and variance VR equal to:

$$\begin{aligned} \overline{ER} &= [(\alpha\hat{\Omega})^{-1} + P'F^{-1}P]^{-1}[(\alpha\hat{\Omega})^{-1}\Pi + P'F^{-1}V] \\ VR &= [(\alpha\hat{\Omega})^{-1} + P'F^{-1}P]^{-1}. \end{aligned} \tag{5.2}$$

$\overline{ER}$ may be expressed in the following form, which is more intuitive from a financial point of view since it is expressed as the deviation from Π (the optimal portfolio being a deviation from the market portfolio):

$$\overline{ER} = \Pi + (\alpha\hat{\Omega})P'(F + \alpha P\hat{\Omega}P')^{-1}(V - P\Pi).$$

If no views are specified, we recover that:

$$\overline{ER} = [(\alpha\hat{\Omega})^{-1}]^{-1}[(\alpha\hat{\Omega})^{-1}\Pi] = \Pi.$$

Finally the allocation uses $\overline{ER}$ instead of $\hat{\mu}$:

$$a^* = (A\hat{\Omega})^{-1}(\bar{E}R)$$

which rewrites :

$$a^* = w_M + \frac{\alpha}{A}P'(\alpha P\hat{\Omega}P' + F)^{-1}(V - P\Pi).$$

The model is Bayesian in several ways. Section 5.4.2.2 gives the mathematical background to understand the model from a Bayesian perspective. Practically, the manager can update at each date its equilibrium vector Π by replacing w_M by any w reflecting current or new information. Then, the allocation can be updated at each date through Π. We see with the latter formula that if both w_M

[5] We do not index R with the time index because, given $r_{0,t}$, the extra return R is a random variable with the same distribution at each time t.

and the views are unchanged at each allocation, the weights will not evolve. The coherence of this approach is to model the distribution of Π conditional on μ since it means that conditional on the same μ, Π is on average equal to μ. If all the agents believe in the views μ, the equilibrium returns will tend to Π itself centered on μ, which confirms the equilibrium justification of Π.

5.4.2.2 A link with Bayesian theory?

The B-L approach looks similar to a Bayesian setting since a prior distribution is specified on μ, which is supposed to be random, and this prior is then updated to form new opinions. This is the commonly accepted view among practitioners. However, in order to have a Bayesian interpretation we cannot use the conditional distribution of $R|\mu$ as sampling mechanism, even if it could seem the more natural candidate. Indeed, as developed in Satchell and Scowcroft (2000), we have to invert the distribution of $\mu|\Pi$ in order to obtain $\Pi|\mu \sim \mathcal{N}(\mu, \alpha\hat{\Omega})$. This distribution will play the role of the sampling mechanism. Hence, Π must be meant as being the implied returns from the equilibrium model, depending on our data. To see this, we suppose that we observe a sample of extra-returns on the N assets and to average them so that we obtain

$$\begin{aligned}\bar{R} &= \frac{1}{T}\sum_{t=1}^{\tau}(r_t - R_{t-1}e) \\ \bar{R}|\mu &\sim \mathcal{N}\Big(\mu, \frac{1}{T}\hat{\Omega}\Big).\end{aligned}$$

Scalar $1/T$ plays the role of α in the original B-L model and $\bar{R}$ is also an estimator of Π. Then, the views give a prior distribution on l linear combinations of the components of μ that is mixed with the information coming from the equilibrium model (*i.e.* in the sampling model). In practice the sampling mechanism must be understood in the sense that, if all the investors hold the same view on μ and invest in a CAPM-type world, then Π represents the equilibrium returns conditional on these common views. We state the following proposition (proof is available in Section 5.6.2):

Proposition 5.4.1. Let $\bar{R}|\mu \sim \mathcal{N}(\mu, \frac{1}{T}\hat{\Omega})$ be the data equilibrium return, given the forecast μ held by the investors. Let $P\mu \sim \mathcal{N}(V, F)$ be the prior beliefs on l linear combinations of expected excess returns, with F a diagonal matrix. Then, the posterior forecast, given the equilibrium information is

$$\mu|\bar{R} \sim \mathcal{N}(\overline{ER}, VR),$$

where $\overline{ER}$ and VR are as given in (5.2).

□

As expressed in Satchell and Scowcroft (2000), "*the interpretation of what is prior and what is sample information may differ from B-L*". Anyway, the Bayesian interpretation is natural in the updating procedure described in Equation 5.4.2.1. In practice, the prior-to-posterior transformation is not the same computation written down by Black and Litterman to obtain the modified expected excess returns $\overline{ER}$, but it gives the same result. Hence, it can be seen as a way to interpret the B-L model from a Bayesian perspective.

5.4.2.3 Effect on the spectrum

The main effort of the model is to focus on the vector of expected returns. Even if the initial source of instability is the inversion of $\hat{\Omega}$ and not the estimation of μ, we will show in this paragraph that B-L model provides in fact a correction on this matrix. Up to our knowledge, such a reading of this correction is not provided in the literature. The expression of a^* is:

$$(A\hat{\Omega})^{-1}[(\alpha\hat{\Omega})^{-1} + P'F^{-1}P]^{-1}[(\alpha\hat{\Omega})^{-1}\Pi + P'F^{-1}V].$$

The multiplying matrix is:

$$M = (A\hat{\Omega})^{-1}[(\alpha\hat{\Omega})^{-1} + P'F^{-1}P]^{-1}.$$

We will explicit its eigenvalues. For this we may study M^{-1}. M is proportional to $M_1(M_1 + M_2)^{-1}$ where $M_1 = (\alpha\hat{\Omega})^{-1}$ and $M_2 = P'F^{-1}P$. Then M^{-1} is proportional, up to a parameter A/α, to the matrix:

$$[M_1(M_1 + M_2)^{-1}]^{-1} = (M_1 + M_2)M_1^{-1} = I_d + M_2M_1^{-1}$$

where $M_2M_1^{-1} = \alpha P'F^{-1}P\hat{\Omega}$. Thus the eigenvalues of M^{-1} have for lower bound :

$$A/\alpha(1 + \inf sp(\alpha P'F^{-1}P\hat{\Omega}))$$

where $sp(.)$ stands for the set of eigenvalues of a matrix. Then, even if the eigenvalues of $P'F^{-1}P\hat{\Omega}$ tends to zero, the set of eigenvalues of M^{-1} is bounded from below (for fixed and calibrated A and α) and the eigenvalues of M cannot diverge, solving the instability of the initial problem.

5.4.2.4 Is Black-Litterman really Bayesian?

We have seen in the former discussion that the B-L technique aims fundamentally not to correct the estimation of the covariance matrix, but rather insists in the modification of the expected returns vector by the incorporation of specific views. The idea here is to exploit the information in the theoretical CAPM model and the prior information given by some views. Hence, the frequentist estimator $\hat{\mu}$ is replaced by the posterior mean estimator $\overline{ER} := \mathbb{E}(\mu|\bar{R})$ and the original inverse problem becomes:

$$A\hat{\Omega}a_* = \overline{ER}.$$

Therefore, a_* can be rewritten:

$$a_* \quad = \quad \frac{1}{A}[\frac{1}{\alpha}I + (P'F^{-1}P)\hat{\Omega}]^{-1}[(\alpha\hat{\Omega})^{-1}\Pi + P'F^{-1}V].$$

We could interpret the first squared bracket as the Tikhonov regularization of the inverse of $P'F^{-1}P\hat{\Omega}$ and $\frac{1}{\alpha}$ as the regularization parameter. Actually, this interpretation is correct for finite sample size, but, as the sample size increases, $\frac{1}{\alpha}$ increases too, since $\frac{1}{\alpha} = T$. Then, the regularization bias does not decrease and this does not make sense.

In a more appropriate way, we could interpret the Black-Litterman approach as a Bayesian regularization scheme. Bayesian analysis of inverse problems takes the posterior distribution (and the quantities recovered from it) as solution of an inverse problem. The ill-posedness is solved through the incorporation of the prior distribution, at least when the dimension of the problem is finite. The Bayes regularization scheme is convergent in the sense that the bias introduced by the prior distribution disappears as the sample size T increases to ∞. To show this, we rewrite $\overline{ER}$ with Π and α replaced by $\bar{R}$ and $\frac{1}{T}$, respectively. Then, (by neglecting the risk-aversion parameter) :

$$\begin{aligned}
\hat{\Omega} a_* &= (T\hat{\Omega}^{-1} + P'F^{-1}P)^{-1}(T\hat{\Omega}^{-1}\bar{R} + P'F^{-1}V) \\
&= (\hat{\Omega}^{-1} + \frac{1}{T}P'F^{-1}P)^{-1}(\hat{\Omega}^{-1}\bar{R} + \frac{1}{T}P'F^{-1}V) \\
a_*^{\alpha} &= (I + \frac{1}{T}P'F^{-1}P\hat{\Omega})^{-1}(\hat{\Omega}^{-1}\bar{R} + \frac{1}{T}P'F^{-1}V)
\end{aligned}$$

that converges to $\hat{\Omega}^{-1}\bar{R}$ and by the law of large number it converges to $a_* = \Omega^{-1}\mu$. We can interpret $\frac{1}{T}$ as the regularization parameter and we clearly see that the prior distribution (that provides the regularization) disappears as the sample size increases. In conclusion, the B-L correction is a regularization scheme **in the Bayesian sense, but not in the classical sense**. Its stabilization effect on the eigenvalues has been however shown in Section 5.4.2.3.

5.4.2.5 Black-Litterman as a shrinking procedure

Equation (5.2) in the B-L approach suggests to interpret $\overline{ER}$ as deriving from a shrinkage procedure. For this being true we have to suppose that there exists an $n \times 1$ vector υ_0 such that the prior mean of the views can be written as $V = P\upsilon_0$. To be more precise, this is equivalent to say that the investors have a prior information that they translate in the specification of a prior mean for the vector μ, that is $\mathbb{E}([\mu] = \upsilon_0$, where the expectation $\mathbb{E}[\cdot]$ is taken with respect to the prior distribution. Then, from this information on the whole vector μ, they derive the prior information (or view) on l linear combinations of the components of μ. We notice that when $l < N$ the reverse is not feasible, i.e. it is not possible to recover υ_0 from V, due to the singularity of $P'P$. Then, (5.2) can be rewritten as

$$\overline{ER} = [(\alpha\hat{\Omega})^{-1} + P'F^{-1}P]^{-1}[(\alpha\hat{\Omega})^{-1}\Pi + P'F^{-1}P\upsilon_0]$$

and it looks like a convex combination of Π and υ, that is of the two priors on μ, one given by the CAPM and another one given by the subjective views. Hence, the B-L approach shrinks (in the matrix sense) the CAPM-asset excess returns Π towards the view υ and then it corrects the instability of the resulting portfolio. We summarize our result in the following proposition.

Proposition 5.4.2. Let $P\mu \sim \mathcal{N}(V, F)$. If there exists an $n \times 1$ real vector υ_0 such that $V = P\upsilon_0$, then the quantity $\overline{ER}$ found with the B-L approach can be interpreted as a shrinkage estimator, with intensity of reduction equal to

$$[(\alpha\hat{\Omega})^{-1} + P'F^{-1}P]^{-1}P'F^{-1}P.$$

$\square$

The difference between the B-L approach and the shrinkage procedure explained in 5.3.1.1 consists in the fact that the first one **is shrinking a mean vector and the second one is shrinking**

a covariance matrix. We have stressed that this interpretation of B-L as a shrinkage procedure requires to have a prior belief on the whole μ and not only on l linear combinations of components of μ. This assumption is plausible from a Bayesian point of view, since μ is a random vector we believe distributed according to a certain measure. Then, from this distribution, is possible to recover a prior distribution for a linear transformation of μ.

This result is important because it gives a formal statistical justification to the B-L solution. The modified estimator $\overline{ER}$ was originally motivated in Black and Litterman (1991) by practical considerations. We have shown that it is possible to find a mathematical/statistical foundation of this estimator in terms of the modification that it produces on the spectrum, a justification from a Bayesian point of view (since $\overline{ER}$ can be seen as posterior mean) and from a frequentist point of view (by interpreting $\overline{ER}$ as a shrinkage estimator). At the best of our knowledge, such an interpretation has not been proposed in previous literature.

5.5 Conclusion

To circumvent the instability of the mean-variance portfolio allocation, practitioners have designed numerous methods. Previous works have clearly shown that numerical problems stem from the inversion of the sample covariance matrix, especially when the number of assets and the number of data are close. In order to unify this framework and to compare those methods from a mathematical perspective, rather from out-of-sample backtests, we have provided a spectral interpretation of those techniques. The inverse writing of the optimization problem opens a field of questions that naturally involve the regularization nature of those techniques.

Methods deriving from spectral cut-off are well adapted to the resolution of this problem since the lowest eigenvalues correspond to “misleading” portfolios that are considered as risk-free but may also be considered as linearly redundant. Random Matrix Theory can be a useful tool in selecting the informative eigenvalues. Performing “pure” spectral cut-off implies however a trade-off for the investor, between portfolio stability and overall variance. The error in the sample eigenvectors should also be a matter of interest. The numerical instability is of course mainly caused by the presence of arbitrarily low eigenvalues, but as eigenvectors are homogenous to portfolios, the study of their behavior should also be taken into account.

In this chapter, we have highlighted the action of shrinkage on the spectrum. We have in particular underlined that shrinkage is is a first order regularization scheme and gave arguments to understand what is the link between sample eigenvalues dispersion and eigenvectors, and highlighted the Bayesian interpretation of the shrinkage estimator.

Alternatively, the Black-Litterman setting has a Bayesian inspiration in several ways, even if we confirmed that the Bayesian reading usually given is false. We have shown that it is possible to find a mathematical justification of the Black-Litterman estimator in terms of the modification that it produces on the spectrum, both from a a Bayesian and from a frequentist point of view. Even if the true source of instability is not directly tackled, the substitution in the estimator of expected returns, bounds in fact from below the eigenvalues of the resulting multiplying matrix. Furthermore, it must be considered technically as a Bayesian regularization scheme, and we show that the B-L procedure may be considered as a shrinkage procedure, under particular specifications of the views, with (schematically) a vector that is shrunken instead of a matrix.

Directions for future work are numerous. The regularization nature of techniques adding a l_1 penalty or a no short-sell constraint would have to be explored. In extension, a systematic lecture of regularization techniques, including Bayesian ones, could extend the list of allocation methods.

5.6 Proofs

5.6.1 Proof of Proposition 5.3.1

Let $(l_j, \varphi_j)_j$ be the eigensystem associated to T_a. For any vector $u \in \mathbb{R}^N$ we have:

$$\Omega_s u = \sum_{k=1}^{N} \Big[\alpha l_k < \varphi_k, u > \varphi_k + (1-\alpha)\hat{\lambda}_k < \hat{\phi}_k, u > \hat{\phi}_k\Big].$$

Since $(\hat{\phi}_k)_{k=1,\ldots,N}$ and $(\varphi_k)_{k=1,\ldots,N}$ are two orthonormal basis, each φ_k can be re-written in the basis made of the $(\hat{\phi}_k)$:

$$\varphi_k = \sum_{j=1}^{N} < \varphi_k; \hat{\phi}_j > \hat{\phi}_j$$

Then:

$$\begin{aligned}
\Omega_s u &= \sum_{j=1}^{N} \Big[\alpha l_j < \sum_{k=1}^{N} < \varphi_j, \hat{\phi}_k > \hat{\phi}_k, u > (\sum_{k=1}^{N} < \varphi_j, \hat{\phi}_k > \hat{\phi}_k) + (1-\alpha)\hat{\lambda}_j < \hat{\phi}_j, u > \hat{\phi}_j\Big] \\
&= \sum_{j=1}^{N} \Big[(1-\alpha)\hat{\lambda}_j < \hat{\phi}_j, u > \hat{\phi}_j + \sum_{k=1}^{N} (< \hat{\phi}_k, \varphi_j > \alpha l_j < \hat{\phi}_k, u >)(\sum_{k'=1}^{N} < \varphi_j, \hat{\phi}_{k'} > \phi_{k'})\Big] \\
&= \sum_{j=1}^{N} \Big[(1-\alpha)\hat{\lambda}_j < \hat{\phi}_j, u > \hat{\phi}_j + \sum_{k'=1}^{N} \sum_{k=1}^{N} < \hat{\phi}_k, \varphi_j > \alpha l_j < \varphi_j, \hat{\phi}_{k'} >< \hat{\phi}_k, u > \hat{\phi}_{k'}\Big] \\
&= \sum_{j=1}^{N} \Big[(1-\alpha)\hat{\lambda}_j < \hat{\phi}_j, u > \hat{\phi}_j\Big] + \sum_{j=1}^{N} \sum_{k'=1}^{N} \sum_{k=1}^{N} \Big[< \hat{\phi}_k, \varphi_j > \alpha l_j < \varphi_j, \hat{\phi}_{k'} >< \hat{\phi}_k, u > \hat{\phi}_{k'}\Big] \\
&= \sum_{j=1}^{N} \Big[(1-\alpha)\hat{\lambda}_j < \hat{\phi}_j, u > + \sum_{k'=1}^{N} \sum_{k=1}^{N} < \hat{\phi}_k, \varphi_{k'} > \alpha l_{k'} < \varphi_{k'}, \hat{\phi}_j >< \hat{\phi}_k, u > \Big]\hat{\phi}_j.
\end{aligned}$$

Taking $u = \hat{\phi}_j$ in the former expression we get:

$$\Omega_s \hat{\phi}_j = \{(1-\alpha)\hat{\lambda}_j + \sum_{u=1}^{N} \alpha l_u < \varphi_u, \hat{\phi}_j >^2\}\hat{\phi}_j.$$

So, the $\hat{\phi}_j$ are also eigenvectors for Ω_s and the eigenvalues of Ω_s are:

$$(1-\alpha)\hat{\lambda}_j + \sum_{u=1}^{N} \alpha l_u < \varphi_u, \hat{\phi}_j >^2 .$$

Hence, the solution a_*^s of $\Omega_s a_*^s = \mu$ is given by:

$$a_*^s = \sum_{j=1}^{N} \frac{1}{(1-\alpha)\hat{\lambda}_j + \sum_{u=1}^{N} \alpha l_u < \varphi_u, \hat{\phi}_j >^2} < \mu, \hat{\phi}_j > \hat{\phi}_j.$$

Considering:

$$q(\alpha, \hat{\lambda}_j) = \frac{\hat{\lambda}_j}{(1-\alpha)\hat{\lambda}_j + \sum_{u=1}^{N} \alpha l_u < \varphi_u, \hat{\phi}_j >^2}$$

we see that shrinkage techniques can be considered as a first-order regularization scheme.

5.6.2 Proof of Proposition 5.4.1

By using Bayes' Theorem and by using Π and $\bar{R}$ without distinction, we have

$$pdf(\mu|\Pi) \propto pdf(\Pi|\mu) pdf(\mu).$$

Hence,

$$\begin{aligned} pdf(\mu|\Pi) &\propto \exp\{-\frac{1}{2}(\bar{R}-\mu)' T\hat{\Omega}^{-1}(\bar{R}-\mu)\} \times \\ &\quad \exp\{-\frac{1}{2}(P\mu - V)' F^{-1}(P\mu - V)\} \\ &\propto \exp\{-\frac{1}{2}(\mu - \overline{ER})'(VR)^{-1}(\mu - \overline{ER})\}, \end{aligned}$$

where $\overline{ER}$ and VR are determined respectively by

$$\begin{aligned} \mu'(VR)^{-1}\mu &= \mu' T\hat{\Omega}^{-1}\mu + \mu' P' F^{-1} P\mu \\ \Leftrightarrow tr(\mu\mu'(VR)^{-1}) &= tr(\mu\mu' T\hat{\Omega}^{-1}) + tr(\mu\mu' P' F^{-1} P) \\ \Rightarrow (VR) &= \left(T\hat{\Omega}^{-1} + P' F^{-1} P\right)^{-1} \end{aligned}$$

and

$$\begin{aligned} \mu'(VR)^{-1}\overline{ER} &= \mu' T\hat{\Omega}^{-1}\bar{R} + \mu' P' F^{-1} V \\ \Rightarrow \overline{ER} &= \left(T\hat{\Omega}^{-1} + P' F^{-1} P\right)^{-1}\left(T\hat{\Omega}^{-1}\bar{R} + P' F^{-1} V\right). \end{aligned}$$

By denoting α for $\frac{1}{T}$ and if we write Π instead of $\bar{R}$ (since $\bar{R}$ is a consistent estimator of Π) we get the results in (5.4.2.1).

$\square$

Conclusion

It is surprising to see how finance can offer to the econometrician numerous pertinent problems with complex theoretical implications. For instance, as shown in the last chapter, Inverse Problems theory provides an elegant reading of portfolio allocation problems. This approach is not really new, but writing explicitly those links has allowed us to draw a clear view of usual allocation methods from a regularization point of view. We clarified the Bayesian nature of some methods (known for long by the practitioners) with original results and potentially open the way for new methods.

Our main financial focus has however been the study of Hedge Funds. This has implied a huge empirical work. The recent global financial turmoils have led to future dramatic changes in their monitoring and their regulation. The goal of the empirical part of this work was however to enhance the understanding of the imperfect set of information available in databases. Maybe in the future, those databases will be in turn improved and deeply structured. In this case, I hope that this work would be a useful contribution. Nevertheless, working on an imperfect, biased set of data is both the nightmare and the pleasure of an econometrician! Therefore, we particularly insist on the importance of censorship under its various forms. First, nonparametric estimation greatly improves the analysis, and the use of covariates accounts for a cohort effect that captures most of the information for the explanation of Hedge Funds lifetimes. But most of all, a crucial feature is that Hedge Funds may have several reasons to exit from a database, not only because of failure. We confirm this old academic intuition by providing new empirical evidences. Consequently, it is pertinent to replace the usual single-risk approach by a competing-risks model that allows to estimate precisely the behaviors of the two competing reasons of exit, failure and...success! Further work (on identifiability conditions, heterogeneity, etc.) would however be needed to complete this approach.

Our main financial contribution has been to prove the relevance of introducing the concept of endogeneity in Hedge Funds study. Many (if not all) models developed in the literature assume exogenous explanatory variables, conditional duration models and dismiss endogeneity in Hedge Funds studies. This is even more pronounced in the analysis of their lifetimes and their survival. We have provided strong evidence for the contrary: endogeneity is a topic of utmost interest in financial studies. By nature, Hedge Funds are in this respect a wonderful playground to test such ideas. The impact of the lockup variable on the lifetimes of the funds is a first example. But even for this simple variable, we gave estimation and sharp figures for effects that have, up to our knowledge, only been discussed theoretically. We hope that this leads the way for other studies in the same spirit.

The fact that endogeneity is simply left aside in studies is not surprising since the concept of endogeneity is neither easy to define nor to understand. Endogeneity is even hard to detect, or to disentangle from subtle concepts like causality or counterfactuality. The theoretical part of this work is a first step towards a better understanding of endogeneity in a dynamic context. The heart of this work was to present two classes of models for stochastic processes with endogenous variables, treated with the instrumental variables method. We reach in a case a dynamic extension of separable model through a generalization of the standard Doob-Meyer decomposition of semi-martingales. The case of duration model is original in the sense that we observe that using a change of time is compulsory in this situation. As in the usual case, the functional parameters of interest are here again characterized as solutions of (potentially sophisticated) integral equations. We have been mainly concerned by modelling and a lot of work has to be done to build a whole estimation framework (estimation procedures, practical aspects, limit properties, etc.). Unfortunately, many objects that have been introduced in our analysis cannot be estimated through their present expression. But I would be happy if one day, someone could exploit this novel framework to develop the concepts introduced here.

Bibliography

Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View*. Statistics for Biology and Health. Springer-Verlag, New-York.

Abbring, J. and Heckman, J. (2008). Dynamical Policy Analysis. In *The Econometrics of Panel Data*, volume 46, chapter 24, pages 795–863. Springer.

Abbring, J. and van den Berg, G. (2003a). A Simple Procedure for the Evaluation of Treatment Effects on Duration Variables. *IZA Discussion Paper*, (810).

Abbring, J. and van den Berg, G. (2003b). The Nonparametric Identification of Treatment Effects in Durations Models. *Econometrica*, 71(5):1491–1517.

Abbring, J. and van den Berg, G. (2004). Analyzing the effect of dynamically assigned treatments using duration models, binary treatment models, and panel data models. *Empirical Economics*, 29:5–20.

Ackermann, C., McEnally, R., and Ravenscraft, D. (1998). The Impact of Regulatory Restrictions on Fund Performance: a Comparative Study of Hedge Funds and Mutual Funds. *University of North Carolina Working Paper*.

Ackermann, C., McEnally, R., and Ravenscraft, D. (1999). The Performance of Hedge Funds: Risk, Return and Incentives. *Journal of Finance*, 54(3):833–874.

Agarwal, V., Daniel, N., and Naik, N. (2009). Role of Managerial Incentives and Discretion in Hedge Fund Performances. *Journal of Finance*, 64(5):2221–2256.

Agarwal, V., Fos, V., and Jiang, W. (2010). Inferring Reporting Biases in Hedge Fund Databases from Hedge Fund Equity Holdings. *CFR Working Papers from University of Cologne, Centre for Financial Research*, (10-08).

Agarwal, V. and Naik, N. (2000). Multi-Period Performance Persistence Analysis of Hedge Funds. *Journal of Financial and Quantitative Analysis*, 35(03):327–342.

Ai, C. and Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, 71(6):1795–1843.

Alt, H. (2006). *Lineare Funktionalanalysis*. Springer-Lehrbuch Masterclass. Springer.

Amin, G. and Kat, H. (2002). Welcome to the Dark Side: Hedge Fund Attrition and Survivorship Bias over the Period 1994-2001. *ICMA Centre Discussion Paper*.

Ang, A. and Bollen, N. (2010). Locked Up by a Lockup : Valuing Liquidity as a Real Option. *Financial Management*, 39(3):1069–1096.

Angrist, J. and Imbens, G. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the Statistical Association*, 91(434):444–455.

Aragon, G. (2007). Share Restrictions and Asset Pricing: Evidence From The Hedge Fund Industry. *Journal of Financial Economics*, 83(1):33–58.

Aragon, G. and Nanda, V. (2009). On Tournament Behavior in Hedge Funds: High Water Marks, Fund Liquidation, and the Backfilling Bias. *Working Paper*.

Aït-Sahalia, Y. (2002). Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, 70(1):223–262.

Avellaneda, M. and Besson, P. (2007). Hedge Funds: How Big is Big. *NYU Mathematics in Finance Working Paper Series*, (2007-2).

Baquero, G., ter Horst, J., and Verbeek, M. (2005). Survival, Look-Ahead Bias and the Performance of Hedge Funds. *Journal of Financial and Quantitative Analysis*, 40(3):493–524.

Barry, R. (2002). Hedge Funds: a Walk through the Graveyard. *Applied Finance Center, Macquarie University, Sydney, Australia, Working Paper*.

Barès, P., Gibson, R., and Gyger, S. (2007). The Style Consistency of Hedge Funds. 13(2).

Beare, B. (2009). Distributional Replication. *UCSD Economics Working Paper*.

Ben-David, I., Franzoni, F., and Moussawi, R. (2010). The Behavior of Hedge Funds during Liquidity Crises. *Working Paper*.

Berk, J. and Green, R. (2004). Mutual Fund Flows and Performance in Rational Markets. *Journal of Political Economy*, 112(6):1269–1295.

Bernanke, B. (2006). Hedge Funds and Systemic Risk. In *Federal Reserve Bank of Atlanta's 2006 Financial Market Conference "Hedge Funds: Creators of Risk?"*.

Bijwaard, G. (2007). Instrumental Variable Esimation of Treatment Effects for Duration Outcomes. *Econometric Institute Report, Erasmus University*, 2007-20.

Bijwaard, G. (2008). Instrumental Variable Estimation for Duration Data. In Engelhardt, H., Kohler, H., and Furnkranz-Prskawetz, A., editors, *Causal analysis in population studies: concepts, methods, applications*, volume 23, chapter 6, pages 111–148. Springer.

Black, F. and Litterman, R. (1991). Global Asset Allocation with Equities, Bonds, and Currencies. *Fixed Income Research, Goldman, Sachs & Company*.

Black, F. and Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal*, 48:28–43.

Blundell, R. and Horowitz, J. (2007). A Nonparametric Test of Exogeneity. *The Review of Economic Studies*, 74(4):1035–1058.

Blundell, R. and Powell, J. (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. In Dewatripoint, M., Hansen, L., and Turnsovsky, S., editors, *Advances in Economics and Econometrics*, chapter 8, pages 312–357. Cambridge University Press.

Bollen, N. and Pool, V. (2009). Do Hedge Fund Managers Misreport Returns? Evidence from the Pooled Distribution. *Journal of Finance*, 64:2257–2288.

Bouchaud, J.-P., Laloux, L., Miceli, A., and Potters, M. (2007). Large Dimension Forecasting Models and Random Singular Value Spectra. *The European Physical Journal B*, 55(2):201–207.

Boyson, N. (2002). How are Hedge Fund Managers Characteristics related to Performance, Volatility and Survival. *Ohio State University Working Paper*.

Boyson, N., Stahel, C., and Stulz, R. (2010). Is There Hedge Fund Contagion and Liquidity Shocks. *Journal of Finance*, 65(5):1789–1816.

Brémaud, P. (1991). Point Processes and Queues, Martingale Dynmamics. *Springer Verlag*.

Broadie, J., Daubechies, I., de Mol, C., Giannone, D., and Loris, I. (2009). Sparse and Stable Markowitz Portfolios. *Proceedings of the National Academy of Science*, 106:12267–12272.

Brown, S., Goetzmann, W., and Ibbostson, R. (1999). Offshore Hedge Funds: Survival and Performance. *Journal of Business*, 72:91–117.

Brown, S., Goetzmann, W., and Park, J. (2001). Careers and Survival : Competition and Risk in the Hedge Fund and CTA industry. *Journal of Finance*, 56(5):1869–1886.

Brown, S., Goetzmann, W., and Park, J. (2004). Conditions for Survival: Changing Risk and the Performance of Hedge Fund Managers and CTAs. *Yale Working Paper*.

Carrasco, M. (2008). A Regularization Approach to the Many Instruments Problem. *Mimeo Montreal*.

Carrasco, M., Florens, J.-P., and Renault, E. (2003a). Estimation of the Solution of an Integral Equation of the Second Kind. *IDEI Working Paper*.

Carrasco, M., Florens, J.-P., and Renault, E. (2003b). Linear Inverse Problems in Structural Econometrics: Estimation based on Spectral Decomposition and Regularization. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6B. Elsevier.

Chabi-Yo, F., Ghyselsa, E., and Renault, E. (2008). On Portfolio Separation Theorems with Heterogeneous Beliefs and Attitudes towards Risk. *Working Paper*, 2008-16.

Chan, N., Getmansky, M., Haas, S. M., and Lo, A. W. (2007). *Systemic Risk and Hedge Funds*. NBER Chapters. National Bureau of Economic Research, Inc.

Chernozhukov, V., Gaglliardini, P., and Scaillet, O. (2007a). Nonparametric Instrumental Variable Estimators of Quantile Structural Effects. *Swiss Finance Institute Working Paper*.

Chernozhukov, V. and Hansen, C. (2005). An IV Model of Quantile Treatment Effects. *Econometrica*, 73(1):245–261.

Chernozhukov, V. and Hansen, C. (2006). Instrumental Quantile Regression Inference for Structural and Treatment Effect Models. *Journal of Econometrics*, 132(2):495–525.

Chernozhukov, V., Imbens, G., and Newey, W. (2007b). Instrumental Variable Identification and Estimation of Nonseparable Models. *Journal of Econometrics*, 139(1):4–14.

Chesher, A. (2002). Semiparametric identification in duration models. *CEMMAP Working Paper (CWP20/02).*

Choueifaty, Y. and Coignard, Y. (2008). Toward Maximum Diversification. *The Journal of Portfolio Management*, 35(1):40–51.

Copas, J. (1983). Regression, Prediction and Shrinkage. *Journal of Royal Statistical Soc. B*, 45(3):311 – 354.

Couderc, F., Renault, O., and Scaillet, O. (2008). Business and Financial Indicators: What are the Determinants of Default Probability Changes? In *Credit Risk: Models, Derivatives, and Management*, chapter 14, pages 235–268. Chapman and Hall.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2010a). Nonparametric Instrumental Regression. *GREMAQ Discussion Paper.*

Darolles, S. and Gouriéroux, C. (2009). The Effects of Management and Provision Accounts on Hedge Fund Returns. *Working Paper.*

Darolles, S., LeFol, G., Ozik, G., and Sadka, R. (2010b). Funds of Hedge Funds Reputation and the High Risk of Low Volatility Funds. *Working Paper.*

Dellacherie, C. and Meyer, P.-A. (1971). Probabilités et Potentiels. *Hermann.*

Demey, P., Maillard, S., and Roncalli, T. (2010). Risk-Based Indexation. *Lyxor White Paper.*

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). How Inefficient Are Simple Asset-Allocation Strategies? *Review of Financial Studies*, 22(5):1915–1953.

Derman, E. (2007). A Simple Model for the Expected Premium for Hedge Fund Lockups. *Journal of Investment Management*, 5(3).

Detemple, J., Garcia, R., and Rindisbacher, M. (2005). Representation Formulas for Malliavin Derivatives of Diffusion Processes. *Finance and Stochastics*, 9(3):349–367.

d'Haultfoeuille, X. (2010). On the Completeness Condition in Nonparametric Instrumental Problems. *Econometric Theory.*

Ding, B., Getmansky, M., Liang, B., and Wermers, R. (2009). Investor Flows and Share Restrictions in the Hedge Fund Industry. *Working Paper.*

Doss, H. (1977). Liens entre équations différentielles stochastiques et ordinaires. *Annales de l'IHP*, 13(2):99–125.

Edwards, F. R. (1999). Hedge Funds and the Collapse of Long- Term Capital Management. *Journal of Economic Perspectives*, 13(2):189–210.

El-Karoui, N. and Meyer, P.-A. (1977). Les Changements de Temps en Théorie Générale des Processus. 8:65–78.

El-Karoui, N. and Weidenfeld, G. (1977). Théorie Générale et Changements de Temps. In *Séminaire de probabilités (Strasbourg)*, volume 11, pages 79–108. Springer.

Engl, H., Hanke, M., and Neubauer, A. (1996). Regularization of Inverse Problems. *Kluwer Academic Publishers, Mathematics and Its Applications.*

Engle, R., Hendry, D., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51(2):277–304.

Fama, E. (1968). Risk, Return and Equilibrium: Some Clarifying Comments. *Journal of Finance*, 23(1):29–40.

Fan, J., Fan, Y., and Lv, J. (2008). High Dimensional Covarance Matrix Estimation Using a Factor Model. *Journal of Econometrics*, 147(1):186–197.

Florens, J.-P. (2003). Inverse Problems in Structural Econometrics. The Example of Instrumental Variables. 2.

Florens, J.-P. (2005). Endogeneity in non separable models. Application to treatment models where the outcomes are durations.

Florens, J.-P. and Fougère, D. (1996). Noncausality in Continuous Time. *Econometrica*, 64(5):1195–1212.

Florens, J.-P., Johannes, J., and van Bellegem, S. (2009). Instrumental Regression in Partially Linear Models. *IDEI Working Paper*, (613).

Florens, J.-P. and Mouchart, M. (1982). A Note on Noncausality. *Econometrica*, 50(3):583–592.

Florens, J.-P., Mouchart, M., and Rolin, J.-M. (1990). Elements of Bayesian Statistics. *Marcel Dekker, New-York.*

Fung, W. and Hsieh, D. (1997a). Empirical Characteristics of Dynamic Trading Strategies. *The Review of Financial Studies*, 10(2):275–302.

Fung, W. and Hsieh, D. (1997b). Survivorship Bias and Investment Style in the Return of CTAs. *Journal of Portfolio Management*, 24(1):30–41.

Fung, W. and Hsieh, D. (2000). Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases. *Journal of Financial and Quantitative Analysis*, 35:291–307.

Fung, W. and Hsieh, D. (2006). Hedge Funds: An Industry in its Adolescence. In *Federal Reserve Bank of Atlanta's 2006 Financial Market Conference "Hedge Funds: Creators of Risk?"*.

Gagliardini, P. and Scaillet, O. (2009). Tikhonov Regularization for Nonparametric Instrumental Variables Estimators. *Swiss Finance Institute Research Paper.*

Gard, T. (1988). Introduction to Stochastic Differential Equations. *Marcel Dekker, New-York.*

Getmansky, M. (2004). *What Drives Hedge Fund Returns? Models of Flows, Autocorrelation, Optimal Size, Limits to Arbitrage and Fund Failures.* PhD thesis, MIT Sloan School of Management.

Getmansky, M., Lo, A., and Igor Makarov, I. (2004a). An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, 74(3):529–609.

Getmansky, M., Lo, A., and Mei, S. (2004b). Sifting Through the Wreckage: Lessons from Recent Hedge-Fund Liquidations. *Journal of Investment Management, Fourth Quarter*, 2(4):6–38.

Gill, R. D. and Robins, J. (2001). Causal Inference for Complex Longitudinal Data: the Continuous Case. *Annals of statistics*, 29(6):1785–1811.

Goldfarb, D. and Iyengar, G. (2003). Robust Portfolio Selection Problems. *Mathematics of Operations Research*, 28(1):1–38.

Gourieroux, C. and Jasiak, J. (2001). *Financial Econometrics: Problems, Models, and Methods.* Princeton University Press.

Grecu, A., Malkiel, B., and Saha, A. (2007). Why do Hedge Funds Stop Reporting their Performance? *Journal of Portfolio Management*, 34(1):119–126.

Gregoriou, G. (2002). Hedge Fund Survival Lifetimes. *Journal of Asset Management*, 3(3):237–252.

Gregoriou, G. and Rouah, F. (2002). Is Size a Factor in Hedge Fund Performances? *Derivatives Use, Trading and Regulation*, 7(4):301–306.

Gupta, B. and Kazemi, H. (2008). Abnormal Return Patterns and Hedge Fund Failures. *Journal of Risk Management in Financial Institutions*, 2(1):88–106.

Haff, L. (1980). Empirical Bayes Estimation of the Multivariate Normal Covariance Matrix. *The Annals of Statistics*, 8(3):586–597.

Hall, P. and Horowitz, J. (2005). Non-Parametric Methods for Inference in the Presence of Instrumental Variables. *Annals of Statistics*, 33(6):2904–2929.

Heckman, J. (2000). Microdata, Heterogeneity and the Evaluation of Public Policy. *Nobel Prize Lecture.*

Heckman, J. (2001). Econometrics Counterfactuals and Causal Models. *Keynote Address, International Statistical Institute, Seoul.*

Heckman, J. and Robb, R. (1986). Econometric Causality. *Journal of Econometrics*, 30(1-2):239–267.

Heckman, J. and Singer, B. (1984). The Identifiability of the Proportional Hazard Model. *Review of Economic Studies*, 51:231–241.

Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738.

Heckman, J. and Vytlacil, E. (2007). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6, chapter 70. Elsevier.

Hendricks, D., Patel, J., and Zeckhauser, R. (1997). The J-Shape Of Performance Persistence Given Survivorship Bias. *The Review of Economics and Statistics, MIT Press*, 79(2):161–166.

HFR (2009). HFR Global Hedge Fund Industry Report, Second Quarter 2009.

Hoderlein, S. and Mammen, E. (2007). Identification of Marginal Effects in Nonseparable Models without Monotonicity. *Econometrica*, 75(5):1513–1518.

Hoerl, A., , and Kennard, R. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67.

Holland, A. S., Kazemi, H. B., and Li, Y. (2010). Compensation Options, Managerial Incentives, and Risk-Taking in Hedge Funds. *Working Paper, University of Washington Bothell.*

Hombert, J. and Thesmar, D. (2009). Limits of Limits of Arbitrage : Theory and Evidence. *Working Paper.*

Horowitz, J. and Lee, S. (2007). Nonparametric Instrumental Variables Estimation of a Quantile Regression Model. *Econometrica*, 75(4):1191–1208.

Jacobsen, M. (2005). *Point Processes Theory and Applications, Marked Point and Piecewise Deterministic Processes.* Birkhäuser Boston.

Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios, Why Imposing the Wrong Constraints Help. *Journal of Finance*, 58(4):1651–1684.

Jondeau, E. and Rockinger, M. (2006). Optimal Portfolio Allocation under Higher Moments. *European Financial Management*, 12(1):29–55.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data, 2nd Edition.*

Kaltenbacher, B., Neubauer, A., and Scherzer, O. (2008). *Iterative Regularization Methods for Nonlinear Ill-Posed Problems.* Walter de Gruyter.

Karr, A. (1991). Point Processes And Their Statistical Inference. *Marcel Dekker Inc., New York and Basel.*

Kazamaki, N. (1972). Changes of Time, Stochastic Integrals and Weak Martingales. *Probability Theory and Related Fields, Springer, Berlin*, 22(1):25–32.

Khandani, A. and Lo, A. (2007). What Happened to the Quants in August 2007? *Journal of Investment Management*, 5:29–78.

Kluve, J. (2004). On the Role of Counterfactuals in Inferring Causal Effects of Treatments. *Foundations of Science*, 9:65–101.

Kundro, C. and Feffer, S. (2003). Understanding and Mitigating Operational Risk in Hedge Fund. *Capco White Paper.*

Lancaster, T. (1992). *The Econometric Analysis of Transition Data.* Cambridge University Press.

Ledoit, O. and Wolf, M. (2002). Some Hypothesis Tests for the Covariance Matrix when the Dimension is Large Compared to the Sample Size. *Annals of Statistics*, 30(4):1081– 1102.

Ledoit, O. and Wolf, M. (2003). A Well-Conditioned Estimator for Large Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2004). Honey, I Shrunk the Sample Covariance Matrix . *Journal of Portfolio Management*, 31(1):1–22.

Lee, W. (2000). *Theory and Methodology of Tactical Asset Allocation* . Wiley.

LeJan, Y. (1979). Martingales et Changements de Temps. *Séminaire de Probabilités, XIII, (Univ. Strasbourg, 1977), Lecture Notes in Mathematics, 721, Springer, Berlin*, pages 385–399.

Lerner, J. and Schoar, A. (2009). The Illiquidity Puzzle: Theory and Evidence. *Journal of Financial Economics*, 72(1):3–40.

Li, Q. and Racine, J.-S. (2006). *Nonparametric Econometrics : Theory and Practice.*

Liang, B. (1999). On the Performance of Hedge Funds. *Financial Analysts Journal*, 55:72–85.

Liang, B. (2000). Hedge Funds: the Living and the Dead. *Journal of Financial and Quantitative Analysis*, 35(3):309–335.

Liang, B. and Park, H. (2008). Share Restrictions, Liquidity Premium, and Offshore Hedge Funds . *Working Paper.*

Lok, J. (2008). Statistical Modelling of Causal Effects in Continuous Time. *Annals of Statistics*, 36(3):1464–1507.

Lok, J., Gill, R., Vaart, A. V. D., and Robins, J. (2004). Estimating the Causal Effect of a Time-Varying Treatment on Time-to-Event using Structural Nested Failure Time Models. *Statistica Neerlandica*, 58(3):271–295.

Lunde, A., Timmermann, A., and Blake, D. (1999). The Hazards of Mutual Fund Underperformance: A Cox Regression Analysis. *Journal of Empirical Finance*, 6:121–152.

Malevergne, Y. and Sornette, D. (2004). Collective Origin of the Coexistence of Apparent RMT Noise and Factors in Large Sample Correlation Matrices. *Physica A: Statistical Mechanics and its Applications*, 331(3-4):660–668.

Marcenko, V. and Pastur, L. (1967). Distribution of Eigenvalues of Some Sets of Random Matrices. *Math. USSR - Sb.*, 1:507 – 536.

Markowitz, H. (1952). Portfolio selection. *Journal Of Finance*, 7(1):77–91.

Martellini, L. (2009). Toward the Design of Better Equity Benchmarks. *Journal of Portfolio Management*, 34(4):1–8.

Matzkin, R. (2003). Nonparametric Estimation of Nonadditive Random Functions. *Econometrica*, 71(5):1339–1375.

Mehta, M. (2004). *Random Matrices*, volume 142. Academic Press, 3 edition.

Merton, R. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, 8(4):323–361.

Meucci, A. (2008). Fully Flexible Views, Theory and Practice. *Risk Magazine*, 21(10):97–102.

Michaud, R. (1989). The Markowitz Optimization Enigma: Is "Optimized" Optimal. *Financial Analysts Journal*, 45:31–42.

Morgan, S. and Winship, C. (2007). Counterfactual and Causal Inference: Methods for Principles for Social Research.

Muirhead, R. (1987). Developments in Eigenvalue Estimation. In Gupta, A. and D.Reidel, editors, *Advances in Multivariate Statistical Analysis*, pages 277–288. Springer.

Newey, W. and Powell, J. (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71(5):1565–1578.

Pafka, S. and Kondor, I. (2004). Estimated Correlation Matrices and Portfolio Optimization. *Physica A-Statistical Mechanics and its Applications*, 343:623–634.

Park, J. Y. (2010). Martingale Regressions for Conditional Mean Models in Continuous Time. *Working Paper*.

Patilea, V. and Rolin, J.-M. (2006). Product-Limit Estimators of the Survival Function with Twice-Censored Data. *The Annals of Statistics*, 34(2):925–938.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Plantin, G. and Makarov, I. (2011). Rewarding Trading Skills Without Inducing Gambling. *Working Paper*.

Pojarliev, M. and Levich, R. (2010). Trades of the Living Dead: Style Differences, Style Persistence and Performance of Currency Fund Managers. *Journal of International Money and Finance*, pages 1752–1775.

Potters, M., Bouchaud, J.-P., and Laloux, L. (2005). Financial Applications of Random Matrix Theory: Old Laces and New Pieces. *Acta Physica Polonica B*, 36(9):623–634.

Protter, P. (2003). Stochastic Integration and Differential Equations (2nd Edition). *Springer*.

Qian, E. (2005). Risk Parity Portfolios: Efficient Portfolios through True Diversification. *PanAgora Asset Management*.

Ramlau-Hansen, H. (1983). Smoothing Counting Processes Intensities by Means of Kernel Functions. *The Annals of Statistics*, 11(2):453–466.

Rouah (2005). Competing Risks in Hedge Funds Lifetimes. *Working Paper*.

Sadka, R. (2010). Liquidity Risk and the Cross-Section of Hedge-Fund Returns. *Journal of Financial Economics*, 98:54–71.

Satchell, S. and Scowcroft, A. (2000). A Demystification of the Black-Litterman Model: Managing Quantitative and Traditional Portfolio Construction. *Journal of Asset Management*, 1(2):138–150.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Specht, K. and Gohout, W. (2003). Portfolio Selection using the Principal Components GARCH model. *Financial Markets and Portfolio Management*, 17(4):450–458.

Stein, C. (1986). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Third Berkeley Symposium on Mathematical and Statistical Probability*, volume 1, pages 197–206. University of California, Berkeley.

Teo, M. (2010). The Liquidity Risk of Liquid Hedge Funds. *Journal of Financial Economics, forthcoming.*

Tikhonov, A. and Arsenin, V. (1977). *Solution of Ill-posed Problems*. H. Winston and Sons (distributed by Wiley, New York).

Tsiatis (1975). A nonidentifiability aspect of competing risks. *Proc. Nat. Acad. Sci USA*, 72(1):20–22.

van den Berg, G. (2000). Duration Models: Specification, Identification and Multiple Durations. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6B, chapter 55. Elsevier.

Vinod, H. D. (1976). Canonical Ridge and Econometrics of Joint Production. *Journal of Econometrics*, 4(2):147–166.

Wang, Q. and Phillips, P. (2009). Structural Nonparametric Cointegrating Regression. *Econometrica*, 77(6):1901–1948.

Windcliff, H. and Boyle, P. (2004). The 1/N Pension Plan Puzzle. *North American Actuarial Journal*, 8:32–45.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. *Wiley-Interscience.*

Appendix A

Appendix

As this work lies at the confluence of various fields of literature, we give in this section some mathematical and financial complements, to make the concepts used in this manuscript more understandable. We will review some fundamental properties and definitions on stochastic processes (Section A.1) and inverse problems (Section A.2). Additionally we will sum up the basics of estimation in duration models (Section A.3). Finally, we provide a financial glossary and a more precise description of the hedge fund industry (Section A.4).

A.1 Stochastic processes: complements

A **filtered probability space** is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ that is to say an increasing family of σ-algebras that are all sub-algebras of $\mathcal{F}$. A probability space is **complete** if $\forall F \in \mathcal{F}$ such as $\mathbb{P}(F) = 0$, then $f \in \mathcal{F}$ as soon as $f \subset F$. A complete filtered probability space is said to satisfy the **usual conditions** if:

- $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets of $\mathcal{F}$;
- the filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous i.e. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.

In continuous time, a **stochastic process** $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ taking its values in $(H, \mathcal{H})$ (called the state space most of the time), is a family of random variables such as for each t, X_t is a mapping between $(\Omega, \mathcal{F})$ and $(H, \mathcal{H})$. For a given $w \in \Omega$, $t \mapsto X_t(w)$ is a trajectory of the process. A process is said to be **continuous** (respectively **right-continuous**, **left-continuous**, **increasing**) if $\mathbb{P}$-almost surely, for events $w \in \Omega$, the trajectories $X_t(w)$ are continuous (respectively right-continuous, left-continuous, increasing). In the same way, a process is said to be **càdlàg** if $\mathbb{P}$-a.s. its trajectories are right-continuous with left limits.

X is said to be **measurable** if $(t, w) \mapsto X_t(w)$ is measurable as a map from $([0;\infty[\times\Omega; \mathcal{B}_0 \bigotimes \mathcal{F})$ to $(H, \mathcal{H})$. Moreover, the process is said to be **adapted** if it is measurable, and for $t \in \mathbb{R}$, X_t is $\mathcal{F}_t$-measurable as a mapping from $(\Omega, \mathcal{F})$ to $(H, \mathcal{H})$. Intuitively, a process is adapted if its value at t may be calculated with the corresponding filtration at date t, i.e. all the available information at date t. In particular, an adapted counting process would be piecewise deterministic.

X is said to be **progressive** or **progressively measurable** if for $t \in]0;\infty[$, $(t,w) \mapsto X_t(w)$ as a mapping from $([0;t] \times \Omega; \mathcal{B}([0;t]) \bigotimes \mathcal{F}_t)$ to $(H, \mathcal{H})$ is measurable. X is said to be **predictable** when X_0 is $\mathcal{F}_0$-measurable and that $(t,w) \mapsto X_t(w)$ as a mapping from $(]0;\infty[\times\Omega; \mathcal{B}_+ \bigotimes \mathcal{F})$ to $(H,\mathcal{H})$ is measurable with respect to the the σ-algebra of predictable sets $\mathcal{P}\int$ which is the sub σ-algebra of $\mathcal{B}_+ \bigotimes \mathcal{F}$ generated by:

$$]t;\infty[\times A \quad \text{with} \quad t \in]0;\infty[, A \in \mathcal{F}_t.$$

It can also be generated by the segments of $[0;\infty[\times\Omega$, of the form: $[s;t] \times A$, $\forall s,t, 0 \leq s \leq t < \infty$ and $A \in \mathcal{F}_s$. Intuitively, when a process is assumed to be predictable, its value at t can be calculated with all the information anterior to t that is on $[0;t[$.

A predictable process is progressive, measurable. A progressive process is adapted. A process which is left-continuous and measurable is predictable. A process which is right-continuous and adapted is progressive.

A stopping time τ is said to be **predictable** if $\tau > 0$ is the limit of an increasing sequence of stopping times τ_n such that $\tau_n < \tau$. An **accessible** stopping time τ is such that there exists a sequence predictable τ_n of stopping times, such that the probability that there exists $n_0 \in \mathbb{N}$ such that $\tau = \tau_{n_0}$ is equal to one. Conversely, a stopping time τ is **totally inaccessible** if no such sequence of stopping times exists. In other terms, the probability that $\tau = \tau'$ is equal to zero for each predictable time τ'.

A.2 Inverse Problems: complements

A space is **locally compact** if each point has a compact neighborhood. A subspace is **relatively compact** is a subset whose closure is compact.

A.2.1 Compact operators

An operator $K^1 : \mathcal{E} \to \mathcal{F}$ is said to be **compact** between two Banach spaces $\mathcal{E}$ and $\mathcal{F}$ as soon as $K(\mathcal{B}_\mathcal{E})$ is relatively compact in $\mathcal{F}$ ($\mathcal{B}_\mathcal{E}$ being $\{x \in \mathcal{E} || ||x|| \leq 1\}$) i.e. $\overline{K(\mathcal{B}_\mathcal{E})}$ is compact in $\mathcal{F}$. An other definition is that for any bounded sequence $\{x_n\}$ of $\mathcal{E}^\mathbb{N}$, $\{Kx_n\}$ has converging subsequences in $\mathcal{F}$.

If $\mathcal{K}(\mathcal{E}, \mathcal{F})$ is the vector space of compact operators, $\mathcal{K}(\mathcal{E}, \mathcal{F})$ is closed in the space of linear operators $\mathcal{L}(\mathcal{E}, \mathcal{F})$ and **contains all the finite rank operators**. In particular, K compact is equivalent to K^*, KK^*, and K^*K compact. It implies also that K is continuous and bounded (equivalence in the linear case).
Moreover, if $\mathcal{F}$ is a Hilbert space, any compact operator $K : \mathcal{E} \longrightarrow \mathcal{F}$ is limit in $\mathcal{L}(\mathcal{E}, \mathcal{F})$ of a sequence of operators of finite rank. This property is the source of main regularization techniques that will be explored throughout the following sections. How a compact operator looks like? We have already seen that finite rank operators are compact. An easy characterization is through the **Hilbert-Schmidt condition** (henceforth H-S) since H-S operators are compact and are often met in common problems.

The study of compact operators is more thrilling in infinite dimension. In finite dimension, an operator is one-to-one as soon as it is surjective. So the injectivity is the only notion to check in order to assess the invertibility of the operator. In infinite dimension however, a bounded operator may be surjective without being injective and reciprocally. What will help us is the precise study of the behavior of the operator as expressed before.

However when $dim(E) = \infty$, then for $T = K$ compact, $0 \in Sp(K)$ where $Sp(K)$ is the spectrum of K. We have moreover either one of the three situations:

1. $Sp(K) = \{0\}$
2. $Sp(K) = \{0\}$ is finite
3. $Sp(K) = \{0\}$ is a sequence that tends to 0.

We will now suppose that $dim(E) = +\infty$. In fact, there is a more convenient object to define which is the singular value. When K is self-adjoint (still compact, linear) its eigensystem exist and is characterized by the sequence (λ_n, ϕ_n) of eigenvalues λ_n and eigenvectors such that $K\phi_n = \lambda_n \phi_n \quad \forall n \in \mathbb{N}$. K may be rewritten:

$$K\phi = \sum_{n=1}^{\infty} \lambda_n < \phi; \phi_n > \phi_n.$$

Unfortunately, such an eigensystem do not exist for an operator K which is not self-adjoint. The singular system generalizes this definition. For $K : \mathcal{H} \to \mathcal{G}$ a compact operator, linear, between two Hilbert spaces, the roots of the eigenvalues of K^*K are the *singular values.* The advantage of compact operators is that their spectrum is *discrete* and that it may be characterized easily. A compact operator

[1]Henceforth, the generic operator T will be noted K when it is assumed to be compact.

is injective iff all its singular values are different from 0. If it is the case, 0 is not an eigenvalue of K^*K but may be one of KK^*. $(\lambda_j, \phi_j, \psi_j)$ is called a *singular system*.

A.2.2 Least-squares solutions and inverse problems

A.2.2.1 Generalities

It's also important to view inverse problems through the study of best-approximate solutions. Let's recall that if $T : \mathcal{H} \to \mathcal{G}$ is a bounded linear operator, $f \in \mathcal{H}$ is the *best-approximate solution* of 1.6 if f is a least-square solution of 1.6 and if :

$$||f|| = inf\{||\phi|| \quad | \quad \phi \quad \text{is a least-square solution of 1.6}\}.$$

The Moore-Penrose inverse of T generalizes the notion of inverse operator and is defined as the operator mapping y with the corresponding best-approximate solution. More precisely, it's the linear extension of $\overline{T}^{-1}$ to the space $\mathcal{R}(T)\dot{+}\mathcal{R}(T)^{\perp}$ (then defining $\mathcal{D}(T^{\dagger})$) with: $\overline{T} := T|_{\mathrm{Ker}(T)^{\perp}} \to \mathcal{R}(T)$. Thus $\mathrm{Ker}(T^{\dagger}) = \mathcal{R}(T)^{\perp}$. An important property is that $T^{\dagger}$ is bounded (or continuous as linear) if and only if the range of T is closed. The Moore-Penrose inverse is closely related to inverse problems and stability topics. An important property is that for $y \in \mathcal{D}(T^{\dagger}$, f is a least-squares solution of 1.6 iff we have:

$$T^*Tf = T^*y \tag{A.1}$$

Equation (A.1) is called the *normal equation* and will play a crucial role in many regularization procedures. A consequence of this is that:

$$T^{\dagger} = (T^*T)^{\dagger}T^* \tag{A.2}$$

Returning to linear compact operators, $T = K$ with non-closed range, 1.6 is ill-posed and the best approximate solution does not depend continuously on y. If $(\lambda_n, \phi_n, \psi_n)$ is the singular system of K (linear compact) we have:

$$K^{\dagger}y = \sum_{i=1}^{\infty} \frac{< y, \phi_n > \psi_n}{\lambda_n}$$

for $y \in \mathcal{D}(K^{\dagger})$ iff:

$$\sum_{n=1}^{\infty} \frac{| < y, \phi_n > |^2}{\lambda_n^2} < +\infty \tag{A.3}$$

Equation (A.3) ensures then that $y \in \mathcal{D}(K^{\dagger})$ and is called the *Picard criterion*, which will be at the heart of the definition of classes of regularity. This criterion ensures that the Fourier coefficients $< y, \phi_n >$ decay sufficiently fast with respect to the sequence of λ_n. This is the first time that we can observe that the relative properties of the operator (through the λ_n) are intrinsically linked to y (through $< y, \phi_n >$) to control stability. The error in y that affects $< y, \phi_n >$ for small eigenvalues is greatly amplified by the $1/\lambda_n$ factor.

A.2.2.2 Order-optimality

At this point, sufficient information is given to understand the aim and the construction of regularization techniques. But an evident problem is the control of the nature of the convergence of the solution of the regularized problem towards the true solution. We could ask oneself whether it is more interesting to control $||f_\alpha - f^\dagger||$ as $\alpha \to 0$ or $||f^\delta_{\alpha(\delta,y^\delta)} - f^\dagger||$ as $\delta \to 0$ where $f^\delta_{\alpha(\delta,y^\delta)} = R_{\alpha(\delta,y^\delta)} y^\delta$. The first one is more tricky to deal with as it will involve the exact data and the regularization operator.

In fact, the quantity $||f^\delta_{\alpha(\delta,y^\delta)} - f^\dagger||$ cannot be bounded by any function of only δ that tends to zero. Thus the convergence of those quantities cannot lead to uniform rates in ill-posed situations. Moreover, we can observe that:

$$||f^\delta_{\alpha(\delta,y^\delta)} - f^\dagger|| \leq ||f^\delta_{\alpha(\delta,y^\delta)} - f_{\alpha(\delta,y^\delta)}|| + ||f_{\alpha(\delta,y^\delta)} - f^\dagger||$$

that shows that the two rates are however closely related. The problem is that as we cannot have access to a uniform bound for the convergence, we face rates of convergence that can be arbitrarily slow. If we want to obtain such bounds or to improve our control the speed of convergence we need more assumptions, or to restrict ourselves to specific subspaces or subsets. The conclusion is that speaking of convergence rate of regularized solutions for ill-posed problems is only possible on subspaces of $\mathcal{D}(T^\dagger)$: defining those subspaces is equivalent as we will see to make a priori assumptions on the exact data or the true solution. Generally, subsets are such as:

$$\{f \in \mathcal{H} \quad | \quad f = Lw, ||w|| \leq \rho\}$$

with L is an operator between Hilbert spaces, with its range in $\mathcal{H}$, which is bounded linear. It is useful to consider $L = (T^*T)^\mu$ for $\mu > 0$. The reasons for this are for the moment rather unclear. This allows to define however sets that are called *source sets*:

$$\mathcal{H}_{\mu,\rho} := \{f \in \mathcal{H} | f = (T^*T)^\mu w, ||w|| \leq \rho\} \quad \text{and} \quad \mathcal{H}_\mu := \bigcup_{\rho > 0} \mathcal{H}_{\mu,\rho}.$$

In fact, $\mathcal{H}_\mu$ is (easily) the range of the operator $(T^+T)^\mu$. If we assume that the true solution of our initial problem is in a set $\mathcal{H}_{\mu,\rho}$, we say that f has a *source representation* and this assumption is called the *source condition*. In fact, in many situations, T is a smoothing operator, specially when T has an expression through an integral equation. In those cases, it is clear that the source condition can be interpreted as a smoothness condition (as for instance the more general translation of f being a function of $\mathcal{C}^k$-class).

When $T = K$ is compact, we can again relate the $\mathcal{H}_\mu$ sets to spectral theory. It can be shown that if K compact, with a corresponding singular system $(\lambda_n, \phi_n, \psi_n)$, if $\mu > 0$, we have:

$$K^\dagger y \in \mathcal{H}_\mu = \mathcal{R}((K^*K)^\mu) \Longleftrightarrow \sum_{n=1}^{\infty} \frac{| < y; \psi_n > |}{\lambda_n^{4\mu+2}} < \infty \tag{A.4}$$

Condition A.4 is the Picard criterion when $\mu = 0$ and imposes a condition on the speed at which the Fourier coefficients decrease. But the greater μ, the stricter is this condition. We see that μ starting from 0 and increasing define a **scale** of conditions. This term is important as we will see it later.
In summary, it is difficult to conclude on the speed of convergence of solutions of ill-posed problems without restricting oneself to specific subsets of solutions, defined via the smoothness of the true solution. Doing this, we get a precise and accurate control on the speed of convergence of the solution (which then depends on this condition). If we assume that the $f^\dagger \in \mathcal{H}^\mu = \mathcal{R}((T^*T)^\mu)$ in the context of an ill-posed problem (T non-closed), then the rate of convergence of the regularization procedure is $\mathcal{O}(\delta^{\frac{2\mu}{1+2\mu}})$.

Definition A.2.1. For an ill-posed problem (i.e. $\mathcal{R}(T)$ non-closed), a regularization $\{R_\alpha\}$ for $T^\dagger$ and α a parameter choice rule for $\mu, \rho > 0$ and $y \in T\mathcal{H}_{\mu,\rho}$.

- (R_α, α) is said to be *optimal* in $\mathcal{H}_{\mu,\rho}$ if $\forall\delta > 0$:

$$\Delta(\delta, \mathcal{H}_{\mu,\rho}, R_\alpha) = \delta^{\frac{2\mu}{1+2\mu}} \rho^{\frac{1}{1+2\mu}}.$$

- (R_α, α) is said to be of *optimal order* in $\mathcal{H}_{\mu,\rho}$ if $\exists c > 0$ such as :

$$\Delta(\delta, \mathcal{H}_{\mu,\rho}, R_\alpha) \leq c\delta^{\frac{2\mu}{1+2\mu}} \rho^{\frac{1}{1+2\mu}} \quad \forall\delta > 0.$$

With μ increasing, the source sets are smaller (stricter condition), and the convergence rates become faster. In fact, we have a more precise result when the parameter choice rule is of the form $\alpha(y^\delta, \tau\delta) = \alpha_\tau(y^\delta, \delta)$ and depends on a bound greater than δ as τ is assumed to be $\tau > 1$. This allows to derive the following theorem:

Theorem A.2.1. *If $\forall\tau > \tau_0 \geq 1$, $(R_\alpha, \alpha))$ is of optimal order in $\mathcal{H}_{\mu,\rho}$ for a given $\mu > 0$ and $\forall\rho > 0$, then all regularization methods (R_α, α_τ) with $\tau > \tau_0 \leq 1$ converges for each $y \in \mathcal{R}(T)$ and are of optimal order for all source set $\mathcal{H}_{\nu,\rho}$ where $0 < \nu \leq \mu$ and $\rho > 0$.*

This is important since it permits to define the *qualification* of a regularization method. The qualification μ_0 of a method is the supremum of the set of parameters μ where the regularization is order-optimal.

A.2.3 Regularization methods

We present briefly the main regularization methods. We do not aim to be exhaustive and refer to Engl et al. (1996) and Kaltenbacher et al. (2008) for a complete treatment.

A.2.3.1 Regularization schemes

An ill-posed inverse problem may be solved by regularization schemes. For a compact, one-to-one operator $K : H_1 \rightarrow H_2$ in the equation $Ka = \mu$, a regularization scheme involves a sequence of operators R^α indexed by a positive real parameter α known as regularization parameter. As approximation of the solution, we then take $a_\alpha = R^\alpha\mu$ which can be computed in a stable way since R^α is assumed to be continuous. The regularization parameter α must be chosen such that, as the noise in the data $\hat{\mu}$ goes to zero, the regularized solution a_α converges to $a^\dagger$. Regularization schemes can be of first or of second order.

The regularization scheme of first order is of the form:

$$R^\alpha\mu = \sum_{j=1}^{\infty} \frac{q(\alpha, \lambda_j)}{\lambda_j} < \mu, \psi_j > \phi_j$$

with:

$$q : \mathbb{R}_+^* \times]0, ||K||] \rightarrow \mathbb{R} \quad , \quad \forall\lambda : |q(\alpha, \lambda)| \leq c(\alpha)\lambda \quad , \quad \lim_{\alpha\rightarrow 0} q(\alpha, \lambda) = 1 \quad , \quad c(\alpha) > 0$$

The regularization scheme of the second order allows to write R^α as the product $R^\alpha = A^\alpha K^*$ with A^α self-adjoint and bounded for a given α and:

$$\forall\lambda : |q(\alpha,\lambda^2)| \leq d(\alpha)\lambda^2 \quad , \quad d(\alpha) > 0 \quad , \quad ||A^\alpha|| \leq d(\alpha)$$

$$A^\alpha\phi = \sum_{j=1}^{\infty} \frac{q(\alpha,\lambda_j^2)}{\lambda_j^2} < \phi, \phi_j > \phi_j$$

A.2.3.2 Spectral cut-off

The equation $Kf = y$, with $f \in \mathcal{E}$ and $y \in \mathcal{F}$, cannot be satisfied exactly if $y \notin \overline{\mathcal{R}(K)}$. Hence, it seems natural to transform y by the orthogonal projection onto $\overline{\mathcal{R}(A)}$. However, when y is affected by noise there is no guarantee that $< y^\delta, \psi_j >$ converges to zero rapidly to satisfy Picard criterion given in Equation A.3.

The idea of the spectral cut-off is to replace the original problem with a projected one, where the projection is on a space with dimension smaller than $dim(\mathcal{F})$. Let P_k be the finite-dimensional orthogonal projector on $span\{\psi_1,\ldots,\psi_k\}$, the spectral cut-off approximation of the problem $Kf = y$ is $Kf = P_k y$, with $f \perp Ker(K)$. This problem has a unique solution that we denote with f_α. We introduce a regularization parameter α defined as a function of k: $\alpha = \alpha(k) := sup\{\tilde{\alpha} \in \mathbb{R}; \lambda_k \geq \sqrt{\tilde{\alpha}}\}$. Hence, the spectral cut-off approximated solution writes:

$$f_\alpha = \sum_{j;\lambda_j \geq \sqrt{\alpha}} \frac{1}{\lambda_j} < y, \psi_j > \phi_j.$$

Spectral cut-off is a first-order regularization scheme and corresponds to set $q(\alpha,\lambda) = \mathbf{1}_{\lambda \geq \sqrt{\alpha}}$, where $\mathbf{1}$ denotes the indicator function.

This technique consists in eliminating the smallest eigenvalues. The sum is reduced to a finite sum up to a sufficient order after which the eigenvalues become too noisy. It seems to be particularly adapted in our framework since as we stressed before, the problem comes from the inversion of eigenvalues of the sample covariance matrix.

Let's precise that in the finite-dimensional case, the convergence stated in the Picard criterion is useless. Consequently, the projected equation $Kf = Py$ always has a solution that writes :

$$f = f_0 + K^\dagger y$$

where $K^\dagger$ denotes the *Moore-Penrose inverse* of K, $f_0 \in Ker(K)$ and spectral cut-off is not necessary.

A.2.3.3 Tikhonov regularization

Tikhonov regularization is a method introduced by Tikhonov in order to remedy the instability of the solution (even the best-approximate solution) of an inverse problem. The idea of this scheme is to control simultaneously the norm of the residual and the norm of the regularized solution (see Tikhonov and Arsenin (1977)). A penalty is added to the norm of the solution to compensate the non-invertibility of T^*T. The solution f_0 is then the solution of:

$$f_0 = argmin_f ||Tf - y^\delta||^2 + \alpha||f||^2 \tag{A.5}$$

where α is the regularization parameter. The right term ensures the regularity and the smoothness of the solution. It's easy to recover a closed form expression for the estimated solution as $m^\alpha : f \mapsto ||Tf - y^\delta||^2 + \alpha||f||^2$ is strictly convex ($\forall \alpha > 0$) and has a Fréchet-derivative equal to:

$$(m_f^\alpha)'\tilde{f} = 2 < Tf - y^\delta, T\tilde{f} > +2\alpha < f, \tilde{f} > .$$

The first order condition $(m_f^\alpha)' = 0$, $\forall \tilde{f} \in \mathcal{H}$ leads to $2 < T^*(y^\delta - Tf), \tilde{f} >= \alpha < f, \tilde{f} >$ which gives finally:

$$\hat{f}_\alpha = (\alpha I + T^*T)^{-1}T^*y^\delta. \tag{A.6}$$

Immediately, with the previous notations, this reads also: $F_\alpha(x) = (\alpha + x)^{-1}$. The spectral expression for the estimated function is thus:

$$\hat{f}_\alpha = \sum_{n=1}^{\infty} \frac{\lambda_n}{\alpha + \lambda_n^2} < y^\delta, \psi_n > \phi_n.$$

Using the singular value decomposition of K, the Tikhonov regularized solution can be rewritten as

$$f_\alpha = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} < \hat{\mu}, \psi_j > \phi_j \tag{A.7}$$

that makes clear that this regularization scheme simply consists in moving the spectrum of K^*K away from 0 by a translation of α. With the previous notation, we get $q(\alpha, \lambda^2) = \frac{\lambda^2}{\alpha+\lambda^2}$, $d(\alpha) = 1/\alpha$ and $A^\alpha = (\alpha I + K^*K)^{-1}$ that shows that Tikhonov regularization is a second order regularization scheme. However, when K is self-adjoint the Tikhonov regularized solution takes the form $f_\alpha = (K + \alpha I)^{-1}\hat{\mu}$ since we do not have to project $\hat{\mu}$, but only to correct for the lack of stability of K^{-1}. In this latter case, the Tikhonov scheme becomes a regularization scheme of first order.

Formula (A.5) is very general and adapts directly for nonlinear operators (see below), whereas A.6 is only suited for linear operators. We have the very important following theorem:

Theorem A.2.2. *Let $\hat{f}_\alpha$ resulting from Equation (A.6) and $y \in \mathcal{R}(T)$ which is observed from a noisy version y^δ such as $||y - y^\delta||_\mathcal{G} \leq \delta$. If the parameter α is chosen as follows:*

1. $\lim_{\delta \to 0} \alpha(\delta) = 0$
2. $\lim_{\delta \to 0} \delta^2/\alpha(\delta) = 0$

we have then $\lim_{\delta \to 0} \hat{f}_{\alpha(\delta)} = T^\dagger y$

This is a very general condition that will be often met under various forms. Technically, the proof of the theorem only requires in that $\delta^2/\alpha(\delta)$ remains bounded with $\delta \to 0$. The qualification of the Tikhonov regularization is $\mu_0 = 1$ (see Section A.2.2.2). This means that the best rate achieved by a Tikhonov regularization is a $\mathcal{O}(\delta^{\frac{2}{3}})$. In a linear ill-posed inverse problem, a faster rate of convergence may not be attained. There is a saturation effect which makes that even under stringent assumptions on the smoothness of the estimated function, the method itself yields a limitation effect on the speed of convergence. Such an observation however concerns *a-priori* parameter choices. For *a-posteriori* parameter choice rules, see Engl et al. (1996) (Chap.5,p.121).

A.2.3.4 Tikhonov with modified penalty

Of course the choice of the penalty may be modified depending on assumptions we may formulate on the solution: smoothness or proximity to a specified function. A first idea could be to replace $||f||$ by $||f - f^*||$ where f^* is a given function. $\hat{f}_\alpha$ becomes easily :

$$\hat{f}_\alpha = (\alpha I + T^*T)^{-1}(T^*y^\delta + \alpha f^*).$$

An other alternative is to modify $||f||$ and to replace if by $||L^\mu f||$ with $\mu > 0$, where L is a differential operator which is more adapted to the problem and takes into account the needed smoothness of a given transform of the solution (e.g. a derivative). It is usual to consider $L^\mu = (T^*T)^{-\mu}$ and then $\hat{f}_\alpha$ becomes:

$$\hat{f}_\alpha = (\alpha(L^\mu)^*(L^\mu) + T^*T)^{-1}(T^*y^\delta).$$

A more general version of the Tikhonov regularization method, called *Generalized Tikhonov*, consists in allowing for a more general specification of the penalty term. The optimization problem is replaced by:

$$F_\alpha(a) = ||Kf - y^\delta||^2 + \alpha G(f) \tag{A.8}$$

where $G : \mathcal{E} \to \mathbb{R}$ is a nonnegative functional. The most common choice for G is

$$G(f) = ||Lf||^2$$

where $L : \mathcal{E} \to \mathcal{F}$ is a linear operator. When $\mathcal{E}$ is an infinite dimensional space, L must be unbounded and it is usually a differential operator We talk about *Tikhonov regularization in the Hilbert scale induced by* L when L is a densely defined, unbounded, self-adjoint and strictly positive operator in an Hilbert space H_1. This choice forces the solution f_α to be smooth. When $\mathcal{E}$ is a finite-dimensional space, the operator L is simply a matrix; in particular, if L is full-column rank $G(f)$ is the generalized norm in the metric induced by the matrix $L'L$. The *generalized Tikhonov regularized solution* f_α takes the form:

$$\begin{aligned} f_\alpha &= (\alpha L^*L + K^*K)^{-1}K^*y^\delta \\ &= \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha \sum_k l_k^2 < \varphi_k, \phi_j >^2} < y^\delta, \psi_j > \phi_j && \text{(A.9)} \\ &= \sum_j \frac{\lambda_j}{\lambda_j^2 + \alpha||L\phi_j||^2} < y^\delta, \psi_j > \phi_j && \text{(A.10)} \end{aligned}$$

where (l_j, φ_j, ρ_j) is the *singular value decomposition* of L and the last equality is true only if $Ker(K) = \{0\}$. The regularized solution depends not only on the eigenvalues of the operator K, but also on the norm of the operator L. When $L = I$ we clearly find (A.7).

When the operator $K : \mathcal{E} \to \mathcal{E}$ in the inverse problem is self-adjoint and $\hat{\mu} \in \mathcal{E}$, we take an unbounded self-adjoint linear operator L and the generalized Tikhonov solution is

$$\begin{aligned} f_\alpha &= (\alpha L + K)^{-1}\hat{\mu} \\ &= \sum_j \frac{1}{\lambda_j^2 + \alpha||L\phi_j||^2} < y^\delta, \phi_j > \phi_j. && \text{(A.11)} \end{aligned}$$

A.2.3.5 Iterated Tikhonov

All the methods that have been presented now are direct in the sense that the solution is computed through a single step. A whole class of alternative methods uses iteration processes to estimate the solution of the problem. This is the principle of iterated Tikhonov regularization, which is briefly presented here. Iterated Tikhonov applies easily to linear operators, whereas the idea of iterative methods will be mainly used in situations where there a nonlinear operator is involved.

To illustrate these procedures, we begin by exploring how the last method adapts to iterative schemes. Suppose that we start with $\hat{f}^0 = 0$ and apply a Tikhonov regularization in the first step:

$$\hat{f}_\alpha^1 = (\alpha I + T^*T)^{-1}(T^*y^\delta + \alpha \hat{f}^0).$$

At a given step $k \in \mathbb{N}^*$ if we have an estimator $\hat{f}_\alpha^k$, the next estimator at step $k+1$, we get:

$$\hat{f}_\alpha^{k+1} = (\alpha I + T^*T)^{-1}(T^*y^\delta + \alpha \hat{f}_\alpha^k).$$

It can be shown that $\hat{f}_\alpha^k$ may be directly estimated as:

$$\hat{f}_\alpha^k = (\alpha I + T^*T)^{-k}(T^*T)^{-1}[(\alpha I + T^*T)^k - \alpha^k I]T^*y^\delta.$$

A.2.3.6 Landweber iteration

Iterative methods consider that the direct expression of the problem is maybe more intuitive than the inverse one, and iteratively update a current estimator $\hat{f}_k$ through $\hat{f}_{k+1} = \hat{f}_k + R_k(\hat{f}_k, y^\delta)$ where R_k depends on the chosen method that will principally use the operator or its dual. The general outline of those methods is to start the process with a given function, and to sequentially update an estimator in order to approach the true solution. They exhibit a self-regularizing effect acting as the stoping criterion of the iteration plays the role of the regularization parameter. The idea is that as f_k is far from the true solution, $y - TF_k$ is huge and this residual is much greater than the data error. The algorithm has then the good descent direction. Iterations are stopped when the residual becomes inferior to the error.

Among iterative methods, Landweber iteration minimizes $||Tf - y||$ using a gradient method. The descent algorithm is along the direction $-T^*(T\hat{f}_k - y)$ at a given step k of iteration. For a given $m > 0$, the iteration procedure is :

$$\hat{f}_{k+1} = \hat{f}_k - mT^*(T\hat{f}_k - y).$$

Recursively, it is easy to obtain:

$$\hat{f}_k = \sum_{i=0}^{k} m(I - mT^*T)^i T^* y.$$

m is a relaxation parameter that satisfies $0 < m < \frac{1}{||T||^2}$. When $||T|| \leq 1$, m is useless and the iteration can be written under the form $\hat{f}_k = \hat{f}_{k-1} - T^*(T\hat{f}_{k-1} - y)$. For a fixed index k, $\hat{f}_k^\delta$ depends continuously on the data, and cannot arbitrarily diverge at a given level of noise. If $||y^\delta - y|| \leq \delta$, we have $||\hat{f}_k - \hat{f}_k^\delta|| \leq \sqrt{k}\delta$.

The iteration is stopped before $\sqrt{k}\delta$ becomes too high (i.e. when k is sufficiently weak, this concept is called semi-convergence). The *discrepancy principle* specifies to stop the iteration for the first index

$k = k(\delta, y^\delta)$ such that : $||y^\delta - T\hat{f}_k^\delta|| \leq \tau\delta$ for a fixed $\tau > 1$. Contrary to Tikhonov, there is no saturation effect. Landweber iteration leads to a huge number of iteration (even if the operator need not to be inverted), and practitioners will prefer other iterative methods.
F_α is then "a F_k" since the regularization parameter is now related to the stopping index of the iteration. Deriving from the Landweber method is the semi-iterative procedures. Rather than using the last iteration, all the previous estimators are mixed. At step k, $\hat{f}_k$ is given by:

$$\hat{f}_k = \sum_{j=0}^{k-1} w_j^{(k)} \hat{f}_j - m_k T^*(T\hat{f}_{k-1} - y)$$

where $\sum_{j=0}^{k-1} w_j^{(k)} = 1$. Here again, the index k plays the role of the regularization parameter. In those cases, F_k is a polynomial of degree $k-1$. As a particular case, the ν-methods use only two previous iterations and $\hat{f}_k$ is given by :

$$\hat{f}_k = w_k \hat{f}_{k-1} + (1 - w_k)\hat{f}_{k-2} - m_k(T\hat{f}_{k-1} - y).$$

A.2.3.7 Nonlinear Landweber

Nonlinear Landweber iteration is the adaptation of Landweber methods for nonlinear problems. Previously for a linear operator T, we had:

$$\hat{f}_{k+1} = \hat{f}_k - T^*(T\hat{f}_k - y)$$

where $T^*(T\hat{f}_k - y)$ was the direction for the gradient taken in $\hat{f}_k$ of $f \mapsto ||y - T\hat{f}_k||^2$. If T is nonlinear and T' is its Fréchet-derivative, this gradient becomes $f \mapsto T'^*_{\hat{f}_k}(y - Tf)$. Consequently, the nonlinear Landweber iteration update is:

$$\hat{f}_{k+1} = \hat{f}_k - T'^*_{\hat{f}_k}(T\hat{f}_k - y).$$

When T is linear, $T = T'$ and the former expression generalizes the approach. To ensure convergence and control on its rate, strong assumptions have to be made on T since at each step the operator $T'_{\hat{f}_k}$ is modified (via $\hat{f}_k$) and it is difficult to constrain the $\{\hat{f}_k\}$ sequence to belong to given smoothness spaces. Kaltenbacher et al. (2008) gives conditions for the convergence, which can only be obtained locally, whether data is exact or noisy.

A.2.3.8 Other methods

Many other methods exist. The Newton-type methods rely on the use of the Fréchet-derivative. The initial equation $Tf = y$ is iteratively linearized around the current estimated solution and the new problem is solved. Provided that $y \approx Tf \approx T\hat{f}_k + T'_{\hat{f}_k}(f - \hat{f}_k)$, the idea is to iterate and find $\hat{f}_{k+1}$ such as $y \approx T\hat{f}_k + T'_{\hat{f}_k}(\hat{f}_{k+1} - \hat{f}_k)$, i.e.

$$T'_{\hat{f}_k}(\hat{f}_{k+1} - \hat{f}_k) = y - T\hat{f}_k. \tag{A.12}$$

We recover here an other linear inverse problem as $T'_{\hat{f}_k}$ is a linear operator. But if T is compact then $T'_{\hat{f}_k}$ is also compact and (A.12) may also be ill-posed. Levenberg-Marquardt method solves A.12 with the conjugate use of Tikhonov regularization. This is a particular choice which specifies the method but any regularization method for linear problems could be chosen alternatively.
Landweber methods can also be extended or modified. Regularizing in Hilbert scales, the adjoint of the Fréchet derivative is replaced by an operator depending on the considered Hilbert scale. In

iteratively regularized Landweber methods, the usual iteration is modified with a term depending on $\hat{f}_k^\delta$, inspired from Newton-methods. This modifies in particular the condition on τ in the stopping criterion. Adaptations are even possible when T' is not available.

Concerning Newton-type methods, iteratively regularized Gauss-Newton methods are the extension of Newton-type methods in the same fashion than for Landweber iteration procedures, and have a lot of generalizations. See Kaltenbacher et al. (2008) for a comprehensive review of those methods, including other kind of procedures such as multilevel, multigrid, level set methods, that are beyond the scope of this presentation.

A.3 Single risk estimation in duration models

We have already expressed the importance of censorship, which can be of two kinds. In our study, we face most of the time a right-censorship[2]. For dead funds, the duration can be observed. For the other ones, their potential exit cannot be observed and then their lifetime duration is truncated. The second kind of censorship is due to the fact that we observe only discrete durations (in months).

A.3.1 Parametric estimation

Suppose that the durations have a continuous support, and are potentially censored. Working with a parametric family for the density f_θ where θ is the parameter to be estimated, we get that a non-censored observation t has a contribution to the likelihood equal to $f_\theta(t)$. However, for a censored observation t, the contribution to the likelihood is equal to:

$$\int_t^\infty f_\theta(u)du.$$

The total likelihood for a set of observations $(\tau_i)_{i\in[1;n]}$ with a censorship indicator $(\delta_i)_{i\in[1;n]}$ (with δ_i equal to 1 if the observation is censored, 0 otherwise) is equal to:

$$L(\tau_1,\dots,\tau_n,\theta)=\prod_{i=1}^n (f_\theta(\tau_i))^{\delta_i}\Big(\int_{\tau_i}^\infty f_\theta(u)du\Big)^{1-\delta_i}. \tag{A.13}$$

Then, a maximum likelihood estimation can be set up to estimate the parameters that fit best the distribution.

The general form of the likelihood is easy to derive in the most general case. We note generically $f(t,x,\theta)$ the density, $\lambda(t,x,\theta)$ the hazard function and $S(t,x,\theta)$ the survivor function, depending on time t, observation of the covariate x and parameters θ (we suppose that x includes the possibility to give information on censorship). The log-likelihood is:

$$\begin{aligned} log(L(\tau_1,\dots,\tau_n,x_1,...,x_n,\theta)) &= \sum_{i\in I_u} log(f(\tau_i,x_i,\theta)) - \sum_{i\in I_c} log(S(\tau_i,x_i,\theta)) \\ &= \sum_{i\in I_u} log(\lambda(\tau_i,x_i,\theta)) - \sum_{i} log(S(\tau_i,x_i,\theta)) \end{aligned} \tag{A.14}$$

where I_c is the set of censored observations, I_u the uncensored ones. The second expression can be obtained thanks to the link between the density, survivor and hazard function, as $f(t)=\lambda(t)S(t)$.

We could have observed durations that are discrete (which is a second kind of censorship). We face this situation since our Hedge Fund lifetimes are estimated as a number of months. A duration τ means that the Hedge Funds lifetimes has failed during the time interval $[\tau;\tau+1[$. We replace in the likelihood of expression (A.13) the contribution of uncensored data $f_\theta(t_i)$, by the quantity $\int_{\tau_i}^{\tau_i+1} f_\theta(u)du$. Then the likelihood to be maximized becomes:

$$L(\tau_1,\dots,\tau_n,\theta)=\prod_{i=1}^n \Big(\int_{\tau_i}^{\tau_i+1} f_\theta(u)du\Big)^{\delta_i}\Big(\int_{\tau_i}^\infty f_\theta(u)du\Big)^{1-\delta_i}.$$

[2]We could also include the problem of left-censored data when we are doubtful on the exact beginning of the lifetime, but such an extension is not done here. On this point see the work of Patilea and Rolin (2006).

If $S_\theta(t) = \int_t^{+\infty} f_\theta(t')dt'$ the likelihood becomes:

$$L_{dis}((t_i)_{i\in[1;n]}, \theta) = \prod_{i=1}^{n} (S_\theta(t_i) - S_\theta(t_i+1))^{\delta_i} (S_\theta(t_i))^{1-\delta_i}$$

A.3.2 Nonparametric estimation

One first estimates $\Lambda(t)$ (or at least its increments) and then one tries to find a smooth expression for its derivative in order to get $\hat{\lambda}(t)$ (see Ramlau-Hansen (1983)). Suppose that we have an estimator $\hat{\Lambda}(t)$ at dates $0 = \tau_0 \leq \tau_1 \leq \ldots \leq \tau_n$. With a gaussian kernel, $\hat{\lambda}(t)$ is estimated through :

$$\hat{\lambda}(t) = \sum_{k=1}^{n} \frac{1}{\sqrt{2\pi}h_n} exp\left(-\frac{(t-\tau_i)^2}{2h_n^2}\right) \times (\Delta\hat{\Lambda}(\tau_i))$$

with h_n an associated bandwidth. We can improve this formula by using boundary kernels: for the terms corresponding to little values of τ_i ($\tau_i \leq h_n$) we make the correction proposed by Li and Racine (2006):

$$\hat{\lambda}(t) = \sum_{k=1}^{n} \frac{1}{\sqrt{2\pi}h_n} exp\left(-\frac{(t-\tau_i)^2}{2h_n^2}\right) \times (\Delta\hat{\Lambda}(\tau_i)) \times \left(\mathbf{1}_{\{\tau_i \geq h_n\}} + \frac{\mathbf{1}_{\{\tau_i < h_n\}}}{1+\Phi(-\frac{\tau_i}{h_n})}\right)$$

with Φ the cumulative function of the standard gaussian distribution. We have not yet defined how to estimate the integrated hazard $\Lambda(t)$. Several equivalent approaches are available. First, the Kaplan-Meier estimator is a product-limit estimator. It is obtained by specifying that being alive at time t is equivalent to being alive just before t and not dying at this date. Its expression is:

$$\hat{\Lambda}_{KM}(t) = -ln(\hat{S}_{KM}(t)) \qquad \text{with} \qquad \hat{S}_{KM}(t) = \prod_{\{j|\tau_j<t\}} \left(1 - \frac{f_j}{a_j}\right)$$

where f_j is the number of individuals failing at date τ_j (thus only uncensored individuals), and a_j is the total number of individuals which have attained at least τ_j, then including censored individuals.

An other possibility is to use the Nelson-Aalen estimator (see Kalbfleisch and Prentice (2002)) which writes:

$$\hat{\Lambda}_{NA}(t) = \sum_{i=1}^{n} \frac{(1-\delta_i)\mathbf{1}_{\{\tau_i \leq t\}}}{\sum_{j=1}^{n} \mathbf{1}_{\{\tau_i \leq \tau_i\}}}$$

where $\delta_i = 1$ if the observation is censored, 0 otherwise.

An other estimator is used in Couderc et al. (2008) which uses Gamma-Ramlau Hansen estimator. Its expression is:

$$\hat{\lambda}_{GRH}(t) = \sum_{i=1}^{n} \frac{\tau_i^{t/h_n} exp(-\tau_i/h_n)}{h_n^{t/(h_n+1)}\Gamma(\frac{t}{h_n}+1)} \times (\Delta\hat{\Lambda}(\tau_i))$$

where h_n is again the smoothing parameter and δ_i the censorship indicator defined as above.

A.4 Hedge Funds: complements

A.4.1 Financial glossary

Alpha: estimated most of the time as the constant term of a linear regression of the historical returns; more generally for a fund, it is often understood as the added value of the manager or as a proxy of her ability.

Assets Under Management (AUM): total size of a fund, determined both by the in-flows of money of the investors and by the market value of its assets.

Backtest(ing): "what-if" analysis of the historical behavior of a fund/trading strategy estimated on the past, whereas the fund/the strategy was not existing at this moment.

Beta: estimated as the coefficient term of a linear regression of the historical returns; more generally for a fund, it refers to its risk exposure towards risk-factors.

Capacity: assumed maximal size of a fund w.r.t. the market it invests in; if the size of the fund is bigger than its capacity, there may be a risk for the fund to be too big to benefit from arbitrages of the current investment strategy.

CAPM: *Capital Asset Pricing Model*; model trying to explain how the risk (variance) of a financial asset may be decomposed between systematic and residual risk. See Sharpe (1964), Fama (1968).

CFTC: *Commodity Futures Trading Commission* is an independent, federal agency in the United States in charge of the regulation of commodity trading.

Directional Strategy: strategy that takes a pure bet on the direction (increase or decrease) of the price evolution of a security, without hedging this position.

Efficient Portfolio: in the portfolio theory, efficient portfolios are defined as being optimal through a mean-variance criterion as for a given return they are sought to have a minimal variance; their set define the efficient frontier.

(Max) Drawdown: (maximal) continuous decrease in value of a stock or a fund; often measured as a percentage of the last local maximum NAV.

Funds of Hedge Funds (FoF or FoHF): funds that invest in several other Hedge Funds; they seek diversification without usual Hedge Fund strategies management; funds of funds of Hedge Funds may also exist; minimum investment required are generally lower.

HFR: HFR (www.hedgefundresearch.com) is a company monitoring the global Hedge Fund industry, selling Hedge Fund databases, industry reports, and indexes (both investable and non-investable).

High-Water Mark: highest value reached by a fund; used to estimate performance and sur-performance fees. For two accounts of classes A (clients) and B (fund) whose values at month t

are A_t and B_t, let's write HMW_t the value of the High-Water Mark. For each t the total investment capacity of the fund is $A_t + B_t$. Let h_t be a predetermined hurdle rate and r_t the return of the fund. The HMW_t value is computed through (see Darolles and Gouriéroux (2009)):

$$HMW_t = \max_{0 \leq t' \leq t} \Big(A_{t'} \prod_{k=t'}^{t} (1 + h_k) \Big).$$

Between two resets of the accounts and with a sur-performance fees equal to f the accounts are recursively defined through:

$$A_{t+1} = A_t(1 + r_{t+1}) - f[A_t(1 + h_{t+1}] - HMW_t]^+$$

$$B_{t+1} = B_t(1 + r_{t+1}) + f[A_t(1 + h_{t+1}] - HMW_t]^+.$$

Hurdle Rate: reference return above which (over-)performance fees are estimated; may be equal to 0, the LIBOR rate, etc.; generally predetermined and known in advance.

Leverage: ratio of debt used to expand an initial capital in order to raise the anticipated gains (or losses) of a strategy; a leverage of 3 means that two thirds of the current portfolio are borrowed.

Lockup: period during which an investor cannot redeem her shares or withdraw her capital, just after initial investment.

Long Position: buying a stock or a security (the investor expects an increase of its price).

Loss Carry Forward: fees collection scheme that is an alternative to the High-Water Mark; at month t, $LCF_t \leq 0$ represents a cumulative negative performance, which is different from 0 when returns are insufficient to compensate past losses. LCF_t is computed through (see Darolles and Gouriéroux (2009)):

$$LCF_{t+1} = -[LCF_t + A_t(r_{t+1} - h_t)]^-$$

Between two resets of the accounts and with a sur-performance fees equal to f the accounts are recursively defined through:

$$A_{t+1} = A_t(1 + r_{t+1}) - fA_t[r_{t+1} - (A_t h_t - LCF_t)/A_t]^+$$

$$B_{t+1} = B_t(1 + r_{t+1}) + fA_t[r_{t+1} - (A_t h_t - LCF_t)/A_t]^+.$$

NAV: *Net Asset Value*; this is the reference value of a share of fund calculated with the market value of its assets, but after fees calculation.

NFA: *National Futures Association*; organization in the United States in charge of the regulation for investors wanting to register as future traders.

Private Equity Funds: funds that invest in non-listed businesses, by opposite to stocks or liquid financial products; they may invest in start-ups, new companies, non-listed companies needing financing, environmental or pioneering firms, distressed companies, etc.

Real Estate Funds: funds that invest in land, buildings, and real property of this kind.

Provision Account: fees collection scheme, different from High-Water Mark and Loss Carry Forward; uses a third account P such as the way the class B account is feeded is remaiend unchanged. At each month t the value of the P account is noted $P_t \geq 0$. P_t is such that $LCF_t + P_t$ equals at each date P_t or LCF_t, the other being equal to 0. If p is the fraction of investment allocated to the provision account we have then:

$$LCF_{t+1} = -[LCF_t + P_t + p(A_t + P_t)(r_{t+1} - h_t)]^-$$
$$P_{t+1} = -[LCF_t + P_t + p(A_t + P_t)(r_{t+1} - h_t)]^+$$
$$A_{t+1} = (A_t + P_t)(1 + r_{t+1}) - P_{t+1}$$
$$B_{t+1} = B_t(1 + r_{t+1}).$$

See Darolles and Gouriéroux (2009).

SEC: *Security Exchange Commission*; this is the federal office in the United States in charge of the regulation and control of the financial markets.

Short Position: selling a stock or a security (the investor expects a decline of its price; the security is generally not owned and must be borrowed under specific conditions.

TASS: Lipper TASS is the name of the commercial database set up by Lipper (www.lipperweb.com) for qualified investors interested in monitoring the global Hedge Fund industry.

Top-Down Approach: starting from macro-economic forecasts or previsions, consists in specifying markets, countries or sectors for investment; converse approach is *bottom-up*, where the focus is made on stock-picking, with an individual analysis of securities.

Turnover: measures for a fund or a portfolio of assets the frequency or the proportion of the portfolio that is changed in average at each redefinition of the composition of the portfolio/fund.

UCITS norms: *Undertakings for Collective Investment in Transferable Securities* ; set of norms aiming at harmonizing and regulating European markets and funds; defined first in 1985; their first goal was to simplify the process for a fund aiming to be promoted anywhere in Europe; UCITS IV norms are currently negotiated.

A.4.2 Hedge Funds strategies

We adopt here the strategy classification and description of HFR available in HFR (2009). We focus on the categorization of single funds. We give here the main strategies with some examples of refinements. Such a classification may be difficult as sometimes overlaps exist (between *Long-Short* and *Relative Value* strategies for example)

Equity Hedge-Long/Short Equity

ex: Equity Market Neutral; Fundamental Value; Fundamental Growth; Long Bias; Short Bias;

This constitutes the most frequent kind of Hedge Fund strategy. *Equity Hedge* or *Long/Short Equity* funds share with Alfred Winslow Jones' first Hedge Fund the same philosophy. They identify under-priced stocks that they buy, and over-priced stocks that they try to short-sell. Diversification is sought by investing in several economies, sectors or markets. Use of beta and of leverage helps to immunize the fund w.r.t. to market movements.

Objectives in terms of return, volatility, and market sensibility, depend deeply on the fund. *Long-Only Absolute Return* funds focus on the buy-side of the investment. *Fundamental Value* refers to funds seeking individual quality of stocks (bottom-up approach) and *Fundamental Growth* is linked with a top-down approach, the manager trying first to identify sectors or economies of interest. *Long Bias* strategy is evidently focusing on identifying under-priced securities, with stocks and has more long than short positions in a portfolio. *Short Bias* is the opposite.

To be more precise, there can be slight distinctions between strategies, even if the underlying traded securities are identical. As explained by Khandani and Lo (2007), Long-Short Equity funds may use techniques that are quantitative or not, where technology may not have an important role to play. Conversely, funds proceeding to *Statistical Arbitrage* use highly technical, heavy computational, short-term mean-reverting strategies, with a large number of assets. Finally *Quantitative Equity Market Neutral funds* seek a lower turnover, with less frequent trades, again using quantitative models but with e.g. economic indicators. This shows that those categories only remain a tool to help for categorization.

Macro

ex: Commodity; Currency; Systematic Diversified;

Macro Funds follow generally a top-down approach, seeking to take benefit from global macro-economic changes, movements of the economies or government decisions. Such funds generally operate on various markets (bonds, equities, commodities, credit, etc.) and take mainly directional bets.

Relative-Value

ex: Fixed Income; Convertible Arbitrage; Volatility;

Relative Value is a very vague category that includes all kind of sub-strategies whose objective is to annihilate market risk using leverage, pairs, or various kinds of assets. Some *Long-Short Equity* strategies could then be considered as *Relative Value* strategies. *Fixed Income* strategies are concentrated in investing in fixed income instruments (bonds, swaps, interest rates, etc.), seeking arbitrages with the use of long and short positions on those securities. *Convertible Arbitrage* funds buy convertible securities while hedging their positions on corresponding underlyings (equity, fixed-income).

Event-Driven

ex: Credit Arbitrage; Distressed Debt; Merger Arbitrage; Special Situations;

Merger Arbitrage takes place following merging announcements as funds may purchase stocks of the target company and short the stocks of the acquiring company. *Event Driven* strategies try to take benefit from special situations in the life of companies: merging, bankruptcies, hostile takeovers, rating degradation, public offerings, etc. They apply to various markets and assets and are usually less concerned by market movements. *Distressed Debt* strategies classically focus on distressed companies that are e.g. on the verge of bankruptcy. Their debt is therefore under-priced and such strategies seek to benefit from it. *Merger Arbitrage* strategies exploit the co-movements of the stocks prices of companies that will nearly merge.

Managed Futures/CTA

As said before, they constitute a particular case. CTA (*Commodity Trading Advisors*) are regulated by the CFTC and the NFA in the US and invest in options, commodity futures, foreign exchange securities. Usually in the databases, they are separated from other single funds.

For each category (*Equity Hedge*, *Relative Value*, etc.) *Multi-Strategy Funds* may be found. Finally, some of those strategies can focus on some economies or specific markets. The categorization is usually the following. *America* represents North, Latin and Pan-American markets. *Europe* refers to UK, Western, Northern, Eastern and Pan-European markets, but also to Russia. *Asia* recovers mainly Japan, Hong-Kong and Singapore, but the main distinction is between Asia with or without Japan. Finally, *Emerging markets* is used to refer African, Latin, Middle-East, Russian and Asian (ex-Japan) markets.