# Endogeneity And Instrumental Variables In Dynamic Models

Jean-Pierre Florens[*], Guillaume Simon[†]

10th December 2009

**THIS DRAFT IS A VERY PRELIMINARY VERSION**
Comments Welcome

**Abstract**

The objective of the paper 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 decompositions. Then we estimate 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.

**Keywords**: Endogeneity; Instrumental Variables; Dynamic Models; Duration Models.

[*] IDEI and GREMAQ, Université de Toulouse 1, 21 allée de Brienne, 31000 Toulouse, France.
[†] Université de Toulouse, and LYXOR AM. Email: `guillaume.simon@ensae.org`.

# 1 INTRODUCTION

## 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 $\mathbb{E}[U|Z]$ characterizes $\phi$ as the conditional expectation and if $Y = \phi(Z, U)$ with $\phi$ monotonous in $U$, $U$ uniform, the condition that $Z$ and $U$ are independent characterizes $\phi$ as the 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 case $Y = \phi(Z) + U$ the assumption becomes $\mathbb{E}[U|W] = 0$ or in the nonseparable model it is assumed that $U \perp W$ (see for a recent literature [Florens, 2002], [Newey and Powell, 2003], [Hall and Horowitz] or for non separable case [Horowitz and Lee]). 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 and in a nonparametric framework has an impact on the speed of convergence of the estimators.

The objective of this paper 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 condition may be discussed. This paper is not concerned by statistical inference but it 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 [Carrasco and an Eric Renault, 2003]) 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 where the structure of the process generating $Y$ is given (counting processes or diffusions for instance). 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 .

## 1.2 Mathematical framework

In this paper 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 $\sigma$-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 $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$).

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:

$$X_t = X_0 + H_t + M_t; \quad (1)$$

- $M_t$ 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 paper. We also simplify the expressions by always assuming $X_0 = 0$. Extension to local martingales and to cases where $X_0 \neq 0$ is straightforward. Let us note that the decomposition 1 is a.s. unique. These concepts are fundamental in the theory of stochastic processes (see in particular [Dellacherie and Meyer (1980)] - Vol II - Chap VII).

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

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

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

$$dX_t = h_t dt + dM_t$$

where $H_t = \int_0^t h_s ds$.

Two particular cases will be analyzed in details. First the case where $X_t$ is a counting process and $h_t$ its stochastic intensity (see e.g. [Karr, 1991], [Andersen, Borgan, Gill, Kieding]). Second, we will explore the situation where $X_t$ is a diffusion process where $dM_t = \sigma_t dB_t$ with $B_t$ a Brownian motion (see e.g. [Gard (1988)]), and $h_t$ and $\sigma_t$ are the drift and the volatility.

# 2 THE ADDITIVELY SEPARABLE CASE

## 2.1 The framework

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

- $i)$ $\mathcal{Y}_t$, $\mathcal{Z}_t$, $\mathcal{W}_t$ are the filtrations generated by each subprocess;
- $ii)$ We call the *endogenous filtration* the filtration generated by $\mathcal{Y}_t$ and $\mathcal{Z}_t$ and the *instrumental filtration* the filtration $\mathcal{Y}_t$v$\mathcal{W}_t$ generated by $\mathcal{Y}_t$ and $\mathcal{W}_t$.

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

**Definition 2.1.** The process $Y_t$ has a Doob-Meyer Instrumental Variable (DMIV) decomposition if:

$$Y_t = \Lambda_t + E_t$$

where:

1. $\Lambda_t$ is $\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t$ predictable ;
2. $\mathbb{E}[E_t - E_s | \mathcal{Y}_t \mathrm{v} \mathcal{W}_t] = 0$ for $0 \leq s < t$.

Equivalently we may say that $Y_t$ is an IV semi-martingale w.r.t. $(\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t)_t$ and $(\mathcal{Y}_t \mathrm{v} \mathcal{W}_t)_t$. First we can note that if $\mathcal{W}_t = \mathcal{Z}_t$ this definition reduces to the usual decomposition definition. If the filtration $\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t$ is included into $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$ 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 $\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t$ and $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$ has no inclusion relation. Moreover, the two filtrations do not need to be generated by processes and $\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t$ and $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$, may be replaced by more general filtrations $\mathcal{F}_t$ and $\mathcal{G}_t$ under the condition that $Y_t$ is adapted to each of them.

Assumption $i)$ means that the predictable process "only depends" on the past of $Y_t$ and on $Z_t$ and its past. Assumption $ii)$ is the independence condition between the "noise" $E_t$ and the instruments $W_t$. Equality in $ii)$ 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 $E_t$ is not assumed to be adapted to $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$. The usual decomposition is unique a.s. but it should be noted that this unicity result is not true in general: this will be precisely the object of the identification condition analyzed below.

## 2.2 Identification

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

**Theorem 2.1.** *Let us assume that $Y_t$ is a special semi-martingale w.r.t. $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$ and that :*

$$dY_t = h_t dt + dM_t$$

*where $H_t = \int_0^t h_s ds$ is $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$-predictable and $M_t$ is a $\mathcal{Y}_t \mathrm{v} \mathcal{W}_t$-martingale. If the following family of integral equations:*

$$h_t = \mathbb{E}[\lambda_t | \mathcal{Y}_t \mathrm{v} \mathcal{W}_t] \quad t \geq 0 \tag{2}$$

$$\lambda_t \text{ is } \mathcal{Y}_t \mathrm{v} \mathcal{Z}_t\text{-measurable and integrable}$$

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

Roughly speaking equation (2) means that we have to solve:

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

This expression has mathematically no sense 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 conditional expectation operator. If we restrict our attention to square integrable variables, this operator is defined on $L^2(\mathcal{Y}_t \mathrm{v} \mathcal{W}_t)$. Note that the conditional expectation operator is compact under minor regularity conditions. Its range is then a strict subspace of $L^2(\mathcal{Y}_t \mathrm{v} \mathcal{W}_t)$ 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 $I\!E[\quad|\mathcal{Y}_t \mathrm{v} \mathcal{W}_t]$ defined on $L^2(\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t)$ are identifiable and then the DMIV decomposition is unique ($\Lambda_t$ is identifiable) if and only if the conditional expectation operator is one-to-one. The following concept extends the full known case of static models.

**Definition 2.2.** The filtration $(\mathcal{Y}_t \mathrm{v} \mathcal{Z}_t)_t$ is strongly identified by the filtration $(\mathcal{Y}_t \mathrm{v} \mathcal{W}_t)_t$ (or $Z_t$ is strongly identified by $\mathcal{W}_t$ given $\mathcal{Y}_t$) if and only if :

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

For a good treatment of conditional strong identification, see [Florens et al (1990)] - Chap 5 and for relation with the completeness concept in statistics. 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 $\Lambda_t$ is identified. We want to illustrate this concept in two examples : discrete-time models and diffusions.

## 2.3 Example 1 : discrete time models

Let's consider three stochastic processes $Y_t$, $Z_t$ and $W_t$ and let's assume that :

$$Z_t = \sigma\{Y_0^t, Z_t^t\} \quad \text{and} \quad W_t = \sigma\{Y_0^t, W_0^{t+1}\}$$

when e.g. $Y_0^t = (Y_0, \dots, Y_t)$. In that case the decomposition of $Y_t$ w.r.t. $Z_t$ is characterized by:

$$Y_t = \sum_{j=1}^{t} I\!E[Y_j - Y_{j-1}|Y_0^{j-1}, Z_0^j] + M_t$$

since :

$$Y_t - Y_{t-1} = I\!E[Y_t - Y_{t-1}|Y_0^{t-1}, Z_0^t] + M_t - M_{t-1} = h_t + (M_t - M_{t-1}).$$

Under our assumptions on $Z_t$, $\lambda_t$ is a function of $Y_0^{t-1}$ and $Z_0^t$ and the central equation becomes :

$$h_t = I\!E[\lambda_t|Y_0^{t-1}, W_0^t] \quad \forall t$$

or equivalently:

$$\begin{aligned} h_t(Y_{t-1}, Y_{t-2}, W_t, \dots) \quad = \quad & \int \lambda_t(Y_{t-1}, Y_{t-2}, \dots, Z_t, Z_{t-1}, \dots) \\ & f_t(Z_t, Z_{t-1}, \dots |W_t, W_{t-1}, \dots, Y_{t-1}, \dots) dW_t dW_{t-1} \dots \end{aligned} \tag{3}$$

Those examples show that the estimation of $\lambda$ usually requires some conditional independence assumptions for practical computation. A first step may be to reduce the number of instruments by assuming e.g.:

$$Z_t \perp (W_0^t, Y_0^{t-1}) | W_{t-k}^t, Y_{t-k}^{t-1}.$$

Using a stationary assumption the right hand side of 3 becomes :

$$\int \lambda(Y_{t-1}, Z_t) f(Z_t | W_{t-k}^t, Y_{t-k}^t) dZ_t.$$

Note that :

$$h_t = I\!E[Y_t - Y_{t-1} | Y_0^{t-1}, W_0^t].$$

Then equation (3) becomes :

$$r := I\!E[Y_t - Y_{t-1} | Y_{t-k}^{t-1}, W_{t-k}^t] = \int \lambda(Y_{t-1}, Z_t) f(Z_t | W_{t-k}^t, Y_{t-k}^t) dZ_t$$

For any $t$, $r$ and $f$ may be estimated nonparametrically and the problem reduces to the same problem considered in [Darolles et al., 2008].

### 2.4 Example 2 : Diffusions

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

$$dY_t = \lambda_t(Y_t, Z_t)dt + \sigma_t(Y_t)dB_t \tag{4}$$

where $B_t$ is a Brownian motion. This means that if $Z_t$ is fixed (and not generated by the distribution mechanism) $Y_t$ follows a diffusion process with a drift equal to $\lambda_t$ 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:

$$I\!E[dB_t | \mathcal{Y}_t \text{v} \mathcal{W}_t] = 0$$

In that case equation (4) characterizes the DMIV decomposition of $Y_t$. In order to identify $\lambda_t$ we need to construct the decomposition of $Y_t$ w.r.t. the filtration $\mathcal{Y}_t \text{v} \mathcal{W}_t$ ($dY_t = h_t dt + dM_t$) and to solve:

$$h_t = I\!E[\lambda_t | \mathcal{Y}_t \text{v} \mathcal{W}_t] \tag{5}$$

Note that the "reduced form" model $dY_t = h_t dt + dM_t$ has no reason to be a diffusion. Conditionally to $\mathcal{W}_t$ the process may be non Markovian and $M_t$ maybe different from a Brownian motion.

## 3 THE NON-SEPARABLE CASE

We present here the general framework for the obtention of an integral equation for solving endogeneity in dynamic models. We will give detailed applications in some cases of interest : diffusions, duration models and counting processes. All technical details are available in [Protter, 2003].

### 3.1 General theory

We give here the intuition of the theorem under its most general form. Let's suppose that we observe three processes: $X_t$, $Z_t$, and $W_t$. The relative and natural filtration relative to the observation of those processes is $\mathcal{X}_t \mathrm{v} \mathcal{Z}_t \mathrm{v} \mathcal{W}_t$. Suppose that $Y_t$ is a function of $X_t$ (the nature of this function will be precised in the applications).

Fundamentally our object of interest will be a function $\phi$ of time $t$ and variables $Z$ which will be conceived as the inverse (as a function of time) of a counterfactual version of the compensator of $X$ without endogeneity. Of course, $\phi$ is not the inverse of the usual compensator of $X$ w.r.t. to the natural filtration of $Z$.

Practically, let $(\phi_s(Z))_s$ be an increasing sequence of $Z_s$ finite stopping times. Let $U$ be a process, which will play the role of a perturbation noise. For ease of presentation, but without loss of generality we suppose that all the processes are equal to 0 at the initial date. We assume that:

**Assumption 3.1.** $Y_{\phi_t(Z)} = U_t$

**Assumption 3.2.** $(U_t)_t \perp (W_t)_t$

We suppose that every process admits a Doob-Meyer decomposition towards their canonical filtrations. Moreover $U_t$ writes: $U_t = H_t^U + M_t^U$ as we suppose that it's a semi-martingale w.r.t. its canonical filtration with $M_t^U$ a local martingale, and $H_t^U$ is known. Moreover, we make the following regularity assumptions :

**Assumption 3.3.** $Y$ is a semi-martingale w.r.t. $\mathcal{X}_t \mathrm{v} \mathcal{Z}_t \mathrm{v} \mathcal{W}_t$ and its finite variation process is differentiable:

$$Y_t = H^Y(t|\mathcal{X}_t \mathrm{v} \mathcal{Z}_t \mathrm{v} \mathcal{W}_t) + M_t^Y.$$

There exists a process $h^Y(t)$ such as:

$$H^Y(t|\mathcal{X}_t \mathrm{v} \mathcal{Z}_t \mathrm{v} \mathcal{W}_t) = \int_0^t h_Y(s)ds.$$

**Assumption 3.4.** There exists a sequence of function $\zeta_t$:$(Z_s)_{0 \leq s \leq t} \mapsto I\!R^k$ (with $k$ fixed) such as $h^Y(t)$ $\mathcal{X}_t \mathrm{v} \mathcal{Z}_t \mathrm{v} \mathcal{W}_t$-measurable is also $\mathcal{X}_t \mathrm{v} \zeta(Z)_t \mathrm{v} \mathcal{W}_t$-measurable. There exists a density of $\xi_t$ conditionally on $X$ and $W$: $g_t(\xi|\mathcal{X}_t \mathrm{v} \mathcal{W}_t)$.

As $\phi_t(Z)$ is a sequence of $Z_t$ stopping times, $\mathcal{Z}_{\phi_t(Z)}$ is clearly defined as the stopping time $\sigma$-algebra generated by:

$$\mathcal{Z}_{\phi_t(Z)} = \{\Lambda \in Z_\infty | \Lambda \cap \{\phi_t(Z) \leq s\} \in \mathcal{Z}_s \quad \forall s \geq 0\}.$$

The definition of such sets for $Y$ and $W$ is not so obvious as $phi_t(Z)$ is not compulsory a $\mathcal{Y}_t$ or a $\mathcal{W}_t$ stopping time. We will adopt the same notation, however $\mathcal{Y}_{\phi_t(Z)}$ and $\mathcal{W}_{\phi_t(Z)}$ will be defined as the smallest $\sigma$-algebra that make respectively $Y_{\phi_t(Z)}$ and $W_{\phi_t(Z)}$ measurable, that is to say:

$$\mathcal{Y}_{\phi_t(Z)} = \sigma\{Y_{\phi_t(Z(w))}(w)|s \leq t, w \in \Omega\}$$

$$\mathcal{W}_{\phi_t(Z)} = \sigma\{W_{\phi_t(Z(w))}(w)|s \leq t, w \in \Omega\}$$

as for any $w \in \Omega$, $\phi_t(Z(w))$ and therefore $\mathcal{Y}_{\phi_t(Z)}$ and $\mathcal{W}_{\phi_t(Z)}$ are always defined. Moreover, we add that as soon as $Y$ and $W$ are supposed to be càdlàg processes, $Y_{\phi_t(Z)}$ and $W_{\phi_t(Z)}$ are $\phi_t(Z)$-adapted. In the following, we will note $A$v$B$ for two $\sigma$-algebras $A$ and $B$ the smallest $\sigma$-algebra containing $A$ and $B$.

**Lemma 3.1.** $(\mathcal{Y}v\mathcal{Z}v\mathcal{W})_{\phi_t(Z)} = \mathcal{Y}_{\phi_t(Z)}v\mathcal{Z}_{\phi_t(Z)}v\mathcal{W}_{\phi_t(Z)} = \mathcal{U}_t v\mathcal{Z}_{\phi_t(Z)}v\mathcal{W}_{\phi_t(Z)}$

**Lemma 3.2.** *Let $Y_t$ be a process with stochastic intensity $h_t^Y$, adapted to the filtration $\mathcal{A}_t^Y$ (containing internal history of $Y_t$). Let $\phi(t)$ be a monotonous function of time, sufficiently smooth. Then if we define $N_t$ to be $N_t = Y_{\phi(t)}$ then $N_t$ has also a stochastic intensity $\lambda^N$ for the filtration $\mathcal{A}_t^N = \mathcal{A}_{\phi(t)}^X$ which is given by:*

$$h_t^N = \phi'(t)h_{\phi(t)}^X$$

We then write:

$$\phi_t'(Z)h^Y(\phi_t(Z)|\mathcal{Y}_{\phi_t(Z)}\text{v}\mathcal{W}_{\phi_t(Z)}\text{v}\mathcal{Z}_{\phi_t(Z)}) = h^U(t|\mathcal{U}_t\text{v}\mathcal{W}_{\phi_t(Z)}\text{v}\mathcal{Z}_{\phi_t(Z)}). \tag{6}$$

Moreover :

$$I\!\!E[h^U(t|U_t\text{v}\mathcal{W}_{\phi_t(Z)})|U_t\text{v}\mathcal{W}_{\phi_t(Z)} = I\!\!E[h^U(t|U_t\text{v}\mathcal{W}_{\phi_t(Z)})|U_t] = h^U(t)$$

as we assumed that $\mathcal{N}_\infty \perp \mathcal{W}_\infty$ Depending on the process $U$ this expression will become explicit and generally speaking: $\int_0^t h^U(s)ds = H^U(t)$. Then we can also rewrite:

$$I\!\!E[h^U(t|\mathcal{U}_t\text{v}\mathcal{W}_{\phi_t(Z)})|\mathcal{U}_t\text{v}\mathcal{W}_{\phi_t(Z)}] = \int_{I\!\!R^k} h^U(t|\mathcal{U}_t\text{v}\xi(Z)_{\phi_t(Z)}\text{v}\mathcal{W}_{\phi_t(Z)})g_{\phi_t(Z)}(\xi|\mathcal{U}_t\text{v}\mathcal{W}_{\phi_t(Z)})d\xi$$

and we have then :

$$\int_0^{t'} dt \int_{I\!\!R^k} h^U(t|\mathcal{U}_t\text{v}\xi(Z)_{\phi_t(Z)}\text{v}\mathcal{W}_{\phi_t(Z)})g_{\phi_t(Z)}(\xi|\mathcal{U}_t\text{v}\mathcal{W}_{\phi_t(Z)})d\xi = H^U(t') \tag{7}$$

but we can replace 6 in 7 and then:

$$\int_0^{t'} dt \int_{I\!\!R^k} \phi_t'(Z)h^Y(\phi_t(Z)|\mathcal{Y}_{\phi_t(Z)}\text{v}\xi(Z)_{\phi_t(Z)}\text{v}\mathcal{W}_{\phi_t(Z)})g_{\phi_t(Z)}(\xi|\mathcal{X}_{\phi_t(Z)}\text{v}\mathcal{W}_{\phi_t(Z)})d\xi = H^U(t')$$

If we switch the integrals and make the change of variables $u = \phi_t(Z)$ we get :

**Theorem 3.1.**

$$\int_{I\!\!R^k} d\xi \int_0^{\phi_t(Z)} h^Y(u|\mathcal{X}_u v\xi(Z)_u v\mathcal{W}_u)g_s(\xi|\mathcal{X}_u v\mathcal{W}_u)du = H^U(t)$$

### 3.2 Diffusions

Before going further, we need to define the concept of quadratic variation of a semi-martingale.

**Definition 3.1.** Let $Y$ be a semi-martingale. The quadratic variation of $Y$, denoted $[Y,Y] = ([Y,Y]_t)_{t\geq 0}$ is defined by:

$$[Y,Y] = Y^2 - 2\int Y_- dY.$$

$[Y,Y]$ is a process which is càdlàg, increasing, adapted, such as is $T$ is a stopping-time, then:

$$[Y, Y_T] = [Y_T, Y] = [Y_T, Y_T] = [Y,Y]_T.$$

Let's recall that a process $(Y_s)_{s\geq 0}$ is of finite variation on $[0;t]$ as soon as $\sup_{t_i}\sum_i |Y_{t_{i+1}} - Y_{t_i}| < \infty$ for each subdivision of $[0;t]$. A process $(Y_t)_{t\geq 0}$ is said to be of finite variation if it is of finite variation on $[0;t]$ for every $t$. It is then equal to the difference of two increasing processes (and reciprocally).

#### 3.2.1 Driftless processes

**Theorem 3.2.** ***Levy's theorem*** *- A stochastic process $(X_t)_{0\leq t}$ is a Brownian motion if and only if it is a continuous local martingale with $[X,X]_t = t$.*

**Proposition 3.1.** Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ a non-decreasing and continuous function. Then there exists a continuous martingale $M$ such that $[M,M]_t = f(t)$.

*We have that for a $X$, $B_t = X_{[X,X]_t}$ is a Brownian motion.*

The compensator of the square of the process is equal to the quadratic variation. Then let's consider the process $Y_t = f(X_t) = X_t^2$ and apply what has been exposed in the previous section with $\phi_t^{-1}(Z) = [X,X]_t$. For this class of process, we have obviously that $U$ is a Brownian motion, which is univocally determined by the fact that $[U,U]_t = H^U(t) = t$. Consequently, the theorem becomes:

$$\int_{\mathbb{R}^k} d\xi \int_0^{\phi_t(Z)} h^{X^2}(u|\mathcal{X}_u \text{v} \xi(Z)_u \text{v} \mathcal{W}_u) g_s(\xi|\mathcal{X}_u \text{v} \mathcal{W}_u) du = t. \tag{8}$$

#### 3.2.2 General case

In the general case, processes with a drift can be handled by adding the approach of section 3.2.1 a second equation to treat the drift term. Let's suppose that $X$ is a diffusion and that there exists a $\mathcal{X}_u \text{v} \mathcal{W}_u$-finite variation process $D$ such that:

$$\mathbb{E}[X_t - D_t | \mathcal{X}_t \text{v} \mathcal{W}_t] = 0$$

Equation 8 is fundamentally unchanged, except that $X_t$ is replaced by $(X-D)_t$:

$$\int_{\mathbb{R}^k} d\xi \int_0^{\phi_t(Z)} h^{(X-D)^2}(u|\mathcal{X}_u \text{v} \xi(Z)_u \text{v} \mathcal{W}_u) g_s(\xi|\mathcal{X}_u \text{v} \mathcal{W}_u) du = t.$$

## 3.3 Durations

### 3.3.1 Definitions and generalities

A duration is the length $\tau$ of a time-period, spent by an observed individual in a given state. We can make for simplicity the assumption that $I\!P(\tau = \infty) = 0$. The distribution function $F$ of the duration $\tau$ is defined as:

$$F(t) = I\!P(\tau \leq t) \quad \text{for} \quad t \geq 0.$$

The survivor function $S$ of the duration $\tau$ is defined as:

$$S(t) = I\!P(\tau \geq t) = 1 - F(t) + I\!P(T = t) \quad \text{for} \quad t \geq 0.$$

The density of $\tau$ is a function $f : I\!R^+ \mapsto I\!R^+$ which verifies:

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

The integrated (or cumulative) hazard function is of the duration variable $\tau$ is the function $\Lambda$ such that:

$$\begin{aligned} \Lambda \quad : \quad & I\!R^+ \mapsto I\!R^+ \\ & t \rightarrow \Lambda(t) = \int_0^t \frac{dF(u)}{S(u)}. \end{aligned}$$

$\Lambda$ is left-continuous, monotone (increasing), and such that $\Lambda(0) = 0, \Lambda(\infty) = \infty$. For a duration variable $\tau$ whose density is $f$, we have:

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

The hazard function $\lambda$ of the duration $\tau$ is defined through:

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

The following equations are then straightforward:

$$S(t) = \exp(-\Lambda(t)) \quad ; \quad \Lambda(t) = \int_0^t \lambda(u)du \quad ; \quad f(t) = \lambda(t)\exp(-\int_0^t \lambda(u)du).$$

When $\lambda(t) = \lambda > 0 \quad \forall t \in I\!R$, we are in the particular case of a Poisson process. Given an increasing sequence of durations $(\tau_i)_i$, the related univariate counting process $N_t$ is such that $N_t = \sum_{i\geq 1} \mathbf{1}(\tau_i \leq t)$. The trajectories of such processes start in 0, are increasing right-continuous, left-limited, with jumps of size one.

A duration model can be linked with counting processes, as we can define $N_t = \mathbf{1}(\tau \geq t)$. The internal history of the process $N_t$ is $\mathcal{F}_t^N = \sigma\{N_s | 0 \leq s \leq t\}$ potentially augmented with the null sets. A counting process $N_t$, in particular as a $\mathcal{F}_t^N$-adapted point process, is a sub-martingale : using the Doob-Meyer decomposition we have always that $N_t = \Lambda_t + M_t$ where $M_t$ is a martingale, and $\Lambda_t$ is the compensator process.

As soon as the compensator exists, and is absolutely continuous relatively to the Lebesgue measure, the intensity process exists. We have the following definition (see [Karr, 1991]] ):

**Definition 3.2. [Stochastic intensity]**
If $N_t$ is a point process, $\mathcal{F}_t$-adapted, and $\lambda_t$ a $\mathcal{F}_t$-progressive process, positive, such as:

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

is a martingale. Then $\lambda_u$ is the stochastic intensity of the process. When it exists, the intensity is unique.

We have also in [Brémaud, 1991] the following characterization:

**Definition 3.3. [Stochastic Intensity (characterization)]**
If $N_t$ is a point process, $\mathcal{F}_t$-adapted, and $\lambda_t$ a $\mathcal{F}_t$-progressive process, positive, such as:

$$\forall t \geq 0 \quad \int_0^t \lambda_s ds < \infty \quad I\!\!P - a.s.$$

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

$$I\!\!E\Big[\int_0^\infty \phi_s dN_s\Big] = I\!\!E\Big[\int_0^\infty \phi_s \lambda_s ds\Big]$$

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

An other consequence is that $\forall s, t, 0 \leq s \leq t$:

$$I\!\!E[N_t - N_s|\mathcal{F}_s] = I\!\!E\Big[\int_s^t \lambda_u du|\mathcal{F}_s\Big] \tag{9}$$

As underlined by ([Bauwens and Hautsch])the previous relation characterizes the stochastic intensity of a process[1].

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

$$h_N(t) = \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} I\!\!E[N_{t+\Delta t} - N_t|\mathcal{F}_t^N]$$

Let's recall that if $\Lambda_t$ is the compensator of an univariate counting process $N_t$ whose stochastic intensity is $\lambda_t$: $\Lambda_t = \int_0^t \lambda_s ds$. The process: $M_t = N_t - \Lambda_t$ is a $\mathcal{F}_t^N$-martingale, with $M_0 = 0$: this gives the Doob-Meyer decomposition of the process $N_t$.

[1]The class of predictable processes (containing progressive processes) is 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 $A \in \mathcal{F}_s$. Moreover we can see that if equation [9] holds for every $0 \leq s \leq t$ then (leaving aside regularity conditions) for $0 < s < t$:

$$\begin{aligned} I\!\!E[N_t - \int_0^t \lambda_u du|\mathcal{F}_s] &= I\!\!E[N_t - N_s + N_s - \Big(\int_0^s \lambda_u du + \int_s^t \lambda_u du\Big)|\mathcal{F}_s] \\ &= I\!\!E[N_t - N_s|\mathcal{F}_s] - I\!\!E[\int_s^t \lambda_u du|\mathcal{F}_s] + N_s - \int_0^s \lambda_u du \\ &= N_s - \int_0^s \lambda_u du \end{aligned}$$

In the second equality we use the fact that $N_s - \int_0^s \lambda_u du$ is $\mathcal{F}_t$-adapted, and in the third we use the equation [9]. Then the process $(N_t - \int_0^t \lambda_u du)_t$ is a martingale and relation [9] characterizes the stochastic intensity process.

The intensity exists iff the jumps are totally unreachables and if the process is of locally integrable variation (see [Protter, 2003]). It turns out that the class of processes made of processes with continuous compensators can be transformed into a Poisson process through a time transformation. We have then (see [Brémaud, 1991]) :

**Theorem 3.3.** ***[Change of time for an univariate point process]***
*Let $Y_t$ be a process with $\lambda_t$, a $\mathcal{F}_t$-intensity, and $\mu_t$, 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 $I\!P$-as, we have $\lim_{t \longrightarrow \infty} Y_t = \infty$. For $t \in I\!R$ we define $\phi(t)$ the $\mathcal{G}_t$-stopping time such as:*

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

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

### 3.3.2 Endogeneity in duration models

The analysis of endogeneity in duration models is particularly relevant in the context of treatment effect using counterfactuals (see [Heckman]). A counterfactual is the hypothetical view of what have might have occurred if what happened could be undone. Let us denote by $\xi$ (continuous or discrete) the level of a treatment and by $(\tau^\xi)_\xi$ the counterfactual process of outcome durations. For each $\xi \in \Theta \subset I\!R$, $\tau^\xi$ is a duration. Actually this process is degenerate in the sense that the $\tau^\xi$ are all functions of a single random element. We assume :

$$\tau^\xi = (\Lambda^\xi)^{-1}(U) = \Phi^\xi(U) \tag{10}$$

where $U$ is an exponential (1) distribution. For each $\xi$, $\Lambda^\xi$ is the integrated hazard function of $\tau^\xi$ and is assumed to be strictly increasing from $I\!R^+$ to $I\!R^+$. We denote by $\Phi^\xi$ its inverse function and, assuming smoothness conditions:

$$\lambda^\xi(t) = \frac{\partial}{\partial t}\Lambda^\xi(t)$$

is the hazard rate of $\tau^\xi$. The parameter of interest of this model is functional and equal to $\Lambda^\xi(t)$ (or $\lambda^\xi(t)$) or to its inverse $\Phi^\xi(u)$. This counterfactual process is completed by an assignment mechanism. We consider a joint distribution of a vector $((\tau^\xi)_\xi, Z, W)$ where $W$ is a vector of instrumental variables and Z is the level of treatment. We assume:

- $(\tau^\xi)_\xi$ has a marginal distribution characterized by 10;
- We suppose for the moment that the instruments are not time dependent. This assumption is introduced in order to simplify the intuition of the presentation, but the aim of the next sections will be to drop this hypothesis and to incorporate dynamic instruments in the study.

The observed data are $(\tau, Z, W)$ where $\tau = \tau^z$ i.e. the value of outcome at the assigned level of treatment $Z = z$. As in most of the treatment models, the dependence between $\tau$ and $Z$ come from two factors: the treatment effect described by $\Lambda^z$ and the assignment bias captured by the dependence between $U$ and $Z$. Without any additional assumption, these two effects may not be separated and the parameter of interest $\Lambda^Z(t)$ is not identified. We may consider three independence (or conditional independence) conditions which allow identification:

- $H_1$ - $Z \perp U$
- $H_2$ - $Z \perp U|W$
- $H_3$ - $U \perp W$

The condition $H_1$ defines the pure randomization case. In this case $W$ may be neglected and the conditional integrated hazard function of $\tau$ given $Z$ is precisely $\Lambda^z(t)$. This parameter is then identified and may be estimated by usual methods.

The assumption $H_2$ is a conditional randomization condition from which the identification follows from the following argument. Let us denote by $S(t|z,w)$ the conditional survivor function of $\tau$ that is to say:

$$\begin{aligned} S(t|z,w) &= P(\tau \geq t|Z=z, W=w) \\ &= P(U \geq \Lambda^z(t)|Z=z, W=w) \\ &= S_U(\Lambda^z(t)|W=w) \end{aligned}$$

where $S_U(u|W=w) = I\!P(U \geq u|W=w) = P(U \geq u|Z=z, W=w)$ with condition $H_2$. Moreover :

$$P(U \geq u|W=w) = \frac{P(U \geq u)p(w|U \geq u)}{p(w)}$$

where $p$ represents both marginal and conditional density of W. Then:

$$\begin{aligned} S(t|z,w) &= \frac{e^{-\Lambda^z(t)}S(t)p(w|z \leq t)}{S(t)p(w)} \\ &= \frac{e^{-\Lambda^z(t)}S(t|w)}{S(t)} \end{aligned}$$

using obviously $S$ for the marginal and the conditional survivor functions of $\tau$ . Then $\Lambda^z(t)$ is identified.

The third condition is the most interesting case and will be explored in the next section.

## 3.4 Model and integral equation

### 3.4.1 The counterfactual approach

We write a counterfactual model $(\xi, \tau_\xi)$ for durations that are now considered as time of jumps of an elementary counting process: $\mathbf{1}(\tau_\xi \geq t)$. Endogeneity arises through the assignment mechanism since observations will be reduced to realizations $(z, \tau^z)$ of $(\xi, \tau_\xi)$. We are interested in the process :

$$X_{\xi,t} = \mathbf{1}(\tau_\xi \geq t) = \Lambda_\xi(t) + E_{\xi,t}$$

where $E_{\xi,t}$ ought to be a martingale. However, if we observe the process $\xi$ at a given value $Z$, we have:

$$X_t = X_{Z,t} = \Lambda_Z(t) + E_{Z,t}$$

Yet the term $E_{Z,t}$, is no more a martingale. The latent model is still depending on $\xi$, but $E_{Z,t}$ depends on the observed endogenous variables : $z$ and $\tau^z$. The previous considerations are no longer valid, since we don't know if $\mathbb{E}[E_{Z,t}|Z] \neq 0$ as no assumption (even conditional) on the independence between $Z$ and $U$ is made. Now, we have to try to look for an other form of dependence, and try to find for the duration $\tau$ an expression of the form $\tau = \phi(Z, U)$, $Z$ being the set of endogenous variables, and $U$ being a random perturbation. The function is then different from the hazard of $\tau$ conditional on $Z$. Let's $U$ be an exponential variable of parameter 1 and let's consider the process $\mathbf{1}(U \geq t)$. It's easy to show that its compensator is $t \wedge U$. We define:

$$M_t = \mathbf{1}(U \geq t) - t \wedge U.$$

Then $M_t$ a $\mathcal{F}_t$-martingale (where $\mathcal{F}_t$ is the natural filtration of the process $\mathbf{1}(U \geq t)$). Now let's turn to the counterfactual process $X_{\xi,t} = \mathbf{1}(\tau_\xi \leq t)$ with differential form :

$$dX_{\xi,t} = \lambda(t, \xi)(1 - X_{t^-})dt + dE_{\xi,t}.$$

We note $\Lambda(t, \xi) = \int_0^t \lambda(s, \xi)ds$. The integrated hazard function of $X_{\xi,t}$ is then equal to $\Lambda(t \wedge \tau, \xi)$. We have:

$$X_{\xi,t} = \Lambda(t \wedge \tau_\xi, \xi) + E_{\xi,t}$$

**Theorem 3.4.** *Under the previous notations:*

$$E_{\xi,t} = M_{\Lambda(t,\xi)} \quad \Leftrightarrow \quad \tau_\xi = \Lambda^{-1}(U, \xi)$$

See proof on appendix . $E_{Z,t}$ is not bound to be a martingale but $E_{\xi,t}$ may be linked with $M_t$, the compensated version of an elementary counting process relative to an exponential random variable $U$. This is interesting since it allows to link $\tau^\xi$ with $U$ and $\xi$ through the inverse counterfactual integrated hazard function of $X_{\xi,t}$. This will be our function of interest, which we will try to estimate.

### 3.4.2 Transformation in time of a Poisson process

In this section we will explore the case of processes that are obtained through a time-transformation of a standard Poisson process. As expressed in 3.3.1, the processes that allow such a transformation are those with continuous compensators (this class includes in particular increasing processes with locally integrable variation, with jumps occurring at stopping times that are totally unreachables).

We want to work with instruments to treat endogeneity. Note that the condition that we obtained in 3.4 is very close to the characterization of 3.3: the times of jump of the counting process and the corresponding Poisson process are related through a compensator function of the process $X_{\xi,t}$ but in the counterfactual universe.

Exponential variables and Poisson processes are closely related. Let's consider $N_t$ a homogenous Poisson process, of parameter 1. We note $M_t = N_t - t$. We define, still in the counterfactual world:

$$Y_{\xi,t} = \Lambda(t, \xi) + M_{\Lambda(t,\xi)}$$

It's a Poisson process of intensity $\lambda(t, \xi) = \Lambda'(t, \xi)$, since $M_{\Lambda(t,\xi)} = N_{\Lambda(t,\xi)} - \Lambda(t, \xi)$ we have easily:

$$Y_{\xi,t} = \Lambda(t, \xi) + [N_{\Lambda(t,\xi)} - \Lambda(t, \xi)] = N_{\Lambda(t,\xi)}$$

We begin by considering $N_t$, a standard Poisson process: $N_t \sim \mathcal{P}(1)$. We will note $M_t$ the martingale (relatively to the usual filtration) such as $M_t = N_t - t$. Then, $\Lambda$ be a function of $(t, \xi)$, continuous and increasing in $t$, with the initial condition $\Lambda(0, \xi) = 0, \quad \forall \xi$. We define $X_{\xi,t}$ as being:

$$X_{\xi,t} = N_{\Lambda(t,\xi)} \quad \text{or} \quad N_t = X_{\Lambda^{-1}(t,\xi)}$$

In the counterfactual view, this is equivalent, thanks to the former discussion, to :

$$X_{\xi,t} = \Lambda(t, \xi) + M_{\Lambda(t,\xi)}$$

However, the assignation mechanism of the observed value ($\xi = z$) introduces endogeneity :

$$X_{\Lambda^{-1}(t,\xi=z)} = X_{\phi(t,z)}.$$

We want to estimate $\phi(.,.)$ and we consider additional instruments $W$ to treat for endogeneity. Those instruments are not supposed to be time-dependent in a first approach. The main hypothesis we make is that:

**Assumption 3.5.** $W$ are independent from $U$ and $N_t$.

### 3.4.3 Integral equation for $\phi$

We suppose that $X_t$ is a one-jump process: $X_t = \mathbf{1}(\tau \geq t)$ where $\tau$ is a random duration. So is $N_u = \mathbf{1}(U \leq u)$, but we make the additional assumption that $U = \Lambda(\tau, Z)$ follows an exponential law of parameter one conditional to the instruments $W$. Our aim is to estimate $\phi(.,.)$, the inverse of $\Lambda$, observing $\tau$, Z, and W. However, we do not observe $N_t$, but assume that $N_t$ is independent from $W$.

Let's write $f(t, z|w)$ the joint law of $(\tau, Z)$ conditional to $W$. Having $\tau = \phi(U, z)$ and $U = \Lambda(\tau, z)$, if we note $g(u, z|w)$ the joint law of $(U, Z)$ conditional to $W$, we have clearly:

$$g(u, z|w) = \phi'(u, z) \times f(\phi(u, z), z|w)$$

$\phi'(u, z)$ being the derivative of $\phi$ towards its first argument. We will note $f(t, z, w)$ and $g(u, z, w)$ the corresponding joint laws, a point replacing a variable when it has been integrated along it, thus leading to marginal laws. Then the survival function $S(t, z|w)$ will be defined through:

$$S_\tau(t, z|w) = I\!P[\tau \geq t, Z = z|W] = \int_t^\infty \frac{f(t', z, w)}{f(., ., w)} dt'$$

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

$$\int g(u, z|w)dz = e^{-u}$$

$$\int \phi'(u, z) f(\phi(u, z), z|W)dz = e^{-u}$$

Second, we have :

$$S_U(u, z|W) = I\!P[U \geq u, Z = z|W]$$

$$I\!P[U \geq u|W] = \int S_u(u, z|W)dz = e^{-u}$$

as a survival function of an exponential variable. Provided that $\phi$ is monotonic in its first argument, we have:

$$\mathbb{P}[U \geq u, Z = z|W] = \mathbb{P}[\phi(U,z) \geq \phi(u,z), Z = z|W]$$

Then integrating along $z$:

$$\int S_u(u,z|W)dz = \int S_\tau(\phi(u,z),z|W)dz$$

Consequently :

$$\boxed{\int S_\tau(\phi(u,z),z|W)dz = e^{-u}}$$

This second equation was already obtained in the static case : this is natural since we explore the same kind of dependence $\tau = \phi(U,Z)$ with $U$ being an exponential variable of parameter 1. The novelty is the first equation, resulting from the expression of the law of $(U,Z)$ conditional to $W$. 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:

$$\begin{aligned} 1 &= \int \frac{\phi'(u,z) f(\phi(u,z),z|W)}{\int S_\tau(\phi(u,z'),z'|W)dz'}dz \\ &= \int \phi'(u,z) \underbrace{\frac{f(\phi(u,z),z|W)}{S_\tau(\phi(u,z),z|W)}}_{Ⓐ} \times \underbrace{\frac{S_\tau(\phi(u,z),z|W)}{\int S_\tau(\phi(u,z'),z'|W)dz'}}_{Ⓑ} \end{aligned}$$

Ⓐis the hazard function of the process $\{X_t\}$ taken in $\phi(u,z)$ conditional on $Z = z$, $W$. Indeed:

$$\frac{f(t,z|W)}{S(t,z|W)} = \frac{f(t|z,W)f(z|W)}{S(t|z,W)f(z|W)} = \frac{f(t|z,W)}{S(t|z,W)} = \lambda^\tau(t|z,W)$$

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

$$\boxed{\int \phi'(u,z)\lambda^\tau(\phi(u,z)|Z = z, W)g(z|U \geq u, W)dz = 1} \quad (11)$$

It would be tempting to integrate with respect to the variable $u$ and to switch both integrals but this will be difficult since there is a conditioning term in $U$ in the densities in $z$. However if we integrate in $u$ for $u$ varying from 0 to $s$ we get :

$$\int_0^s du \int_z \phi'(u,z)\lambda^X(\phi(u,z)|Z = z, W)g(z|W, \mathcal{F}_u^N)dz = s$$

Then if we switch the integrals and make the following change in variables $t = \phi(u,z)$ we get :

$$\int_z \int_{t=0}^{\phi(s,z)} \lambda_T(t|Z = z, W)g(z|U \geq \phi^{-1}(t,z), W)dzdt = s$$

and as $g(z|U \geq \phi^{-1}(t,z), W) = g(z|\tau \geq t, W)$ we conclude that:

$$\boxed{\int_z \int_{t=0}^{\phi(s,z)} \lambda_T(t|Z = z, W)g(z|\tau \geq t, W)dzdt = s} \quad (12)$$

### 3.4.4 Working with intensities

We can address this question in a more general framework and extend the formula 12 by using intensities rather than using survivorship functions. We need the following lemma:

**Lemma 3.3.** *Let $X_t$ be a process with stochastic intensity $\lambda_t^X$, adapted to the filtration $\mathcal{A}_t^X$ (containing internal history of $X_t$). Let $\phi(t)$ be a monotonous function of time, sufficiently smooth. Then if we define $N_t$ to be $N_t = X_{\phi(t)}$ then $N_t$ has also a stochastic intensity $\lambda^N$ for the filtration $\mathcal{A}_t^N = \mathcal{A}_{\phi(t)}^X$ which is given by:*

$$\lambda_t^N = \phi'(t)\lambda_{\phi(t)}^X$$

The proof is straightforward when we use characterization 9. For $0 < s < t \in I\!R$ we have:

$$\begin{aligned} I\!E[N_t - N_s|\mathcal{A}_s^N] &= I\!E[X_{\phi t} - X_{\phi(s)}|\mathcal{A}_{\phi(s)}^X] \\ &= I\!E[\int_{\phi(s)}^{\phi(t)} \lambda_u du|\mathcal{A}_{\phi(s)}^X] \\ &= E\Big[\int_s^t \phi'(v)\lambda_v dv|\mathcal{A}_s^N\Big] \end{aligned}$$

by making the change of variables $v = \phi(u)$ in the last equation. This therefore characterizes the intensity of $N$. This will help us to extend our result. Indeed, if we note $\lambda_t^N$ and $\lambda_t^X$ the intensities of jump of $N_t$ and $X_t$ respectively, we have that:

$$\lambda^N[u|z, W, \mathcal{F}_u^N] = \lambda^X[\phi(u,z)|z, W, \tau \geq \phi(u,z)] \times \phi'(u,z)$$

Then, as $N_t$ is a process of intensity one:

$$\int \lambda^N[u|z, W, \mathcal{F}_u^N] g(z|W, \mathcal{F}_u^N) dz = 1$$

which rewrites using the previous relation:

$$\boxed{\int \phi'(u,z)\lambda^X(\phi(u,z)|Z = z, W) g(z|W, \mathcal{F}_u^N) dz = 1} \quad (13)$$

Consequently, integrating in $Z$, the general following result still holds:

$$\boxed{\int_z \int_0^{\phi(u,z)} \lambda^X(t|Z = z, W) g(z|F_t^X, W) dt dz - u = 0} \quad (14)$$

### 3.4.5 Generalization

The former section gives us insights to generalize what has been done to the case of dynamic variables and / or instruments. Let's observe that it is always easier to work on intensities rather than on compensators.

Let's suppose that we have for the process $X$ the Doob-Meyer decomposition : $X_t = H_t^X + M_t^X$ for the filtration $X_t v Z_t v W_t$. We suppose that $\phi(t, Z)$ is a family of increasing, finite stopping times, $Z_t$ measurable. We make the additional assumption that for each $t : \phi(t, Z) \geq t$.

We suppose that $N_t = X_{\phi(t,Z)}$. Then : $N_t = H_t^N + M_t^N$ for $X_{\phi(t,Z)} v Z_{\phi(t,Z)} v W_{\phi(t,Z)} = N_t v Z_{\phi(t,Z)} v W_{\phi(t,Z)}$ with $H_t^N = H_{\phi(t,Z)}^X$ and $M_t^N = M_{\phi(t,Z)}^X$.

In this case we need a stronger assumption between $W$ and $N$:

**Assumption 3.6.** $N_\infty$ and $W_\infty$ are independent that is to say, both all the trajectory of $N$ and all the trajectory of $W$ are independent. We suppose that:

$$\mathbb{E}[h_t^N | N_t v W_{\phi(t,Z)}] = \mathbb{E}[h_t^N | N_t]$$

We have that:

$$\int_Z \phi'(s,Z) h^X(\phi(s,Z)|X_{\phi(s,Z)} v Z_{\phi(s,Z)} v W_{\phi(s,Z)}) g(Z_{\phi(s,Z)}|X_{\phi(s,Z)} v W_{\phi(s,Z)}) dZ = 1$$

$$\in_0^t ds \int_Z \phi'(s,Z) h^X(\phi(s,Z)|X_{\phi(s,Z)} v Z_{\phi(s,Z)} v W_{\phi(s,Z)}) g(Z_{\phi(s,Z)}|X_{\phi(s,Z)} v W_{\phi(s,Z)}) dZ = t$$

If we make the change of variables $u = \phi(s,Z)$ then:

$$\int_Z dZ \int_0^{\phi(t,Z)} h^X(u|X_u v W_u v Z_u) g(Z_u|X_u v W_u) du = t$$

**Dynamic covariates with all the trajectory :** suppose that we have dynamic covariate but that we can observe the whole trajectory of the processes (instruments and variables). The equation becomes:

$$\int_Z dZ \int_0^{\phi(t,Z)} h^X(u|X_u v W v Z) g(Z|X_u v W) du = t$$

### 3.5 Estimation

We have to estimate $\lambda^X()$ and $g()$ that are unknown. It will be possible to give estimators $\hat{\lambda}$ of $\lambda(t|z,w)$ but the situation for $g(z|\mathcal{F}_t^X, W)$ will be slightly different, since we will be unable to derive the speed of convergence of a potential estimator $\hat{g}$ without precisely defining the nature of the dependence between $Z$ and $\mathcal{F}_t^X$, even in the static case.

We will work with fixed values of "u". We suppose that for any $(u,z)$ we have that $z \mapsto \phi(u,z) \in L_Z^2$ and is $\mathbb{R}^+$-valued. We consider $T$ as expressed formerly, and that for any $u$ fixed, $T\phi \in L_W^2$. If $\phi_0$ stands for the true solution of equation 12 we will note $\hat{T}_n$ the proposed estimator of $T$ for a sample made of $n$ observations. If $\hat{\lambda}$ and $\hat{g}$ are estimators of $\lambda$ and $g$, a form of $\hat{T}_n$ could be:

$$\hat{T}_n = \int\int \hat{\lambda}\hat{g}$$

In the following, we will drop the indexation by $n$. As $n$ goes to infinity, we will adopt the generic notation $\delta_n$ for the speed of convergence that is to say:

$$||\hat{T}_n - T||_{L_W^2} = O(\delta_n)$$

### 3.5.1 Speed of convergence

The easiest way to see that $\delta_n$ strongly relies on the assumption made on $Z$ conditionally on $\mathcal{F}_t^X$ is to linearize and to see that :

$$||\hat{\lambda}\hat{g} - \lambda g|| \leq ||(\hat{\lambda} - \lambda)g|| + ||(\hat{g} - g)\lambda|| + o(||\hat{\lambda} - \lambda|| \times ||\hat{g} - g||)$$

Then the speed of convergence of $\hat{\lambda}\hat{g}$ is driven by the lowest speed reached by either $\hat{\lambda}$ or $\hat{g}$. As both $\lambda$ and $g$ are conditional densities or intensities, those speed of convergence will rely on the dimension of the conditioning variable. For example, as soon as the dimension of $(Z, W)$ is greater than the dimension of $(\mathcal{F}_t^X, W)$, we will only have to control for the speed of convergence of $\lambda$. For instance, in the case of a discrete time sample of $X$, keeping all the past observations of $X$ would lead to a dimensionality problem concerning the speed of $\hat{g}$.

### 3.5.2 Example

We consider equation (12) and suppose that $\mathcal{F}_t^X = \sigma(\{\tau \geq t\})$. We will use kernel density estimation, kernels will be noted $K(.)$ and will sometimes stand for a generalized multi-dimensional kernel, depending on the dimension of the concerned variable, with associated bandwidth (generically denoted by $h_.$, depending on the size $n$ of the sample but we drop this indexation for sake of simplicity). The joint density $f(t, z, w)$ can be estimated through:

$$\hat{f}(t, z, w) = \frac{1}{nh_t h_z^p h_w^q} \sum_{i=1}^{n} K(\frac{t - t_i}{h_t})K(\frac{z - z_i}{h_z})K(\frac{w - w_i}{h_w})$$

for a set of observations $(t_i, z_i, w_i)_{i \in [1;n]}$, where $p$ is the dimension of $Z$, $q$ the dimension of $W$. We suppose moreover that there exist an analytical expression of the survival function of the kernel in $t$, noted $\bar{K}$, such as : $\bar{K}(t) = \int_t^\infty K(t')dt'$. Consequently, the survival function is estimated through:

$$\hat{S}(t|z, w) = \frac{1}{h_z^p h_w^q} \sum_{i=1}^{n} \bar{K}(\frac{t - t_i}{h_t})K(\frac{z - z_i}{h_z})K(\frac{w - w_i}{h_w}).$$

Finally:

$$\hat{\lambda}(t|z, w) = \frac{1}{h_t} \frac{\sum_{i=1}^n K(\frac{t-t_i}{h_t})K(\frac{z-z_i}{h_z})K(\frac{w-w_i}{h_w})}{\sum_{i=1}^n \bar{K}(\frac{t-t_i}{h_t})K(\frac{z-z_i}{h_z})K(\frac{w-w_i}{h_w})}$$

and:

$$\hat{g}(z|\tau \geq t, W) = \frac{1}{h_z^p} \frac{\sum_{i=1}^n K(\frac{z-z_i}{h_z})K(\frac{w-w_i}{h_w})\mathbf{1}_{t \leq t_i}}{\sum_{i=1}^n K(\frac{w-w_i}{h_w})\mathbf{1}_{t \leq t_i}}.$$

Then, $\hat{T}(\phi)$ as a function of $W$ for a fixed $u$ may be estimated by:

$$\int_z \int_0^{\phi(u,z)} \hat{\lambda}(t|z, W)\hat{g}(z|\tau \geq t, W)dtdz - u$$

We summarize here the hypotheses on the kernels and on the densities that we have to make in order to derive speed of convergence. First, we suppose that the variables of interest take their values in compact sets: $\tau \in [0; T]$, $Z \in [0; 1]^p$ and $W \in [0; 1]^q$. Second, we consider nonparametric kernels $K$ with corresponding bandwidths $h_t$, $h_z$, and $h_w$. Kernels for estimation of $Z$ and $W$ are respectively $p$ and $q$ dimensional generalized product kernel functions. We suppose moreover that:

- $\int K(u)du = 1$;

- $\int K^2(u)du < \infty$;
- $\int_t^{+\infty} K(u)du$ has a closed form expression.

We suppose that for each kernel of bandwidth $h$ there exists an order $r \in \mathbb{N}$ such as $h^{-j+1} \int u^j K(u/h)du$ is equal to 1 for $j = 0$, 0 if $j \in [1; r-1]$, and is different from 0 for $j = r$. In the following, $r$ will denote the smallest order of those kernels. Each bandwidth depends on the sample size $n$ and $h_{n,t} \to 0$, $h_{n,z} \to 0$, $h_{n,w} \to 0$, and $nh_{n,t} \to \infty$, $nh^p_{n,z} \to \infty$, $nh^q_{n,w} \to \infty$. Third, we suppose the intensity $\lambda(t|z,w)$ is $d_\lambda$-times continuously differentiable with a bounded $d_\lambda$-derivative and that $g(z|\tau \geq t, w)$ is $d_g$ times continuously differentiable with a bounded $d_g$-derivative. Under those hypotheses, and if $\rho$ is the minimum between the order $r$ of the kernels, $d_\lambda$ and $d_g$, the estimator $\hat{\lambda}\hat{g}$ tends to $\lambda g$ (in a $L^2$ sense), at a speed $v_n$ which is :

$$\boxed{v_n = O\left(\frac{1}{nh^p_{n,z}h^q_{n,w}} + (h^p_{n,z}h^q_{n,w})^{2\rho}; \frac{1}{nh_t h^q_{n,w}} + (h^q_{n,w})^{2\rho}\right)}$$

### 3.5.3 Simple forms of $\phi$

We can also make some simplifying models for $\phi$. In the case of a risk proportional model, we suppose that there is a semi-parametric model on $\phi$, and that:

$$\phi(u,z) = \gamma(\psi(z)u) \quad \text{and} \quad \psi(z) = \psi_\beta(z) = \psi(z,\beta)$$

with for example $\psi(z,\beta) = \exp(\beta z)$. This implies that the model becomes:

$$\int_z \int_0^{\gamma(u)} \psi(z)\lambda_X(\psi(z)s|z,W)g(z|\tau \geq \psi(z)s, W)dsdz - u = 0 = T(\gamma,\beta)$$

This assumption allows to switch the integrals in the former expression, which is not possible in the general case. It is even sometimes possible to consider $\gamma(x,y) = xy$.

# 4 The Non-separable case : identification and convergence

## 4.1 An ill-posed inverse problem

### 4.1.1 Definition

We want to show that the problem we are studying is an ill-posed linear one. We first assume without loss of generality that $Z$ and $W$ have a compact support and take their values on $[0;1]^{d_Z}$ and $[0;1]^{d_W}$. We recall that $U \sim Exp(1)$. The function $(u,z) \mapsto \phi(u,z)$ to be estimated satisfies :

$$\int_z \int_{t=0}^{\phi(u,z)} \lambda_T(t|Z=z,W)g(z|\tau \geq t, W)dzdt = u$$

and is defined on $\mathbb{R}^+ \times [0;1]^{d_Z}$. Let's define $L^2(U,Z)$, $L^2(U,W)$ the spaces of functions such as :

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

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

We define:

$$\begin{aligned} T : L^2(U,Z) &\mapsto L^2(U,W) \\ \phi &\mapsto T(\phi) : (u,w) \to \int_z \int_0^{\phi(u,z)} a(s,z,w) ds dz - u \end{aligned}$$

where $a(s,z,w) = \lambda^X(s|Z=z,W) g(z|\tau \geq s,w)$. We try to solve the following inverse problem: $T(\phi) = 0$. In fact this problem has to be solved for fixed values of $u$. To simplify notations, we will sometimes note in the following $(X) = L^2(U,Z)$ and $(Y) = L^2(U,W)$.

This problem is nonlinear : to analyze its ill-posedness the first step is to check if the operator T is Frechet-differentiable. We estimate the Gateaux-derivative of the operator which 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 this derivative in $\phi$ is the operator :

$$\begin{aligned} \tilde{\phi} \mapsto T'_\phi(\tilde{\phi}) &= \int_z \tilde{\phi}(u,z) \underbrace{a(\phi(u,z),z,w)}_{f_\tau(\phi(u,z),z|\tau \geq \phi(u,z),W)} dz \\ &= \int_z \frac{\tilde{\phi}(u,z)}{\phi'(u,z)} \underbrace{\phi'(u,z) f_\tau(\phi(u,z),z|\tau \geq \phi(u,z),W)}_{=f_U(u,z|U\geq u,W)} dz. \end{aligned}$$

For each function $\phi \in L^2(U,Z)$, $T'_\phi$ is linear from $L^2(U,Z)$ to $L^2(U,W)$.
To ensure that $T$ is Frechet-differentiable, we have to check whether:

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

We make the following assumption:

**Assumption 4.1.** $\forall \phi \in L^2(U,Z)$ we have:

$$\int_{I\!R^+ \times [0;1]^{d_W}} \int_z a^2(\phi(u,z),z,w) f_U(u) f_W(w) du dw dz < +\infty.$$

Assumption 4.1 ensures that $\forall \phi$, $T'_\phi$ is bounded, as it is linear, $T'_\phi$ is continuous for every $\phi$. We still need to prove that the mapping $\phi \mapsto T'_\phi$ is continuous on $L^2(U,Z)$. In fact, for our analysis of ill-posedness, we only need this mapping to be continuous on the true solution $\phi = \phi_0$.

As $T'_\phi$ is linear, $||T'_\phi||_{\mathfrak{L}(\mathfrak{X},\mathfrak{Y})} = sup_{\tilde{\phi}\in\mathfrak{X},||\tilde{\phi}||=1}||T'_\phi||_{\mathfrak{Y}}$. $\forall\phi_1 \in \mathfrak{X}$, we have for $\tilde{\phi} \in \mathfrak{X}$ and $u, w$:

$$(T'_{\phi_0} - T'_{\phi_1})(\tilde{\phi}) = \int_z \tilde{\phi}(u,z)\left\{\lambda(\phi_0(u,z),z,w)g(z|\tau \geq \phi_0(u,z),w)\right\}dz$$

which can be rewritten:

$$\begin{aligned}(T'_{\phi_0} - T'_{\phi_1})(\tilde{\phi}) &= \int_z \tilde{\phi}(u,z)\lambda(\phi_0(u,z),z,w)\left\{g(z|\tau \geq \phi_0(u,z),w) - g(z|\tau \geq \phi_1(u,z),w)\right\}dz \\ &+ \int_z \tilde{\phi}(u,z)g(z|\tau \geq \phi_1(u,z),w)\left\{\lambda(\phi_0(u,z),z,w) - \lambda(\phi_1(u,z),z,w)\right\}dz\end{aligned}$$

□

We may need the expression of the adjoint operator of the Fréchet derivative. For functions $\tilde{\phi} \in L^2(U,Z)$ and $\tilde{\psi} \in L^2(U,W)$ we seek the linear operator $T'^{*}_{\phi_0}$ from $L^2(U,Z)$ to $L^2(U,W)$ such that:

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

Writing explicitly:

$$\begin{aligned}< T'_{\phi_0}(\tilde{\phi}), \tilde{\psi} >_{L^2(U,W)} &= \int_w (T'^{*}_{\phi_0}(\tilde{\phi})(u,z))(\tilde{\psi}(u,w))f_U(u)f_W(w)dudw \\ &= \int_w (\int_z \tilde{\phi}(u,z)\lambda_T(\phi_0(u,z)|z,w)g(z|\tau \geq \phi_0(u,z),w)dzdt) \\ &\quad \times(\tilde{\psi}(u,w))f_U(u)f_W(w)dudw \\ &= \int_z (\int_w \tilde{\psi}(u,w)\lambda_T(\phi_0(u,z)|z,w)g(z|\tau \geq \phi_0(u,z),w)f_W(w)) \\ &\quad \times(\tilde{\phi}(u,z))f_U(u)dzdtdudw \\ &= \int_z (\int_w \tilde{\psi}(u,w)\lambda_T(\phi_0(u,z)|z,w)g(w|\tau \geq \phi_0(u,z),z) \\ &\quad \times f_Z(z))(\tilde{\phi}(u,z))f_U(u)dzdtdudw \\ &= \int_z (\int_w \tilde{\psi}(u,w)\lambda_T(\phi_0(u,z)|z,w)g(w|\tau \geq \phi_0(u,z),z)dudw) \\ &\quad \times(\tilde{\phi}(u,z))f_Z(z)f_U(u)dzdt \\ &= < \tilde{\phi}, T'^{*}_{\phi_0}(\tilde{\psi}) >_{L^2(U,Z)}\end{aligned}$$

where:

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

### 4.1.2 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 [Chernozhukov et al., 2008], and try to work directly on $T$.

Alternatively, one can try to show that $T'$ computed on the true solution is compact. This condition will be sufficient only in the case of infinite dimension of the range of $T'$ (which is straightforward since the arrival space of $T'_{\phi_0}$ is $L^2(U,W)$.

Suppose then that we want to show the compactness of $T'_\phi$. A first 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, 1992] used by [Chernozhukov et al., 2008].

We could also examine under which conditions the operator is an Hilbert-Schmidt one.

**Assumption 4.2.** We assume that if $\phi_0$ is the true solution of the problem :

$$\int_u \int_w \int_z a^2(\phi(u,z),z,w) \frac{f_W(w) f_U(u)}{f_Z(z)} dz dw du < +\infty.$$

With condition 4.2, 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 $T'_\phi$ since:

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

with:

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

Under those conditions, $T'_{\phi_0}$ is Hilbert-Schmidt and therefore compact.

## 4.2 Identification

We now want to explore the conditions for the identification of $\phi$ defined by equation 14. Let's assume that there exists two functions $\phi_1$ and $\phi_2$ such that 11 holds. We assume moreover that for normalization conditions, $\phi_1(0,z) = \phi_2(0,z) = 0 \quad \forall z$. Then:

$$\int \{\phi_1'(u,z)\lambda^X(\phi_1(u,z)|Z=z,W) - \phi_2'(u,z)\lambda^X(\phi_2(u,z)|Z=z,W)\} g(z|U \geq u, W) dz = 0$$

**Assumption 4.3.** $Z << W \quad | \quad U \geq u$ which means that for all function $\rho(.,.)$:

$$\mathbb{E}[\rho(Z,u)|W, U \geq u] = 0 \quad \Longrightarrow \quad \rho(z,u) = 0$$

Under these conditions, the previous equality implies that:

$$\forall u, \quad \phi_1'(u,z)\lambda^X(\phi_1(u,z)|Z=z,W) = \phi_2'(u,z)\lambda^X(\phi_2(u,z)|Z=z,W).$$

Each term is the derivative of $\Lambda^\tau(\phi_i(u,z)|z,W)$ w.r.t. $u$ where $i \in \{1,2\}$. Then there exists $\alpha \in \mathbb{R}$ such that:

$$\Lambda^X(\phi_1(u,z)|z,W) = \Lambda^X(\phi_2(u,z)|z,W) + \alpha \quad \forall u, z.$$

Especially for u=0:

$$\Lambda^X(\phi_1(0,z)|z,W) = \Lambda^X(\phi_2(0,z)|z,W) = \Lambda^X(0|z,W) = 0$$

then $\alpha = 0$ and:

$$\boxed{\Lambda^X(\phi_1(u,z)|z,W) = \Lambda^X(\phi_2(u,z)|z,W) \quad \forall u,z}.$$

We see that the second assumption in order to get identification, i.e. $\phi_1 = \phi_2$ is that $\Lambda^X$ has to be injective. We supposed earlier that the compensator of $X$ had to be continuous. This assumption is fundamental, but not to ensure injectivity. However, we supposed that $\Lambda$ was increasing: this is not sufficient, it has to be strictly increasing. This is equivalent to the fact that $\lambda^X$ never vanishes. When $\lambda^X = 0$, the process becomes deterministic as we know it cannot jump. The hypothesis is formally similar to the identification condition in the static case: however it is far stronger since the identification is controlled on the trajectory as long as $U \geq u$.

We can also recover this condition by using operators. We study the operator $T : \phi \mapsto T(\phi)$ such as for a function $\phi$ we have:

$$T(\phi) : (u,w) \mapsto \int_z \int_{s=0}^{\phi(u,z)} a(s,z,w)dzds - u$$

(where the function $a(.,.,.)$ is expressed through equation 14). We are looking for functions $\phi$ such as $T(\phi) = 0$. $\phi(.,.)$ takes for argument $s$ and $Z$, and $T$ maps $\phi$ into a function of $s$ and $W$. Additionally, we will have to constrain $\phi$ to be monotonic, increasing. Therefore $T$ will be considered as a mapping between the two Hilbert spaces:

$$T : L^2(U,Z) \rightarrow L^2(U,W)$$

If we want $T$ to be one-to-one, that is to say $T'_\phi(\tilde{\phi}) = 0$, this implies the same condition than in the former approach, that is to say : $Z$ has to be strongly identified by $W$ conditional to $U$.

## 4.3 Convergence

Operator $T$ is defined as :

$$T : \phi \rightarrow \int_z \int_{t=0}^{\phi(s,z)} \lambda_T(t|Z=z,W) g(z|\tau \geq t, W) dzdt - s$$

and we try to solve $T(\phi) = 0$. $\phi$ is a transformation of time, takes its values in $I\!R^+$, and its arguments in $[0;T] \times I\!R^n$ or $I\!R^+ \times I\!R^n$. Moreover we assume that $\phi$ belongs to $L^2(U,Z)$. In the following, we will note $\hat{T} = \hat{T}_n$ an estimator of $T$. Generically, $\hat{T}$ is :

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

where $\hat{a}(z, W, t)$ is an estimator of $a(z, W, t) = \lambda_T(t|Z = z, W)g(z|\tau \geq t, W)$. We face a non linear inverse problem, where $T$ and even $T'$ are not known[2]) which implies the use of a regularization technique, and present first the case of regularization in Hibert Scales.

#### 4.3.1 Regularization in Hilbert scales

We want to solve in $\phi$ the following problem:

$$\min_{\phi} ||\hat{T}(\phi)||^2 + \alpha_n||\phi - \phi^*||_s^2 \quad \text{with} \quad \phi(u, z) \in L_Z^2 \quad \text{for u fixed} \tag{15}$$

where $\alpha_n$ is a regularization parameter and $\phi^*$ is an arbitrary function. In particular, we will have to control the behavior of the parameter $\alpha_n$ 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 Frechet-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 15 towards the true solution of the initial problem 12.

**Assumption 4.4.** Assume that:

- if $\phi_0$ is a solution of the problem 12 then there exists a sequence $\delta_n$ such as

$$||\hat{T}(\phi_0) - T(\phi_0)|| = ||\hat{T}(\phi_0)|| \leq \delta_n$$

- $\delta_n$, $\alpha_n$, and $\delta_n^2/\alpha_n$ tends to 0 as $n$ increases to infinity;

We can show that under hypothesis 4.4 we have:

**Lemma 4.1.** *If $(\hat{\phi}_n^\alpha)$ is a sequence of solutions of the related minimization problems (15), then there exists a subsequence $(\hat{\phi}_{n,2}^\alpha)$ of $(\hat{\phi}_n^\alpha)$ which converges towards a function $\phi_l$ (in a $L_W^2$ sense). Moreover $\phi_l$ is a solution of the problem 12.*

See proof in appendix A.3. We will only need to suppose that $||\hat{T} - T|| \to 0$ which is not a too strong assumption.

**Unicity :** some conditions may be examined to ensure unicity of $\phi_l$. If the problem 12 is identified, $\phi_0$ is unique and $\phi_0 = \phi_l$. As soon as $\phi_l$ 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 $\phi_0$. 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 $\phi^*$-minimal norm solutions. Then $\phi_0$ is taken as the unction, among the set of solutions $\phi$ of 12, which minimizes the quantity $||\phi - \phi^*||_s$. Then, we have that $\phi_0 = \phi_l$. Indeed:

[2]As we are looking for $\phi$ which appears in the expression of $T'$.

$$
\begin{aligned}
||\phi_l - \phi^*|| &\leq \limsup_{n\to\infty} ||\hat{\phi}_{2,\alpha_n} - \phi^*|| \\
&\leq ||\phi_0 - \phi^*|| \\
&\leq ||\phi_l - \phi^*||.
\end{aligned}
$$

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 $\phi_0$ is a $\phi^*$-minimal norm solution and that $\phi_l$ is itself a solution.

**Speed of convergence :** we suppose that the solution initial problem is identified and want to derive the speed of convergence of the solutions $(\hat{\phi}_n^\alpha)$ to the true solution $\phi_0$ (when we suppose that that the initial problem is identified). We mainly need conditions that are similar to those of [Engl et al., 1996], with additional assumptions concerning the Frechet derivative of $\hat{T}$ and $T$ in $\phi_0$.

**Assumption 4.5.** We suppose that:

- (i) - the problem (12) is identified with a true solution $\phi_0$
- (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 $\gamma_n$ such as $||\hat{T}'_{\phi_0} - T'_{\phi_0}|| \leq \gamma_n$
- (v) - there exists $\beta \in I\!R$ such as $\phi_0 - \phi^* \in H_{-\beta}$ where $(H_s)_{s\in I\!R}$ 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 4.4 and A.5 we have the following lemma:

**Lemma 4.2.**

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

See the proof in appendix A.4.

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 $\gamma_n$ and the convergence of the Frechet 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 solutions. In our case, as $T$ is estimated along two integrals, and $T'$ only one, $\gamma_n$ will probably be slower than $\delta_n$ and can not be left aside. Yet, we will need in the following the next assumption:

**Assumption 4.6.** We suppose that:

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

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

**Lemma 4.3.** *The best choice for $\alpha_n$ is:*

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

*If moreover we make assumption 4.6, we get a speed of convergence equal to:*

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

The rate of $\alpha_n$ is chosen according to the usual case and leads to a speed of convergence that is similar to the standard situation. However, if hypothesis 4.6 is not verified, the terms in $\gamma_n$ are too slow and the result does not hold, those terms driving the speed of convergence of the sequential solutions towards the true solution.

### 4.3.2 Using a $L^2$ penalization

We can reformulate the problem with a more traditional kind of penalization. We study the same kind of minimization objective with $||\phi - \phi^*||$ taken under a $L^2$ norm. That is to say :

$$\min_{\phi} ||\hat{T}(\phi)||^2 + \alpha_n ||\phi - \phi^*||^2 \quad \phi(u,z) \in L^2_Z \quad \text{for u fixed} \tag{16}$$

where $\alpha_n$ is still the regularization parameter and $\phi^*$ is an arbitrary function. The study of 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 hypothesis than 4.4 and the demonstration is not affected by the change of penalization and result 4.1 still holds (see appendix A.6).

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.

**Assumption 4.7.** We suppose that:

- (i) - the problem (12) is identified with a true solution $\phi_0$

- (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 $\gamma_n$ 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 4.4 and 16, we can show that :

**Lemma 4.4.** *The best choice for $\alpha_n$ is:*

$$\alpha_n \sim \delta_n$$

*and the resulting speed of convergence is :*

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

$$||\hat{T}(\hat{\phi})|| = O(\delta_n)$$

### 4.3.3 Regularization by iteration

We may also try to regularize this problem by iterative methods. The principle of the method is to linearize the operator around the current solution and to update it (see [Kaltenbacher et al., ]). 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 $\phi_k$ at the iteration $k$, we try to solve the linearized problem:

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

However in practice, this linear problem may also be ill-posed and has to be regularized. If the Tikhonov regularization is applied, then this is the Levenberg-Marquardt method. With an additional penalty term this is the iteratively regularized Gauss-Newton method. The Levenberg-Marquardt method leads to the expression:

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

A stopping criterion has to be applied to assess that there is no need to further iterate. The Tikhonov regularization parameter $\alpha_k$ has also to be controlled and depends on the iteration parameter $k$. When we use the iteratively regularized Gauss-Newton method, the iteration is the following:

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

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

How to apply this framework to our data? Suppose that a sample of size $N$ is available: $(\tau_j, z_j, w_j)$. We need to express $T'_{\phi_k}\phi$ for given $\phi_k, \phi \in L^2(U,Z)$. In the following as we will work with a fixed, we will omit $u$ in the expression of the $\phi_{...}$ functions. $\phi$ may be represented with the help of $(z_j, \phi^{(j)})$ where $\phi^{(j)} = \phi(u, z_j)$. In this case $\phi$ is estimated through

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

$K()$ is here taken as the standard notation for an appropriated kernel, adapted to the modelled object (despite the notations, kernels may differ depending on $z$, $w$, etc.). Similarly:

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

Suppose that we have two estimators $\hat{\lambda}$ and $\hat{g}$ of $\lambda$ and $g$. Then:

$$T'_{\hat{\phi}_k}\hat{\phi} = \int_z \sum_{j=1}^{N} \frac{\phi^{(j)} K(z - z_j)}{\sum_{j=1}^{N} K(z - z_j)} \hat{\lambda}(\hat{\phi}(z)|z, w)\hat{g}(z|w)dz.$$

For instance, we may define $\hat{\lambda}(\hat{\phi}(z)|z, w) = \lambda_0(\hat{\phi}(z))exp(\beta' z + \gamma' w)$ where $\lambda_0$, $\beta$ and $\gamma$ result from the estimation of a general Cox model. $\hat{g}(z|w)$ may be derived with a classical estimator of a conditional density with a given kernel. At this stage, $T'_{\hat{\phi}_k}\hat{\phi}$ is still a function of $w$. We then conclude by applying this function at the points $(w_j)$ of the sample, obtaining then a linear problem using a matrix formulation:

$$A_k\phi = v_k$$

were $\phi$ is the vector of the $(\phi^{(j)})_{j\in[1;N]}$ and $v_k$ is the vector of the $((-T\phi_k)(w_j))_{j\in(1;N]}$. $A_k$ is the matrix made of the terms $a_{ij}$:

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

# A APPENDIX

## A.1 Proof of theorem 2.1

We assume that : $X_t = H_t + E_t$ w.r.t. filtration $Z_t$ and $X_t = \Lambda_t + M_t$ w.r.t. filtration $W_t$ and that both $H_t$ and $\Lambda_t$ are smooth and have an intensity, respectively $h_t$ and $\lambda_t$. Let $t_0 \in I\!\!R$. For $t > t_0$ we have:

$$X_t - X_{t_0} = \int_{t_0}^{t} h_s ds + E_t - E_{t_0}.$$

If we take the conditional expectation of this expression with respect to the filtration $W_{t_0}$, considering that $I\!\!E[E_t - M_{t_0}|W_{t_0}] = 0$ by assumption, we get:

$$I\!\!E[X_t - X_{t_0}|W_{t_0}] = I\!\!E[\int_{t_0}^{t} h_s ds|W_{t_0}].$$

Equivalently:

$$I\!\!E[X_t - X_{t_0}|W_{t_0}] = I\!\!E[\int_{t_0}^{t} \lambda_s ds|W_{t_0}].$$

by hypothesis. We confront the two expressions of $I\!\!E[X_t - X_{t_0}|W_{t_0}]$ and divide by $t - t_0$ to get:

$$I\!\!E[\frac{1}{t-t_0}\int_{t_0}^{t} h_s ds|W_{t_0}] = I\!\!E[\frac{1}{t-t_0}\int_{t_0}^{t} \lambda_s ds|W_{t_0}].$$

As a property of the conditional expectation operator, we can switch the integral and the expectation as soon as for each equation, at least one of two members exists :

$$I\!\!E[\frac{1}{t-t_0}\int_{t_0}^{t} h_s ds|W_{t_0}] = \frac{1}{t-t_0}\int_{t_0}^{t} I\!\!E[h_s|W_{t_0}]ds$$

$$I\!\!E[\frac{1}{t-t_0}\int_{t_0}^{t} \lambda_s ds|W_{t_0}] = \frac{1}{t-t_0}\int_{t_0}^{t} I\!\!E[\lambda_s|W_{t_0}]ds$$

Finally, we want to take the limit under the expectation operator for $t \to t_0$. We will use the Lebesgue theorem of dominated convergence. Then, we must make an additional assumption on $(h_s)$ and $(\lambda_s)$. Although it may not be necessary, it may be sufficient to assume that both $h_s$ and $\lambda_s$ are bounded. Moreover, to take the limit $\lim_{t \searrow t_0}$ of these expressions, we must suppose that both processes $h_s$ and $\lambda_s$ are right-continuous. We have then:

$$I\!\!E[h_{t_0}|W_{t_0}] = I\!\!E[\lambda_{t_0}|W_{t_0}].$$

As $I\!\!E[h_{t_0}|W_{t_0}] = h_{t_0}$. We conclude that :

$$\lambda_{t_0} = I\!\!E[h_{t_0}|W_{t_0}]$$

□

## A.2 Proof of theorem 3.4

($\Leftarrow$) Let's express $X_{\xi,t} = \mathbf{1}(\tau_\xi \leq t)$ with: $X_{\xi,t} = \Lambda(t \wedge \tau_\xi, \xi) + E_{\xi,t}$ which rewrites $E_{\xi,t} = X_{\xi,t} - \Lambda(t \wedge \tau_\xi, \xi)$. Then, by definition of $M_{\Lambda(t,\xi)}$:

$$\begin{aligned} M_{\Lambda(t,\xi)} &= \mathbf{1}(U \leq \Lambda(t,\xi)) - U \wedge \Lambda(t,\xi) \\ &= \mathbf{1}(\underbrace{\Lambda^{-1}(U,\xi)}_{=\tau_\xi} \leq t) - \underbrace{\Lambda(\tau_\xi,\xi) \wedge \Lambda(t,\xi)}_{=\Lambda(t\wedge\tau_\xi,\xi)}) \\ &= X_{\xi,t} - \Lambda(t \wedge \tau, \xi) \\ &= E_{\xi,t} \end{aligned}$$

($\Rightarrow$) Let's assume that $E_{\xi,t} = M_{\Lambda(t,\xi)}$. We have $X_t = \mathbf{1}(\tau_\xi \geq t) = \Lambda(t \wedge \tau_\xi, \xi) + E_{\xi,t}$ with $\tau_\xi \geq 0$. By definition:

$$E_{\xi,t} = \mathbf{1}(\tau_\xi \geq t) - \Lambda(t \wedge \tau_\xi, \xi)$$

$$M_{\Lambda(\xi,t)} = \mathbf{1}(U \geq \Lambda(\xi,t)) - \Lambda(\xi,t) \wedge U$$

As $\Lambda(t,\xi) \wedge \Lambda(\tau_\xi,\xi) = \Lambda(t \wedge \tau_\xi, \xi)$, the equality $E_{\xi,t} = M_{\Lambda(\xi,t)}$ holds as soon as $U = \Lambda(\tau_\xi,\xi)$ i.e. $\tau_\xi = \Lambda^{-1}(U,\xi)$ (being the inverse function in $U$, $\xi$ being fixed).

□

## A.3 Proof of lemma 4.1

We note $\hat{\phi}_n^\alpha$ the sequence of solutions of problems 15 (with $\phi^*$ an arbitrary function) and $\phi_0$ a true solution of the initial problem. We want now to show that under hypothesis 4.4, there exists a subsequence of solutions which converges to a function which is solution of 12. 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' 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 $\phi_l$ the limit of this subsequence. The question is to know if $\phi_l$ is itself a solution of 12. First let's remark that $\hat{T}_n(\hat{\phi}_{\alpha_n,2})$ 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:

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

with the right hand side which tends to $0$. Indeed, $||\phi_l||_s$ is fixed, $||\phi_l - \hat{\phi}_{\alpha_n,2}||_s \to 0$, and we just have to suppose 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 towards $\phi_l$, which is solution of the initial problem.

□

## A.4 Proof of lemma 4.2

In the following, for sake of simplicity, we will skip indexation by $n$, working with $\hat{\phi}$, $\alpha$, $\hat{T}$, and $\delta$ instead of $\hat{\phi}_n^\alpha$, $\alpha_n$, $\hat{T}_n$, and $\delta_n$. The problem is to minimize :

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

In particular we have :

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

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 can suppose in the following without loss of generality that $\phi^* = 0$. We write the decomposition of $\hat{T}$ in $\phi_0$ with the Frechet 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 being negative, this becomes:

$$\begin{aligned}||\hat{T}'_{\phi_0}(\hat{\phi}-\phi_0)||^2+\alpha||\hat{\phi}-\phi_0||_s^2 &\leq \delta^2+\alpha<\hat{\phi}-\phi_0,\phi_0>_s -2<\hat{T}(\phi_0)+\hat{r},\hat{T}'_{\phi_0}(\hat{\phi}-\phi_0)> \\ &\quad -2<(\hat{T}'_{\phi_0}-T'_{\phi_0})(\hat{\phi}-\phi_0),T'_{\phi_0}(\hat{\phi}-\phi_0)> \qquad (17)\end{aligned}$$

Using hypothesis A.5-(vi), we have that $||T'_{\phi_0}(\hat{\phi}-\phi_0)||^2 \sim ||\hat{\phi}-\phi_0||^2_{-a}$. Moreover the source condition (hypothesis A.5-(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 as:

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

as $L^s$,$L^{-\beta}$ are self-adjoint, the last equality comes 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 get simply by Cauchy-Schwartz inequality:

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

Using hypothesis A.5-(iii) we have consequently:

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

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

Then:

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

Moreover it's easy to obtain:

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

by applying the Cauchy-Schwartz inequality and hypotheses A.5-(iv) and (vi). With all of this, equation (17) becomes:

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

□

### A.5 Proof of lemma 4.3

Let's present 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:

$$||x||_r \leq ||x||_q^{\frac{s-r}{s-q}} ||x||_s^{\frac{r-q}{s-q}} \tag{18}$$

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

$$c^p \leq e + dc^q \rightarrow c^p = O(e + d^{\frac{p}{p-q}}) \tag{19}$$

We now want to express equation of lemma 4.2 only with norms of Hilbert scales $||.||_{-a}$ and $||.||_s$. Consequently, applying 18 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:

$$||\hat{\phi} - \phi_0||_{-a} ||\hat{\phi} - \phi_0||^2 \leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{a+3s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{2a}{s+a}}$$

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

$$|||\hat{\phi} - \phi_0||^3 \leq ||\hat{\phi} - \phi_0||_{-a}^{\frac{3s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{3a}{s+a}}$$

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 18 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:

$$\begin{aligned} ||\hat{\phi}_n^\alpha - \phi_0||_{-a}^2 + \alpha_n ||\hat{\phi}_n^\alpha - \phi_0||_s^2 \quad \leq \quad & \delta^2 + \delta ||\hat{\phi} - \phi_0||_{-a} + \alpha(||\hat{\phi} - \phi_0||_{-a}^{\frac{\beta-s}{a+s}} ||\hat{\phi} - \phi_0||_s^{\frac{a+2s-\beta}{a+s}}) \\ & + \gamma \times (\delta ||\hat{\phi} - \phi_0||_{-a}^{\frac{s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{a}{s+a}} + ||\hat{\phi} - \phi_0||_{-a}^{\frac{a+2s}{s+a}} ||\hat{\phi} - \phi_0||_s^{\frac{a}{s+a}}) \end{aligned} \tag{20}$$

Applying 19 four times to the terms of the right hand side of equation 20 we get:

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

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

$$||\hat{\phi}_n^\alpha - \phi_0||_{-a} = O(\delta + \alpha^{\frac{a+s}{2a+3s-\beta}} ||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{2a+3s-\beta}} + (\gamma\delta)^{\frac{a+s}{2a+s}} ||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{2a+s}} + \gamma^{\frac{a+s}{a}} ||\hat{\phi}_n^\alpha - \phi_0||_s) \tag{21}$$

Then we wish to replace this order obtained in 21 for $||\hat{\phi}_n^\alpha - \phi_0||_{-a}$ in the right hand side of equation 20. We use that $x = O(u+v)$ implies that $x^\xi = O(u^\xi + v^\xi)$.

If we use hypothesis -(viii) we get that:

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

and finally that:

$$
\begin{aligned}
||\hat{\phi}_n^\alpha - \phi_0||_s^2 \;=\; & O\Big(\delta^2\alpha^{-1} + \delta\alpha^{\frac{\beta-a-2s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{2a+3s-\beta}} \\
& +[\delta^{\frac{\beta-s}{a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a+2s-\beta}{a+s}} + \alpha^{\frac{\beta-s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2(\beta-2s-a)}{\beta-2a-3s}}] \\
& +\Big\{\gamma^{\frac{a+s}{2a+s}}\delta^{\frac{3a+2s}{2a+s}}\alpha^{-1}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{2a+s}} + (\gamma\delta)^{\frac{\beta-s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2s+2a-\beta}{2a+s}} \\
& +(\gamma^{\frac{a+s}{a}}\delta\alpha^{-1} + \gamma^{\frac{\beta-s}{a}})||\hat{\phi}_n^\alpha - \phi_0||_s + \alpha^{-1}\gamma\delta^{\frac{a+2s}{a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{a}{a+s}} \\
& \gamma\delta\alpha^{\frac{\beta-2a-2s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a+2s-\beta}{2a+3s-\beta}} + \alpha^{-1}(\gamma\delta)^{\frac{2a+2s}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{2a}{2a+s}} \\
& \gamma\alpha^{\frac{\beta-a-s}{2a+3s-\beta}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{4s+3a-2\beta}{2a+3s-\beta}} + \gamma^{\frac{3a+3s}{2a+s}}\alpha^{\frac{s-a}{2a+s}}||\hat{\phi}_n^\alpha - \phi_0||_s^{\frac{3a}{2a+s}}\Big\}\Big)
\end{aligned} \tag{22}
$$

The first part of the right hand side of equation 22 is identical to the case of standard nonlinear inverse problem, however a rather complicated term depending on $\gamma$ appears. When we don't know anything about $\gamma_n$ it's difficult to simplify the former expression. However, we can apply the property 19 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 one, 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 $o(\delta^2\alpha^{-1})$ using the assumption -(viii). The term in $||\hat{\phi}_n^\alpha - \phi_0||_s$leads to a term that is only a $O(\gamma^{\frac{2(\beta-s)}{a}})$. Finally, we obtain (after having taken the root of every expression) :

$$
\begin{aligned}
||\hat{\phi}_n^\alpha - \phi_0||_s \;=\; & O(\delta\alpha^{-\frac{1}{2}} + \delta^{\frac{2a+3s-u}{3a+4s-u}}\alpha^{\frac{u-a-2s}{3a+4s-u}} + \delta^{\frac{u-s}{a+u}} + \alpha^{\frac{u-s}{2a+2s}} \\
& +(\gamma\delta)^{\frac{(\beta-s)}{2a+\beta}} + \gamma^{\frac{(\beta-s)}{a}} \\
& +(\gamma\delta)^{\frac{(2a+3s-\beta)}{2a+4s-\beta}}\alpha^{\frac{(\beta-2a-2s)}{2a+4s-\beta}} \\
& +\gamma^{\frac{(2a+3s-\beta)}{a+2s}}\alpha^{\frac{(\beta-a-s)}{a+2s}} + \gamma^{\frac{3(a+s)}{a+2s}}\alpha^{\frac{(s-a)}{a+2s}})
\end{aligned} \tag{23}
$$

The choice of $\alpha$ must be made by examining the first part of the expression and then, to control for conditions on $\gamma$ concerning the second part. Then, the good choice for $\alpha$ is:

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

This quantity replaced in the first line of expression 23 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 $\gamma$, with $\alpha$ replaced by 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 condition $\gamma^{\frac{a+\beta}{a}}\delta^{-1}$.

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

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

the former condition 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 know if it's a $o(\delta^{\frac{\beta-s}{a+\beta}})$ we estimate:

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

so we still get the desired convergence.

Studying the fourth term, $\gamma^{\frac{3s-\beta+2a}{a+2s}}\delta^{\frac{-2(s+a)(-\beta+a+s)}{(a+\beta)(a+2s)}}$ must be a $o(\delta^{\frac{\beta-s}{a+\beta}})$. 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 $2a + 3s - beta$ and $a + 2s$ are supposed to be positive.

For the fifth term, $\gamma^{\frac{3(a+s)}{a+2s}}\delta^{\frac{2(a+s)(s-a)}{(a+\beta)(a+2s)}}$ must be $o(\delta^{\frac{\beta-s}{a+\beta}})$. 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 $\beta$. Then the condition $\gamma^{\frac{a+\beta}{a}}\delta^{-1} \to 0$ is stronger and this is verified.

□

### A.6 Proof of lemma 4.4

We study the problem 16. All the norms and the scalar products will be relative to $L^2$-spaces, and we will note $\hat{\phi}$, $\alpha$ and $\delta$ instead of $\hat{\phi}_n^\alpha$, $\alpha_n$ and $\delta_n$ and take $\phi^* = 0$ for sake of simplicity. As in the demonstration of the lemma 4.3 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 reexpress 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 that :

$$< w, (T'_{\phi_0} - \hat{T}'_{\phi_0})(\hat{\phi} - \phi_0) > \quad \leq \quad ||w|| \quad ||(T'_{\phi_0} - \hat{T}'_{\phi_0})(\hat{\phi} - \phi_0)|| \quad \leq \quad \gamma_n||w||||\hat{\phi} - \phi_0||$$

$$< w, \hat{T}(\hat{\phi}) > \quad \leq \quad ||w|| \quad ||\hat{T}(\hat{\phi})||$$

$$- < w, \hat{T}(\phi_0) + \hat{r} > \quad \leq \quad ||w||||\hat{T}(\phi_0) + \hat{r}|| \quad \leq \quad ||w||(\delta + ||\hat{r}||) \quad \leq \quad ||w||(\delta + C||\hat{\phi} - \phi_0||^2)$$

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 :

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

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 $\alpha$ 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, it's easy to conclude that $\hat{T}(\hat{\phi})|| = O(\delta_n)$.

□

# References

[Alt, 1992] Alt, H. (1992). Lineare Funktional Analysis. *Springer Verlag, Berlin*.

[Brémaud, 1991] Brémaud, P. (1991). Point Processes and Queues, Martingale Dynmamics. *Springer Verlag*.

[Carrasco and an Eric Renault, 2003] Carrasco, M. and an Eric Renault, J.-P. F. (2003). Estimation of the solution of an Integral Equation of the Second Kind.

[Chernozhukov et al., 2008] Chernozhukov, V., Galgliardini, P., and Scaillet, O. (2008). Nonparametric Instrumental Variable, Estimation of Quantile Structural Effects. *Working Paper*.

[Darolles et al., 2008] Darolles, S., Florens, J.-P., and Renault, E. (2008). Nonparametric Instrumental Regression.

[Engl et al., 1996] Engl, H., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*, volume 375.

[Engle et al., 1983] Engle, R., Hendry, D., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51(2):277–304.

[Florens, 2002] Florens, J.-P. (2002). Inverse Problems and Structural Econometrics: The Example of Instrumental Variables.

[Florens and Fougère, 1996] Florens, J.-P. and Fougère, D. (1996). Noncausality in Continuous Time. *Econometrica*, 64(5):1195–1212.

[Florens and Mouchart, 1982] Florens, J.-P. and Mouchart, M. (1982). A Note On Noncausality. *Econometrica,* 50(3):583–592.

[Kaltenbacher et al., ] Kaltenbacher, B., Neubauer, A., and Scherzer, O. Iterative Regularization Methods for Nonlinear Ill-Posed Problems.

[Karr, 1991] Karr, A. (1991). Point Processes And Their Statistical Inference. *Marcel Dekker Inc., New York and Basel*.

[Newey and Powell, 2003] Newey, W. and Powell, J. (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71:1565–1578.

[Protter, 2003] Protter, P. (2003). Stochastic Integration and Differential Equations. *Springer Verlag*.