Paris-Saclay : Université d'Évry

« *ATT* » et Contre-factuelle dans le cas discret

Leyth Akrout

Automne 2021

Rapport de Stage dans le cadre de l'obtention d'un diplôme de Master

Juillet - Novembre 2021

Tuteur de Stage : Sandrine Katsahian

Superviseur académique : Agathe Guilloux

Établissement : Université d'Évry

Entreprise d'accueil : INSERM

Table des matières

Introduction

Dans le cadre d'une étude clinique longitudinale, la question de l'efficacité du traitement est centrale. Il existe dans ce contexte différents indicateur permettant d'y répondre, l'ATT (average treatment effect on treated) en est un exemple important.
Comme son nom l'indique cet indicateur permet de calculer l'effet moyen d'un traitement sur le groupe des individus qui ont été traités. Dans le cadre d'une étude non randomisée cet indicateur peut être estimer à l'aide de méthodes statistiques.
L'une des méthodes est développée par Gran et al. dans l'article "Estimating the treatment effect on the treated under time-dependent confounding in an application to the Swiss HIV Cohort Study" [3]. Essayons d'illustrer le contexte mathématique et la solution proposée par ces auteurs :
Dans ce contexte longitudinal, pour un individu donné on note $t \to D(t)$ le processus à valeur dans $\{0{,}1\}$ correspondant à l'état du traitement et S la date du début de traitement.
Un individu est défini par différentes variables :

— $Z = (Z_1, \ldots, Z_{d_Z})$ des covariables indépendantes du temps.
— $X(t) = (X_1(t), \ldots, X_{d_X}(t))$ des processus continu dépendant du temps.

On observe $t \to N(t)$ le processus de comptage correspondant au nombres d'évènements d'intérêts avant le temps t dont l'intensité est notée :

$$\mu(t)dt = \mathbb{E}(dN(t) = 1 \mid F_{t-})$$

Avec $(F_t)_{t\geq 0}$ la filtration du passé associé aux variables Z_k et aux processus X_k et N.
La supposition principale du modèle est que l'intensité s'exprime pour tout $t \geq 0$ comme suit :

$$\mu(t) = \alpha_0(t) + \alpha_Z(t)Z + \alpha_X(t)X(t) + \alpha_D(t)D(t)$$

Avec $\alpha_0(t)$,$\alpha_Z(t)$,$\alpha_X(t)$,$\alpha_D(t)$ les paramètres de la régression. Le but étant d'estimer l'ATT, commençons par le définir. Pour ce faire on défini les contre-factuelles $X^{0|S}$

comme étant le processus X si le traitement n'avait pas été administré au temps S et $X^{1|S}$ le processus X après le temps S si le traitement a été administré au temps S. Plus précisément :

— $\forall t < S$, $X^{0|S}(t) = X^{1|S}(t) = X(t)$
— $\forall t \geq S$, $X^{1|S}(t) = X(t)$

De manière analogue on peux définir les intensités respectives, on obtient donc que $\forall t \geq S$:

$$\mu^{1|S}(t) = \alpha_0(t) + \alpha_Z(t)Z + \alpha_X(t)X^{1|S}(t) + \alpha_D(t)D(t)$$
$$\mu^{0|S}(t) = \alpha_0(t) + \alpha_Z(t)Z + \alpha_X(t)X^{0|S}(t)$$

Ainsi l'ATT étant la moyenne de l'effet du traitement sur le traité on obtient que $\forall t \geq S$:

$$\text{ATT}(t) = \mathbb{E}_S(\mu^{1|S}(t) - \mu^{0|S}(t)) = \alpha_D(t) + \alpha_X(t)\Big[\mathbb{E}_S(X^{1|S}(t) - X^{0|S}(t))\Big]$$

La seule inconnu pour calculer l'ATT est donc $\mathbb{E}_S(X^{0|S}(t))$. Dans le cadre d'une étude randomisé, le groupe témoin nous permet de calculer aisément cette espérance et donc l'ATT.
Dans un contexte non randomisé le processus $(X^{0|S}(t))_{t\geq S}$ n'est pas observé, par contre le processus $(X^{0|S}(t))_{t<S}$ est observée et vaut $(X(t))_{t<S}$.
L'idée de l'article [3] est de supposer le processus $(X^{0|S}(t))_{t\geq 0}$ suit un modèle spécifique (un modèle VAR vector autoregression est choisit dans l'article) d'estimer les paramètres de ce modèle sur ce que l'on observe c'est à dire $(X^{0|S}(t))_{t<S}$ et ensuite de simuler $(X^{0|S}(t))_{t\geq S}$.
Cette idée est inspirée des incréments linéaire de Farewell dans le cadre de données manquantes.
Cela permet de calculer $\mathbb{E}_S(X^{0|S}(t))$ et donc en particulier de trouver un estimateur de l'indicateur d'intérêt ici l'ATT. Une restriction de l'article [3] est qu'il ne concerne que des processus X_k à espace d'état continu.

Chapitre 1

Présentation de l'Entreprise

Le stage que j'ai réalisé s'est fait entre l'INSERM, les Hôpitaux de Paris en particulier l'hôpital européen Georges-Pompidou, le campus Paris-Saclay ou se situe l'Université ou j'ai fais mon Master l'université d'Évry et l'Université Paris-Descartes.

1.1 L'INSERM

Créé en 1964, l'Inserm est un établissement public à caractère scientifique et technologique, placé sous la double tutelle du ministère de la Santé et du ministère de la Recherche. Dédié à la recherche biologique, médicale et à la santé humaine, il se positionne sur l'ensemble du parcours allant du laboratoire de recherche au lit du patient. Sur la scène internationale, il est le partenaire des plus grandes institutions engagées dans les défis et progrès scientifiques.
L'INSERM travaille avec les hôpitaux Français et internationaux en investissant dans la recherche clinique sur le terrain mais aussi avec des Universités pour le recherche théorique et appliquée en santé.

1.2 L'APHP

L'Assistance Publique - Hôpitaux de Paris (APHP) est le centre hospitalier universitaire de la région Île-de-France et le premier CHU de France. A ce titre elle assure des missions de soins, d'enseignement, de recherche, de prévention, d'éducation à la santé et d'aide médicale urgente. Elle est implantée dans une zone urbaine de 11,5 millions de personnes, et emploie 90000 personnes. Elle se compose de 12 groupes hospitaliers, d'une structure d'hospitalisation à domicile, et de services généraux.

1.3 L'hôpital George-Pompidou

L'hôpital européen Georges-Pompidou (HEGP) est un hôpital de l'Assistance publique - hôpitaux de Paris de portée internationale. Dernier né des grands hôpitaux parisiens, l'hôpital, qui ouvre en 2000, reçoit le nom de l'ancien président de la République Georges Pompidou. Il est réalisé dans les années 1990 sur les plans de l'architecte Aymeric Zublena.
Le Pôle du cœur, le Pôle Cancérologie et le Pôle Urgences Réseau constituent les trois pôles cliniques de l'HEGP. L'HEGP est l'un des hôpitaux les plus performants d'Europe, qui s'illustre entre autres grâce à ses services de chirurgie cardiaque et de cardiologie.
De nombreuses équipes de recherche travaillent sur place dans différents domaines de la santé. En particulier l'unité de recherche clinique dirigé par Madame Arnoux Armelle est basée à l'HEGP.
Je tiens à remercier Madame Armelle Arnoux et toute son équipe pour leur accueil chaleureux et leur gentillesse. Qui m'ont permis lors de ce stage de travailler dans leurs locaux. Je tiens aussi à remercier particulièrement Madame Juliette Murris qui a bien voulu me prêter son bureau.

1.4 L'Université Paris-Descartes

L'université Paris-Descartes dont le nom officiel est Paris-V ést une université française créée en 1971. Elle est l'une des treize universités parisiennes, héritières de l'université de Paris et de la Sorbonne. Elle est pluridisciplinaire et fait partie de la communauté d'universités et établissements (Comue) université Sorbonne-Paris-Cité. Elle se définit comme « l'université des sciences de l'homme et de la santé».
L'université a fusionné le 1^{er} janvier 2020 avec l'université Paris-Diderot et l'Institut de physique du globe pour donner naissance à Université de Paris.
J'ai travailler lors de mon stage plus particulièrement à l'UFR de médecine de Paris-Descartes situé au 15 rue des Cordelier. Je tiens à remercier chaleureusement Madame Camille Nevoret qui m'a accorder de son temps et m'a aidé à comprendre certaines partie théorique et plus globalement mon rôle lors de ce stage. Les différents échanges que nous avons eu se faisaient dans cet UFR.
Enfin je tiens à remercier Madame Agathe Guilloux pour son temps, ses explications, ses conseils ,la relecture du rapport, ses remarques et l'opportunité offerte en proposant ce stage.

Chapitre 2

Présentation du Sujet

2.1 Mise en évidence du Biais

Madame Agathe Guilloux et Madame Camille Nevoret ont mis en évidence un biais dans l'article [3] "Estimating the treatment effect on the treated under time-dependent confounding in an application to the Swiss HIV Cohort Study" de Gran et al dans l'estimateur de l'ATT.
Nous allons essayer dans cette section de décrire succinctement ce biais, un article de Camille Nevoret non encore publié concernera ce travail réalisé par les deux auteurs (G-N) cité plus haut.
Pour ce faire revenons à l'intensité de notre processus de comptage :

$$\mu(t) = \alpha_0(t) + \alpha_Z(t)Z + \alpha_X(t)X(t) + \alpha_D(t)D(t)$$

On peux écrire l'intensité différemment en introduisant les processus $X^{0|S}(t))_{t\geq 0}$ et $X^{1|S}(t))_{t\geq 0}$ ainsi :

$$\mu(t) = \alpha_0(t) + \alpha_Z(t)Z + \alpha_X(t)X^{0|S}(t) + d^S(t)D(t)$$

Avec :

$$d^S(t) = \alpha_D(t) + \alpha_X(t)\big[X^{1|S}(t) - X^{0|S}(t)\big]$$

En effet :

— Si $t < S$ alors $D(t) = 0$ et $X^{0|S}(t) = X(t)$ le terme en $d^S(t)$ n'apparaît pas dans l'intensité on a donc égalité.

— Si $t \geq S$ alors $D(t) = 0$ et $X^{1|S}(t) = X(t)$ de plus les deux termes $\alpha_X(t)X^{0|S}(t)$ se simplifient et on à bien égalité.

On peux maintenant estimer le vecteur des paramètres $A(t) = (\alpha_0(t),\alpha_Z(t),\alpha_X(t),d^S(t))^T$ de cette régression linéaire par la méthode des moindres carrés en fonction de la variable $W(t) = (1,Z(t),X^{0|S}(t),D(t))^T$.
Comme vu dans l'introduction pour $t \geq S$ la variable $X^{0|S}(t)$ n'est pas observé, par

conséquent la variable $W(t)$ ne l'est pas non plus. Elles sont générer à l'aide d'un modèle choisi en fonction du processus X, dans le cadre de l'article [3] initial de Gran et al. et de l'article de Guilloux et Nevoret, les processus étant à espace d'état continue, le modèle choisi est un modèle VAR.
Si on note $\widetilde{X}^{0|S}$ le processus contre-factuelles du processus X on obtient $\tilde{W}$ qui vérifie :

$$W(t) = \widetilde{W}(t) + \xi(t)$$

Avec $\xi(t) = (0_1, 0_{d_Z}, \epsilon(t), 0_1)^T$ et $\epsilon(t)$ l'erreur $X^{0|S}(t) - \widetilde{X}^{0|S}(t)$ dépendant du modèle choisit (ici du modèle VAR).
Si on note $r_n(A,W)$ le risque quadratique recherché, Madame Agathe Guilloux et Madame Camille Nevoret ont montré que :

$$r_n(A,W) = r_n(A,\widetilde{W}) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_i} \alpha_X(t)^T \mathbb{E}\big[\epsilon(t)\epsilon(t)^T\big]\alpha_X(t)dt$$

Avec n le nombre d'individus et τ_i le dernier temps ou l'on observe l'individu i.
De plus pour un individu i dans le contexte d'un modèle VAR(1) on a que :

$$X_i^{0|S}(t) = \beta_0^i + \Pi X_i^{0|S}(t-1) + \omega_i(t)$$

Avec Π la matrice des coefficients et ω_i le bruit blanc centré non observable pour $t \geq S$ de matrice de variance-covariance $\Sigma_t = \Sigma$ constante dans le temps.
Dans ce contexte les auteurs ont montré que l'on peux écrire $\mathbb{E}\big[\epsilon(t)\epsilon(t)^T\big]$ de manière explicite pour $t \geq S$:

$$\mathbb{E}\big[\epsilon(t)\epsilon(t)^T\big] = \sum_{j=0}^{t} \Pi^j \Sigma \Pi^{jT}$$

Dans un contexte pratique, les matrices Σ et Π est à estimer à l'aide de la partie observable du processus $X_i^{0|S}(t) = X(t)$ c'est à dire pour les temps $t < S$.
Si on note ces deux matrices $\widehat{\Sigma}$ et $\widehat{\Pi}$, un estimateur non biaisé du risque quadratique s'écrit :

$$\widehat{r_n}(A,W) = r_n(A,\widetilde{W}) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_i}\sum_{j=0}^{t} \alpha_X(t)^T \widehat{\Pi}^j \widehat{\Sigma} \widehat{\Pi}^{jT} \alpha_X(t)dt$$

Finalement on peux estimer les paramètres $(\alpha_0(t), \alpha_Z(t), \alpha_X(t), d^S(t))$ de la régression noté $(\widehat{\alpha_0}(t), \widehat{\alpha_Z}(t), \widehat{\alpha_X}(t), \widehat{d^S}(t))$ et donc finalement calculer un estimateur de l'ATT, pour tout $t \geq S$:

$$\widehat{\mathrm{ATT}}(t) = \widehat{\alpha_D}(t) + \widehat{\alpha_X}(t)\Big[\mathbb{E}_S(X^{1|S}(t) - \widetilde{X}^{0|S}(t))\Big] = \mathbb{E}_S\Big[\widehat{d^S}(t)\Big]$$

2.2 But du stage

Le contexte étant posé, le but de ce stage est de généraliser ce qui a été fait dans l'article initial [3] de Gran et al. et dans l'article de Madame Guilloux et Madame Nevoret. On rappelle que ce qui a été décrit précédemment à été défini pour X un processus à espace d'état continu.
Plus précisément le but est de modifier l'espace d'état du processus X et d'arriver à définir l'ATT et le biais décrit plus haut pour un processus à espace d'état discret.
Pour ce faire nous pouvons suivre l'article initial [3], il nous faut par contre trouver un modèle discret permettant de générer le processus contre-factuelle $(X^{0|S}(t))_{t\geq 0}$ non observé pour $t \geq S$ pour pouvoir estimer l'ATT.
Il faudra ensuite avec ce modèle décrire et estimer le biais.
Pour ce faire Madame Agathe Guilloux m'avait suggéré de considérer des modèles du type INGARCH(p,q) (Integer Generalized Autoregressive Conditional Heteroskedasticity) Poissonien qui sont une possible généralisation des modèles auto-régressif AR(p). Ces modèles permettent dans le cas de la dimension 1 de générer un processus à espace d'état discret discret dépendant de p valeurs antérieures et de q moyennes conditionnelles antérieures.
Plus précisément on dit que $(X_t)_{t\geq 0}$ suit un modèle INGARCH(p,q), s'il existe β_0,b_1,b_2,...,b_p,c_1,c_2,...,$c_q \in \mathbb{R}^+$ tel que pour tout $t \geq max(p,q)$:

$$\lambda_t = \beta_0 + \sum_{i=1}^{p} b_i X_{t-i} + \sum_{i=1}^{q} c_i \lambda_{t-i}$$

Avec $\lambda_t = \mathbb{E}(X_t \mid F_{t-1})$ et $F_t = \sigma(X_{t-j},\lambda_0 \ / \ j = 1,..,t)$, de tel sorte que $X_t \mid F_{t-1}$ suive une loi de Poisson de paramètre λ_t.
Certains mathématiciens dont Rydberg and Shephard [6], Streett [1] ou encore Ferland et al. [2] se sont penchés sur la question de la stationarité et l'ergodicité de ce genre de modèle et ont montré que si $\sum_{i=1}^{p} b_i + \sum_{i=1}^{q} c_i < 1$ alors le modèle est stationnaire et ergodique.
Pour réaliser le travail demandé il nous faut généraliser ce genre de modèle et de résultats au cas de la dimension d, trouver un moyen d'estimer les paramètres du modèle et enfin écrire le biais dans le cas d'un processus à espace d'état discret et de dimension d.
Le travail réalisé ici commence par se poser la question de la stationnarité et l'ergodicité d'un tel modèle, dans une seconde partie elle traite la question du biais en donnant une écriture explicite de ce dernier. Enfin une troisième et dernière partie vise à estimer les paramètres d'un tel modèle.
La question étant à visée applicatives, l'ecriture d'un programme sous R est en cours de réalisation.

Chapitre 3

Présentation du travail réalisé

3.1 Modèle VINGARCH

On commence comme annoncé précédemment par étendre le modèle INGARCH(p,q) à un modèle vectoriel que nous appelons VINGARCH(p,q).

Definition 1 On dit que $(X_t)_{t\geq 0}$ suit un modèle VINGARCH(p,q) Poissonnien s'il existe un vecteur $\beta_0 \in \mathbb{R}^d$ et des matrices $B_1,..B_p,C_1,..,C_q \in M_d(\mathbb{R}_+)$ tel que pour tout $t \geq max(p,q)$:

$$\lambda_t = \beta_0 + \sum_{i=1}^{p} B_i X_{t-i} + \sum_{i=1}^{q} C_i \lambda_{t-i}$$

avec $\lambda_t^i = \mathbb{E}(X_t^i \mid F_{t-1})$ et $F_t = \sigma(X_{t-j}^i,\lambda_0^i \ / \ j = 1,..,t, i = 1,..,d)$, de tel sorte que $X_t^i \mid F_{t-1}$ suive une loi de Poisson de paramètre λ_t^i et que $X_t^i \perp\!\!\!\perp X_t^j \mid F_{t-1}$, pour tout $t \geq 0$ et tout $i,j \in [\![1,d]\!]$

Remarque 1 : Les λ_t pour $t > 0$ n'apparaissent pas dans la tribu F_t car il serait redondant de les ajouter. En effet $\lambda_t^i = \mathbb{E}(X_t^i \mid F_{t-1})$ étant mesurable selon F_{t-1}, de plus $F_{t-1} \subseteq F_t$, on peux le montrer par récurrence.
Pour t=0 on note : $\widetilde{F_0} = \sigma(X_0^i,\lambda_1^i,\lambda_0^i \ /i = 1,..,d)$ on veux montrer que $\widetilde{F_0} = F_0$.
Il est clair que $F_0 \subseteq \widetilde{F_0}$, de plus $(\lambda_0^i)^{-1}(A) \in F_0$ pour tout ensemble A mesurable par rapport à la tribu de Lebesgue. Donc $\widetilde{F_0} \subseteq F_0 \implies \widetilde{F_0} = F_0$. On peux étendre ce raisonnement par récurrence, ainsi $F_t = \sigma(X_{t-j}^i,\lambda_0^i \ / \ j = 1,..,t, i = 1,..,d)$.

Pour prouver la stationnarité, l'ergodicité et par la suite le biais de notre processus nous allons utiliser la représentation matricielle suivante :

$$\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix} = \begin{pmatrix} \beta_0 \\ 0 \\ 0 \\ \vdots \\ \beta_0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \\ \lambda_{t-1} \\ \vdots \\ \lambda_{t-q+1} \\ \lambda_{t-q} \end{pmatrix} + \begin{pmatrix} \omega_t \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Avec ω_t l'erreur d'espérance nulle entre X_t et l'intensité λ_t de la loi de Poisson et M la matrice suivante :

$$\begin{pmatrix} B1 & B2 & \dots & B_p & C_1 & C_2 & \dots & C_q \\ Id_d & 0_d & \dots & 0_d & 0_d & 0_d & \dots & 0_d \\ 0_d & Id_d & \dots & 0_d & 0_d & 0_d & \dots & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ B1 & B2 & \dots & B_p & C_1 & C_2 & \dots & C_q \\ 0_d & 0_d & \dots & 0_d & Id_d & 0_d & \dots & 0_d \\ 0_d & 0_d & \dots & 0_d & 0_d & Id_d & \dots & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_d & 0_d & \dots & 0_d & 0_d & \dots & Id_d & 0_d \end{pmatrix}$$

Proposition 1 *On a les égalités suivantes :*

(i) $\mathbb{V}(\omega_t) = diag(\mathbb{E}(X_t))$

(ii) $Cov(\omega_{t_1},\omega_{t_2}) = 0_d$ *pour tout* $t_1 \neq t_2, t_1,t_2 \geq 0$

(iii) $Cov(X_t,\omega_{t+h}) = 0_d$ *pour tout* $h >$ *0 et* $t \geq 0$

Preuve :
(i) Soient i,j $\in [\![1,d]\!]$, i$\neq$j et $t \geq 0$, on a que :

$$\mathbb{E}(\omega_t^i\omega_t^j \mid F_{t-1}) = \mathbb{E}((X_t^i - \lambda_t^i)(X_t^j - \lambda_t^j) \mid F_{t-1}) = \mathbb{E}(X_t^iX_t^j - X_t^i\lambda_t^j - X_t^j\lambda_t^i + \lambda_t^i\lambda_t^j) \mid F_{t-1}) =$$

$$= \mathbb{E}(X_t^iX_t^j \mid F_{t-1}) - \mathbb{E}(X_t^i\lambda_t^j \mid F_{t-1}) - \mathbb{E}(X_t^j\lambda_t^i \mid F_{t-1}) + \mathbb{E}(\lambda_t^i\lambda_t^j) \mid F_{t-1})$$

Comme $X_t^i \perp\!\!\!\perp X_t^j \mid F_{t-1}$ et que λ_t est mesurable par rapport à F_{t-1} :

$$\mathbb{E}(\omega_t^i\omega_t^j \mid F_{t-1}) = \lambda_t^i\lambda_t^j - \lambda_t^j\mathbb{E}(X_t^i \mid F_{t-1}) - \lambda_t^i\mathbb{E}(X_t^j \mid F_{t-1}) + \lambda_t^i\lambda_t^j = = 2\lambda_t^i\lambda_t^j - 2\lambda_t^i\lambda_t^j = 0$$

Donc la matrice $\mathbb{V}(\omega_t)$ est diagonale, de plus :

$$\mathbb{E}((\omega_t^i)^2 \mid F_{t-1}) = \mathbb{E}((X_t^i)^2 \mid F_{t-1}) - 2\mathbb{E}(X_t^i\lambda_t^i \mid F_{t-1}) + \mathbb{E}((\lambda_t^i)^2) \mid F_{t-1})$$

or

$$\mathbb{E}((\lambda_t^i)^2 \mid F_{t-1}) = \mathbb{E}(\mathbb{E}(X_t^i \mid F_{t-1})\lambda_t^i \mid F_{t-1}) = \mathbb{E}(\mathbb{E}(X_t^i\lambda_t^i \mid F_{t-1}) \mid F_{t-1}) = \; = \mathbb{E}(X_t^i\lambda_t^i \mid F_{t-1})$$

donc

$$\mathbb{E}((\omega_t^i)^2 \mid F_{t-1}) = \mathbb{E}((X_t^i)^2 \mid F_{t-1}) - \mathbb{E}(X_t^i\lambda_t^i \mid F_{t-1}) = \; = \mathbb{V}(X_t^i \mid F_{t-1}) = \mathbb{E}(X_t^i \mid F_{t-1}) = \lambda_t^i$$

car $X_t^i \mid F_{t-1}$ suit une loi de Poisson.
Finalement

$$\mathbb{E}((\omega_t^i)^2) = \mathbb{E}(\mathbb{E}((\omega_t^i)^2 \mid F_{t-1})) = \mathbb{E}(\lambda_t^i) = \mathbb{E}(X_t^i)$$

donc

$$\mathbb{V}(\omega_t) = \mathrm{diag}(\mathbb{E}(X_t))$$

(ii) Soient i,j $\in [\![1,d]\!]$, et $t_1,t_2 \geq 0$, $t_1 > t_2$, on a que :

$$\mathbb{E}(\omega_{t_1}^i\omega_{t_2}^j \mid F_{t_1-1}) = \mathbb{E}(X_{t_1}^iX_{t_2}^j \mid F_{t_1-1}) - \mathbb{E}(X_{t_1}^i\lambda_{t_2}^j \mid F_{t_1-1}) - \mathbb{E}(\lambda_{t_1}^iX_{t_2}^j \mid F_{t_1-1}) + \mathbb{E}(\lambda_{t_1}^i\lambda_{t_2}^j) \mid F_{t_1-1}) =$$

$$= X_{t_2}^j\mathbb{E}(X_{t1}^i \mid F_{t_1-1}) - \lambda_{t2}^j\mathbb{E}(X_{t_1}^i \mid F_{t_1}) - X_{t2}^j\mathbb{E}(\lambda_{t_1}^i \mid F_{t_1}) + \lambda_{t_1}^i\lambda_{t_2}^j = X_{t_2}^j\lambda_{t_1}^i - \lambda_{t_2}^j\lambda_{t_1}^i - X_{t_2}^j\lambda_{t_1}^i + \lambda_{t_1}^i\lambda_{t_2}^j = 0$$

Donc

$$Cov(\omega_{t_1},\omega_{t_2})_{ij} = \mathbb{E}(\omega_{t_1}^i\omega_{t_1}^j) = 0$$

(iii) Soient i,j $\in [\![1,d]\!]$ et $t,h \geq 0$, on a :

$$\mathbb{E}(X_t^i\omega_{t+h}^j \mid F_{t+h-1}) = \mathbb{E}(X_t^iX_{t+h}^j - X_t^i\lambda_{t+h}^j \mid F_{t+h-1}) = X_t^i\mathbb{E}(X_{t+h}^j \mid F_{t+h-1}) - X_t^i\mathbb{E}(\lambda_{t+h}^j \mid F_{t+h-1}) =$$

$$= X_t^i(\mathbb{E}(X_{t+h}^j \mid F_{t+h-1}) - \lambda_{t+h}^j) = 0$$

Donc $Cov(X_t,\omega_{t+h}) = 0_d$ pour tout h > 0 et $t \geq 0$. ■

Notations : Soient Vec : $M_{s,k}(\mathbb{R}) \to \mathbb{R}^{sk}$ l'application qui à une matrice associe le vecteur composé de la succession de ces colonnes et le produit
$\otimes : M_{s,k}(\mathbb{R}) \times M_{m,n}(\mathbb{R}) \to M_{sm,kn}(\mathbb{R})$ dit de Kronecker entre matrices.

Proposition 2 *Soit $(X_t)_{t\geq 0}$ suivant un modèle VINGARCH(p,q) tel que :*

(i) Les valeurs propres de la matrice M sont de module strictement plus petite que 1

(ii) Pour $t \in [\![0,max(p,q)-1]\!]$:

$$\mathbb{E}(X_t) = (Id_d - \sum_{i=1}^{p} B_i - \sum_{i=1}^{q} C_i)^{-1}\beta_0$$

(iii) Pour $t = max(p,q) - 1$ *:*

$$Vec(\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})) = (Id - M \otimes M)^{-1} \, Vec(\begin{pmatrix} diag(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})$$

Alors la série $(X_t)_{t\geq 0}$ *est stationnaire et ergodique pour la moyenne.*

Preuve :

I) Stationnarité :

Pour $t \geq max(p+q)$, par la supposition (ii) on a que :

$$\mathbb{E}(X_t) = \beta_0 + \sum_{i=1}^{p} B_i (Id_d - \sum_{i=1}^{p} B_i - \sum_{i=1}^{q} C_i)^{-1} \beta_0 + \sum_{i=1}^{q} C_i (Id_d - \sum_{i=1}^{p} B_i - \sum_{i=1}^{q} C_i)^{-1} \beta_0 =$$

$$= (Id_d - \sum_{i=1}^{p} B_i - \sum_{i=1}^{q} C_i + \sum_{i=1}^{p} B_i + \sum_{i=1}^{q} C_i)(Id_d - \sum_{i=1}^{p} B_i - \sum_{i=1}^{q} C_i)^{-1} \beta_0$$

. On obtient donc bien une espérance constante :

$$\forall t \geq 0, \ \mathbb{E}(X_t) = (Id_d - \sum_{i=1}^{p} B_i - \sum_{i=1}^{q} C_i)^{-1} \beta_0$$

Concernant la variance on reprend l'écriture matricielle.
Par la supposition (iii) pour $t = max(p,q) - 1$ on a que :

$$\mathrm{Vec}(\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})) = (Id - M \otimes M)^{-1} \mathrm{Vec}(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})$$

$Id - M \otimes M$ est bien un matrice inversible, en effet par supposition M n'admet que des valeurs propres de module strictement plus petites que 1. Or les valeurs propres de $M \otimes M$ sont les produits répétés avec multiplicité des valeurs propres de M, elles sont donc aussi de module strictement plus petites que 1. En particulier 1 n'est pas valeur propre de $M \otimes M$, par conséquent $Id - M \otimes M$ est inversible. Pour $t \geq max(p,q)$:

$$\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = \mathbb{V}(\begin{pmatrix} \beta_0 \\ 0 \\ 0 \\ \vdots \\ \beta_0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \\ \lambda_{t-1} \\ \vdots \\ \lambda_{t-q+1} \\ \lambda_{t-q} \end{pmatrix} + \begin{pmatrix} \omega_t \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix})$$

A l'aide de la proposition 1 on en déduit que :

$$\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = M\mathbb{V}(\begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \\ \lambda_{t-1} \\ \vdots \\ \lambda_{t-q+1} \\ \lambda_{t-q} \end{pmatrix})M^T + \begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})$$

En prenant l'application Vec et en simplifiant en utilisant la compatibilité avec le produit tensoriel on obtient :

$$\mathrm{Vec}(\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})) = M \otimes M \, \mathrm{Vec}(\mathbb{V}(\begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \\ \lambda_{t-1} \\ \vdots \\ \lambda_{t-q+1} \\ \lambda_{t-q} \end{pmatrix})) + \mathrm{Vec}(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}) =$$

$$= M \otimes M \left[(Id - M \otimes M)^{-1} \text{Vec}(\begin{pmatrix} \text{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}) \right] + \text{Vec}(\begin{pmatrix} \text{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}) =$$

$$= [M \otimes M + Id - M \otimes M] \, (Id - M \otimes M)^{-1} \text{Vec}(\begin{pmatrix} \text{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}) =$$

$$= (Id - M \otimes M)^{-1} \text{Vec}(\begin{pmatrix} \text{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}) = \text{Vec}(\mathbb{V}(\begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \\ \lambda_{t-1} \\ \vdots \\ \lambda_{t-q+1} \\ \lambda_{t-q} \end{pmatrix}))$$

On a donc bien que Vec de la variance est constant au cours du temps. Nous pouvons maintenant reconstruire la variance à partir de ce vecteur. En notant K= $\begin{pmatrix} 1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & \dots & 0 \end{pmatrix}$ une matrice de taille $d \times d^2(p+q)^2$ on a que :

$$\forall t \geq 0, \ \text{Vec}(\mathbb{V}(X_t)) = K \otimes K (Id - M \otimes M)^{-1} \text{Vec}(\begin{pmatrix} \text{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

La variance est donc bien constante au cours du temps, refaisons le même travail pour déterminer la fonction de covariance du modèle.

Pour $t \geq max(p,q)$ et h>0 on regarde la covariance suivante :

$$Cov(\begin{pmatrix} X_{t+h} \\ X_{t+h-1} \\ \vdots \\ X_{t+h-p+1} \\ \lambda_{t+h} \\ \vdots \\ \lambda_{t+h-q+2} \\ \lambda_{t+h-q+1} \end{pmatrix} , \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = Cov(\begin{pmatrix} \beta_0 \\ 0 \\ 0 \\ \vdots \\ \beta_0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} X_{t+h-1} \\ X_{t+h-2} \\ \vdots \\ X_{t+h-p} \\ \lambda_{t+h-1} \\ \vdots \\ \lambda_{t+h-q+1} \\ \lambda_{t+h-q} \end{pmatrix} + \begin{pmatrix} \omega_{t+h} \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} , \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})$$

A l'aide de la proposition 1 il vient que :

$$Cov(\begin{pmatrix} X_{t+h} \\ X_{t+h-1} \\ \vdots \\ X_{t+h-p+1} \\ \lambda_{t+h} \\ \vdots \\ \lambda_{t+h-q+2} \\ \lambda_{t+h-q+1} \end{pmatrix}, \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = MCov(\begin{pmatrix} X_{t+h-1} \\ X_{t+h-2} \\ \vdots \\ X_{t+h-p} \\ \lambda_{t+h-1} \\ \vdots \\ \lambda_{t+h-q+1} \\ \lambda_{t+h-q} \end{pmatrix}, \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})$$

On réitère cette opération h-1 fois pour obtenir l'expression suivante :

$$Cov(\begin{pmatrix} X_{t+h} \\ X_{t+h-1} \\ \vdots \\ X_{t+h-p+1} \\ \lambda_{t+h} \\ \vdots \\ \lambda_{t+h-q+2} \\ \lambda_{t+h-q+1} \end{pmatrix}, \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = M^hCov(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}, \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = M^h\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})$$

En appliquant l'application Vec on obtient finalement :

$$\mathrm{Vec}(Cov(\begin{pmatrix} X_{t+h} \\ X_{t+h-1} \\ \vdots \\ X_{t+h-p+1} \\ \lambda_{t+h} \\ \vdots \\ \lambda_{t+h-q+2} \\ \lambda_{t+h-q+1} \end{pmatrix}, \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})) = \mathrm{Vec}(M^h\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix})) =$$

$$= (M^h \otimes Id)\mathrm{Vec}(\mathbb{V}(\begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \\ \lambda_t \\ \vdots \\ \lambda_{t-q+2} \\ \lambda_{t-q+1} \end{pmatrix}) = (M \otimes Id)^h (Id - M \otimes M)^{-1}\mathrm{Vec}(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})$$

Cette quantité ne dépend que de h, en reprenant la matrice K la fonction de covariance Γ admet donc l'expression suivante :

$$\forall h \geq 0,\ \mathrm{Vec}(\Gamma(h)) = K{\otimes}K\,(M{\otimes}Id)^h(Id{-}M{\otimes}M)^{-1}\mathrm{Vec}(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})$$

Remarque 2 : Si l'on remplace la matrice K par la matrice qui permet d'obtenir les mêmes quantités mais pour la série $(\lambda_t)_{t\geq 0}$, on obtient avec le même raisonnement la stationnarité de cette série temporelle.

Nous avons donc bien prouvé la stationnarité du modèle VINGARCH(p,q), traitons maintenant l'ergodicité.

II) Ergodicité :

Nous allons montrer que les séries $(\lambda_t)_{t\geq 0}$ et $(X_t)_{t\geq 0}$ sont ergodique pour la moyenne. Nous allons nous concentrer sur la série $(X_t)_{t\geq 0}$ l'autre cas étant identique. Il nous faut donc montrer pour tout i $\in [\![1,d]\!]$ que :

$$\frac{1}{T}\sum_{t=0}^{T} X_t^i \xrightarrow[T\to\infty]{} \mathbb{E}(X^i)$$

en moyenne quadratique, autrement dit que :

$$\mathbb{E}\left(\left[\frac{1}{T}\sum_{t=0}^{T} X_t^i - \mathbb{E}(X^i)\right]^2\right) \xrightarrow[T\to\infty]{} 0$$

Pour ce faire observons que la condition suivante est suffisante :

$$\mathbb{E}\left(\left[\frac{1}{T}\sum_{t=0}^{T} X_t - \mathbb{E}(X)\right]\left[\frac{1}{T}\sum_{t=0}^{T} X_t - \mathbb{E}(X)\right]^T\right) \xrightarrow[T\to\infty]{} 0$$

En effet ce que l'on recherche correspond à la diagonale, de plus par stationnarité :

$$\mathbb{E}\left(\left[\frac{1}{T}\sum_{t=0}^{T} X_t - \mathbb{E}(X)\right]\left[\frac{1}{T}\sum_{t=0}^{T} X_t - \mathbb{E}(X)\right]^T\right) = \mathbb{V}\left(\frac{1}{T}\sum_{t=0}^{T} X_t\right) = \frac{1}{T^2}\mathbb{V}\left(\sum_{t=0}^{T} X_t\right) =$$

$$= \frac{1}{T^2}\sum_{t=0}^{T}\mathbb{V}\left(X_t\right) + \frac{1}{T^2}\sum_{0\leq t_1<t_2\leq T} Cov\left(X_{t_1},X_{t_2}\right) + \frac{1}{T^2}\sum_{0\leq t_1<t_2\leq T} Cov\left(X_{t_2},X_{t_1}\right)$$

Par stationnarité on obtient que l'espérance recherchée vaut :

$$\frac{T+1}{T^2}\mathbb{V}\left(X\right) + \frac{1}{T^2}\sum_{t=1}^{T} t\Gamma\left(T-t+1\right) + \frac{1}{T^2}\sum_{t=1}^{T} t\Gamma\left(T-t+1\right)^T$$

La convergence de cette matrice vers 0 étant équivalente au fait que ces coefficients tendent vers 0, il revient au même de montrer que le vecteur composé des colonnes de cette matrice tend vers 0 autrement dit que Vec de cette matrice tend vers 0. Pour $l > 0$ on note sans distinction $\|.\|_l$ la norme l sur $\mathbb{R}^k$ ou sur $M_k(\mathbb{R})$, on a donc que :

$$\left\| Vec\left(\frac{T+1}{T^2}\mathbb{V}(X) + \frac{1}{T^2}\sum_{t=1}^{T} t\Gamma(T-t+1) + \frac{1}{T^2}\sum_{t=1}^{T} t\Gamma(T-t+1)^T\right)\right\|_l =$$

$$= \left\| \frac{T+1}{T^2} Vec(\mathbb{V}(X)) + \frac{1}{T^2}\sum_{t=1}^{T} t\mathrm{Vec}(\Gamma(T-t+1)) + \frac{1}{T^2}\sum_{t=1}^{T} t\mathrm{Vec}(\Gamma(T-t+1)^T)\right\|_l \leq$$

$$\leq \frac{T+1}{T^2}\|Vec(\mathbb{V}(X))\|_l + \frac{1}{T^2}\sum_{t=1}^{T} t\,\|\mathrm{Vec}(\Gamma(T-t+1))\|_l + \frac{1}{T^2}\sum_{t=1}^{T} t\left\|\mathrm{Vec}(\Gamma(T-t+1)^T)\right\|_l \leq$$

$$\leq \frac{T+1}{T^2}\|Vec(\mathbb{V}(X))\|_l + \frac{1}{T^2}\sum_{t=1}^{T} t\,\|\mathrm{Vec}(\Gamma(T-t+1))\|_l + \frac{1}{T^2}\sum_{t=1}^{T} t\left\|\mathrm{Vec}(\Gamma(T-t+1)^T)\right\|_l \leq$$

$$\leq \frac{T+1}{T^2}\|Vec(\mathbb{V}(X))\|_l + \frac{1}{T^2}\sum_{t=1}^{T} t\,\|\mathrm{Vec}(\Gamma(T-t+1))\|_l + \frac{1}{T^2}\sum_{t=1}^{T} t\left\|\mathrm{Vec}(\Gamma(T-t+1)^T)\right\|_l \leq$$

$$\leq \frac{T+1}{T^2}\left(\|Vec(\mathbb{V}(X))\|_l + \sum_{t=1}^{T}\|\mathrm{Vec}(\Gamma(T-t+1))\|_l + \sum_{t=1}^{T}\left\|\mathrm{Vec}(\Gamma(T-t+1)^T)\right\|_l\right) =$$

$$\leq \frac{T+1}{T^2}\left(\|Vec(\mathbb{V}(X))\|_l + \sum_{t=1}^{T}\|\mathrm{Vec}(\Gamma(T-t+1))\|_l + \sum_{t=1}^{T}\left\|\mathrm{Vec}(\Gamma(T-t+1)^T)\right\|_l\right) =$$

$$= \frac{T+1}{T^2}\left(\|Vec(\mathbb{V}(X))\|_l + \sum_{h=1}^{T}\|\mathrm{Vec}(\Gamma(h))\|_l + \sum_{h=1}^{T}\left\|\mathrm{Vec}(\Gamma(h)^T)\right\|_l\right)]$$

Or :

$$\|\mathrm{Vec}(\Gamma(h))\|_l = \left\| K\otimes K\,(M\otimes Id)^h (Id - M\otimes M)^{-1}\mathrm{Vec}(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})\right\|_l \leq$$

$$\leq \|K\otimes K\|_l\, \|(M\otimes Id)\|_l^h\, \|(Id - M\otimes M)\|_l^{-1} \left\| \mathrm{Vec}(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix})\right\|_l = Cst_1\,\|(M\otimes Id)\|_l^h$$

Les valeurs propres de $M \otimes Id$ étant les mêmes que celles de M elles sont donc strictement plus petites que 1, on peux donc trouver $l_0 > 0$ tel que :

$$\|(M\otimes Id)\|_{l_0} \leq 1$$

De plus si on note l^* le conjugué de l, l'espace $(M_k(\mathbb{R}),\|.\|_l)$ devient le dual topologique de l'espace $(M_k(\mathbb{R}),\|.\|_{l^*})$ ou l'application transposition est une isométrie.

Par équivalence des normes dans un espace de dimension fini il existe une constante Cst_2 ne dépendant que de l tel que pour toutes matrice M :

$$\|A\|_l \leq Cst_2 \|A\|_{l^*}$$

Ainsi les quantités suivantes :

$$\frac{T+1}{T^2} \|Vec\left(\mathbb{V}\left(X\right)\right)\|_{l_0}$$

$$\frac{T+1}{T^2} \sum_{h=1}^{T} \|\mathrm{Vec}(\Gamma\left(h\right))\|_{l_0} \leq \frac{T+1}{T^2} Cst_1 \sum_{h=1}^{T} \|(M \otimes Id)\|_{l_0}^h$$

$$\frac{T+1}{T^2} \sum_{h=1}^{T} \left\|\mathrm{Vec}(\Gamma\left(h\right)^T)\right\|_{l_0} \leq \frac{T+1}{T^2} Cst_2 \sum_{h=1}^{T} \left\|\mathrm{Vec}(\Gamma\left(h\right)^T)\right\|_{l_0^*}^h \leq \frac{T+1}{T^2} Cst_2 Cst_1 \sum_{h=1}^{T} \|(M \otimes Id)\|_{l_0}^h$$

Convergent vers 0. Finalement on obtient bien que :

$$\frac{1}{T} \sum_{t=0}^{T} X_t^i \xrightarrow[T \to \infty]{} \mathbb{E}(X^i)$$

La série $(X_t)_{t \geq 0}$ est donc ergodique pour la moyenne. Comme mentionné précédemment un raisonnement similaire permettra d'obtenir l'ergodicité pour la série $(\lambda_t)_{t \geq 0}$

■

Remarque 3 : Notre but étant d'apprendre les paramètres de modèle pour une série statistique donnée il nous faudra par la suite écrire le maximum de vraisemblance et optimiser les paramètres qui sont ici les coefficients du vecteur β_0 et des matrices B_i, C_i.

On doit en particulier montrer que la condition (i) de la proposition 2 définit bien un ouvert de l'espace des paramètres de telle sorte à pouvoir optimiser les paramètres sans se soucier des valeurs aux bords.

Autrement dit nous voulons faire de l'optimisation sans contraintes donc montrer que l'ensemble des paramètres dans $\mathbb{R}_+^{d+(p+q)d^2}$ tel que la matrice M n'admet que des valeurs propres de module strictement plus petit que 1 est ouvert.

Regardons de plus près pour $x \in \mathbb{C}^*$ la condition :

$$det(xId - M) = 0 \implies \|x\| < 1$$

Or :

$$det(xId{-}M) = det\begin{pmatrix} xId_d - B1 & -B2 & \dots & -B_p & -C_1 & -C_2 & \dots & -C_q \\ -Id_d & xId_d & \dots & 0_d & 0_d & 0_d & \dots & 0_d \\ 0_d & -Id_d & \dots & 0_d & 0_d & 0_d & \dots & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -B1 & -B2 & \dots & -B_p & xId_d - C_1 & -C_2 & \dots & -C_q \\ 0_d & 0_d & \dots & 0_d & -Id_d & xId_d & \dots & 0_d \\ 0_d & 0_d & \dots & 0_d & 0_d & -Id_d & \dots & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_d & 0_d & \dots & 0_d & 0_d & \dots & -Id_d & xId_d \end{pmatrix}$$

Soit C_f la dernière colonne, on remplace $C_{f-1} \leftarrow xC_{f-1} + C_f$ ainsi :

$$det(xId{-}M) = x^d det\begin{pmatrix} xId_d - B1 & -B2 & \dots & -B_p & -C_1 & -C_2 & \dots & -xC_{q-1} - C_q \\ -Id_d & xId_d & \dots & 0_d & 0_d & 0_d & \dots & 0_d \\ 0_d & -Id_d & \dots & 0_d & 0_d & 0_d & \dots & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -B1 & -B2 & \dots & -B_p & xId_d - C_1 & -C_2 & \dots & -xC_{q-1} - C_q \\ 0_d & 0_d & \dots & 0_d & -Id_d & xId_d & \dots & 0_d \\ 0_d & 0_d & \dots & 0_d & 0_d & -Id_d & \dots & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_d & 0_d & \dots & 0_d & 0_d & \dots & -Id_d & x^2Id_d \end{pmatrix}$$

On remplace $C_{f-1} \leftarrow x^2C_{f-1} + C_f$ et on réitère l'opération en multipliant à chaque étape par x, ainsi :

$$det(xId - M) = x^{qd} det\begin{pmatrix} xId_d - B1 & -B2 & \dots & -B_p & -\sum_{i=1}^{q} x^{q-i}C_i \\ -Id_d & xId_d & \dots & 0_d & 0_d \\ 0_d & -Id_d & \dots & 0_d & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_d & 0_d & \dots & xId_d & 0_d \\ -B1 & -B2 & \dots & -B_p & x^qId_d - \sum_{i=1}^{q} x^{q-i}C_i \end{pmatrix}$$

On continue en remplaçant maintenant $C_{f-2} \leftarrow xC_{f-2} + C_{f-1}$, ainsi :

$$det(xId{-}M) = x^{qd+1} det\begin{pmatrix} xId_d - B1 & -B2 & \dots & -xB_{p-1} - B_p & -\sum_{i=1}^{q} x^{q-i}C_i \\ -Id_d & xId_d & \dots & 0_d & 0_d \\ 0_d & -Id_d & \dots & 0_d & 0_d \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_d & 0_d & \dots & x^2Id_d & 0_d \\ -B1 & -B2 & \dots & -xB_{p-1} - B_p & x^qId_d - \sum_{i=1}^{q} x^{q-i}C_i \end{pmatrix}$$

On remplace $C_{f-2} \leftarrow x^2 C_{f-2} + C_{f-1}$ et on réitère l'opération en multipliant à chaque étape par x, ainsi :

$$det(xId - M) = x^{d(p+q)} det \begin{pmatrix} x^p Id_d - \sum_{i=1}^{p} x^{p-i} B_i & -\sum_{i=1}^{q} x^{q-i} C_i \\ -\sum_{i=1}^{p} x^{p-i} B_i & x^q Id_d - \sum_{i=1}^{q} x^{q-i} C_i \end{pmatrix}$$

Les Matrices $x^p Id_d - \sum_{i=1}^{p} x^{p-i} B_i$ et $-\sum_{i=1}^{p} x^{p-i} B_i$ commutant on peux conclure que le déterminant vaut :

$$x^{d(p+q)} det \left(x^{p+q} Id_d - \sum_{i=1}^{p} x^{p+q-i} B_i - \sum_{i=1}^{q} x^{p+q-i} C_i + \sum_{i=1}^{p} x^{p-i} B_i \sum_{i=1}^{q} x^{q-i} C_i - \sum_{i=1}^{p} x^{p-i} B_i \sum_{i=1}^{q} x^{q-i} C_i \right) =$$

$$= x^{d(p+q)} det \left(x^{p+q} Id_d - \sum_{i=1}^{p} x^{p+q-i} B_i - \sum_{i=1}^{q} x^{p+q-i} C_i \right)$$

Donc finalement la condition sur M devient une condition plus explicite sur nos matrices de départ, pour $x \in \mathbb{C}^*$:

$$det \left(x^{p+q} Id_d - \sum_{i=1}^{p} x^{p+q-i} B_i - \sum_{i=1}^{q} x^{p+q-i} C_i \right) = 0 \implies \|x\| < 1$$

De plus pour x fixé, l'application qui aux coefficients de β_0 et de nos matrices associe la matrice :

$$x^{p+q} Id_d - \sum_{i=1}^{p} x^{p+q-i} B_i - \sum_{i=1}^{q} x^{p+q-i} C_i)$$

est continue. Par continuité du déterminant il vient que l'application qui aux coefficients de β_0 et de nos matrices associe le polynôme en x :

$$det(x^{p+q} Id_d - \sum_{i=1}^{p} x^{p+q-i} B_i - \sum_{i=1}^{q} x^{p+q-i} C_i)$$

est continue. Par le théorème de Rouché l'application qui a un polynôme associe sa racine de module maximal est continue pour la topologie du maximum des modules des coefficients du polynôme sur $\mathbb{C}[x]$.

Finalement l'application f qui aux coefficients de β_0 et de nos matrices associe la racine de module maximal du polynôme :

$$det(x^{p+q} Id_d - \sum_{i=1}^{p} x^{p+q-i} B_i - \sum_{i=1}^{q} x^{p+q-i} C_i)$$

est continue.

Si on note D_1 le disque ouvert de rayon 1 dans $\mathbb{C}$, $f^{-1}(D_1)$ est ouvert par continuité et est précisément l'espace des paramètres que l'on cherche. L'optimisation se fait donc bien dans un espace ouvert donc sans contraintes.

On peux vérifier que dans le cas d=1 (modèle INGARCH(p,q)) on retrouve bien la condition classique que l'on peux retrouver dans les articles [1]-[2]-[6], à savoir pour β_0,b_1,..,b_p,c_1,..,$c_q \geq 0$:

$$\sum_{i=1}^{p} b_i + \sum_{i=1}^{q} c_i < 1$$

Notre condition est que :

$$det(x^{p+q} - \sum_{i=1}^{p} x^{p+q-i} b_i - \sum_{i=1}^{q} x^{p+q-i} c_i) = 0 \implies \|x\| < 1$$

Pour $x \in \mathbb{R}^*$ comme ici le déterminant vaut l'identité, que le polynôme converge vers $+\infty$ quand x tend vers $+\infty$ et qu'il ne s'annule pas pour $x = 1$, il est donc positif en 1, on en déduit donc que :

$$1 - \sum_{i=1}^{p} b_i - \sum_{i=1}^{q} c_i > 0 \Longleftrightarrow \sum_{i=1}^{p} b_i + \sum_{i=1}^{q} c_i < 1$$

Ainsi notre condition implique la condition classique.

3.2 Biais discret

Cette section est à imaginer à la suite de la partie "Example with a vector autoregressive model VAR(1)" de l'article de Camille Nevoret et devrait porter le nom "Example with a count vector model VINGARCH(p,q)".
Soit $(X_i^0(0),X_i^0(1),...,X_i^0(t_i))$ un échantillon de variables discrètes de dimension d d'un individu i avant une date t_i correspondant à la date de traitement de cet individu et $(X_i^1(t_i+1),X_i^1(t_i+2),...,X_i^1(t_i+t_{k_i}))$ ces mêmes variables après traitement.
Nous allons modéliser les variables avant traitement par $(X_i^0(t))_{t\geq 0}$ en supposant que ces dernières suivent un modèle VINGARCH(p,q) de dimension d.
Notre but (cf article C.Nevoret) étant d'estimer l'effet du traitement sur le patient (l'ATT) en suivant la démarche de l'article [3] "Estimating the treatment effect on the treated under time-dependent confounding in an application to the Swiss HIV Cohort Study" de Gran et al., il nous faut simuler les contre-factuelles non observées $(X_i^0(t_i+1),X_i^0(t_i+2),...,X_i^0(t_i+t_{k_i}))$ notées :

$$(\tilde{X}_i^0(0)=X_i^0(0),\tilde{X}_i^0(1)=X_i^0(1),...,\tilde{X}_i^0(t_i)=X_i^0(t_i),\tilde{X}_i^0(t_i+1),\tilde{X}_i^0(t_i+2),...,\tilde{X}_i^0(t_i+t_{k_i}))$$

.

Par supposition la série des variables avant traitement $(X_i^0(t))_{t\geq 0}$ suit un modèle VINGARCH(p,q), on peux donc simuler les contre-factuelles $(\tilde{X}_i^0(t_i+t))_{t>0}$ en reprenant l'écriture matricielle comme suit :

$$\begin{pmatrix} \tilde{X}_i^0(t_i+t) \\ \tilde{X}_i^0(t_i+t-1) \\ \vdots \\ \tilde{X}_i^0(t_i+t-p+1) \\ \tilde{\lambda}_i^0(t_i+t) \\ \vdots \\ \tilde{\lambda}_i^0(t_i+t-q+2) \\ \tilde{\lambda}_i^0(t_i+t-q+1) \end{pmatrix} = \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} \tilde{X}_i^0(t_i+t-1) \\ \tilde{X}_i^0(t_i+t-2) \\ \vdots \\ \tilde{X}_i^0(t_i+t-p) \\ \tilde{\lambda}_i^0(t_i+t-1) \\ \vdots \\ \tilde{\lambda}_i^0(t_i+t-q+1) \\ \tilde{\lambda}_i^0(t_i+t-q) \end{pmatrix}$$

Comme on sait que pour $t \leq t_i, \tilde{X}_i^0(t) = X_i^0(t)$ et que $\tilde{\lambda}_i^0(t) = \lambda_i^0(t)$, l'équation précédente peut être définit par récurrence en commençant par $t=1$ ce qui donne :

$$\begin{pmatrix} \tilde{X}_i^0(t_i+1) \\ X_i^0(t_i) \\ \vdots \\ X_i^0(t_i-p+2) \\ \tilde{\lambda}_i^0(t_i+1) \\ \vdots \\ \lambda_i^0(t_i-q+3) \\ \lambda_i^0(t_i-q+2) \end{pmatrix} = \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} X_i^0(t_i) \\ X_i^0(t_i-1) \\ \vdots \\ X_i^0(t_i-p+1) \\ \lambda_i^0(t_i) \\ \vdots \\ \lambda_i^0(t_i-q+2) \\ \lambda_i^0(t_i-q+1) \end{pmatrix}$$

Et réitérer l'opération pour obtenir les contre-factuelles au temps voulu. Une manière équivalente de définir les contre-factuelles est d'effectuer le calcul suivant qui sera utile pour obtenir le biais :

$$\begin{pmatrix} X_i^0(t_i+t) \\ X_i^0(t_i+t-1) \\ \vdots \\ X_i^0(t_i+t-p+1) \\ \lambda_i^0(t_i+t) \\ \vdots \\ \lambda_i^0(t_i+t-q+2) \\ \lambda_i^0(t_i+t-q+1) \end{pmatrix} = \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} X_i^0(t_i+t-1) \\ X_i^0(t_i+t-2) \\ \vdots \\ X_i^0(t_i+t-p) \\ \lambda_i^0(t_i+t-1) \\ \vdots \\ \lambda_i^0(t_i+t-q+1) \\ \lambda_i^0(t_i+t-q) \end{pmatrix} + \begin{pmatrix} \omega_i^0(t_i+t) \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} =$$

$$= \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M^2 \begin{pmatrix} X_i^0(t_i+t-2) \\ X_i^0(t_i+t-3) \\ \vdots \\ X_i^0(t_i+t-p-1) \\ \lambda_i^0(t_i+t-2) \\ \vdots \\ \lambda_i^0(t_i+t-q) \\ \lambda_i^0(t_i+t-q-1) \end{pmatrix} + M \begin{pmatrix} \omega_i^0(t_i+t-1) \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega_i^0(t_i+t) \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

En réitérant cette opération t fois on obtient que :

$$= \sum_{j=0}^{t-1} M^j \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M^t \begin{pmatrix} X_i^0(t_i) \\ X_i^0(t_i-1) \\ \vdots \\ X_i^0(t_i-p+1) \\ \lambda_i^0(t_i) \\ \vdots \\ \lambda_i^0(t_i-q+2) \\ \lambda_i^0(t_i-q-1) \end{pmatrix} + \sum_{j=0}^{t-1} M^j \begin{pmatrix} \omega_i^0(t_i+t-i) \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Ainsi on peux définir les contre-factuelles à l'aide de cette formule sans les erreurs autrement dit pour $t > 0$:

$$\begin{pmatrix} \tilde{X}_i^0(t_i+t) \\ \tilde{X}_i^0(t_i+t-1) \\ \vdots \\ \tilde{X}_i^0(t_i+t-p+1) \\ \tilde{\lambda}_i^0(t_i+t) \\ \vdots \\ \tilde{\lambda}_i^0(t_i+t-q+2) \\ \tilde{\lambda}_i^0(t_i+t-q+1) \end{pmatrix} = \sum_{j=0}^{t-1} M^j \begin{pmatrix} \beta_0^i \\ 0 \\ 0 \\ \vdots \\ \beta_0^i \\ \vdots \\ 0 \\ 0 \end{pmatrix} + M^t \begin{pmatrix} X_i^0(t_i) \\ X_i^0(t_i-1) \\ \vdots \\ X_i^0(t_i-p+1) \\ \lambda_i^0(t_i) \\ \vdots \\ \lambda_i^0(t_i-q+2) \\ \lambda_i^0(t_i-q-1) \end{pmatrix}$$

Soit $t > 0$ pour calculer le biais discret il nous faut tout d'abord exprimer la variance de la différence des "vrai" contre-factuelle et de celle simulée.

En notant U la matrice suivante $\begin{pmatrix} 1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & \dots & 0 \end{pmatrix}$, de taille $d \times d(p+q)$ et

$\epsilon_i(t)$ une partie du biais recherché (cf article C.Nevoret) on obtient que :

$$\mathbb{V}(\epsilon_i(t)) = \mathbb{E}\left(\epsilon_i(t)\epsilon_i(t)^T\right) = \mathbb{V}(X_i^0(t_i+t) - \tilde{X}_i^0(t_i+t)) =$$

$$\mathbb{V}\Big(U \begin{pmatrix} X_i^0(t_i+t) \\ X_i^0(t_i+t-1) \\ \vdots \\ X_i^0(t_i+t-p+1) \\ X_i^0(t_i+t) \\ \vdots \\ X_i^0(t_i+t-q+2) \\ X_i^0(t_i+t-q+1) \end{pmatrix} - U \begin{pmatrix} \tilde{X}_i^0(t_i+t) \\ \tilde{X}_i^0(t_i+t-1) \\ \vdots \\ \tilde{X}_i^0(t_i+t-p+1) \\ \tilde{\lambda}_i^0(t_i+t) \\ \vdots \\ \tilde{\lambda}_i^0(t_i+t-q+2) \\ \tilde{\lambda}_i^0(t_i+t-q+1) \end{pmatrix}\Big) = \mathbb{V}\Big(\sum_{j=0}^{t-1} U M^j \begin{pmatrix} \omega_i^0(t_i+t-i) \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}\Big)$$

Or d'après les points (i) et (ii) de la proposition 1 on peut en déduire que :

$$\mathbb{V}(\epsilon_i(t)) = \sum_{j=0}^{t-1} UM^j \mathbb{V}\Bigg(\begin{pmatrix} \omega_i^0(t_i+t-i) \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}\Bigg) M^{i^T} U^T = \sum_{j=0}^{t-1} UM^j \begin{pmatrix} \text{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} M^{j^T} U^T$$

On peux donc conclure (avec les notations de l'article de C.Nevoret) que le biais vaut :

$$\text{biais}(A) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_i} \alpha_X(t)^T \sum_{j=0}^{t-1} UM^j \begin{pmatrix} \text{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} M^{j^T} U^T \alpha_X(t) dt$$

Avec A le vecteur des coefficients de la régression linéaire (Additive Hazard Regression), n le nombre d'individus, α_X une partie du vecteur A qui est concerné par le biais et τ_i le dernier temps ou l'on observe l'individu i.

Remarque 4 : Cette remarque qui vise à expliciter encore plus le biais s'applique aussi pour un modèle VAR(p). Comme le biais est réel il est égal à son Vec, de plus l'intégrale est en fait une somme (temps discret) on peux donc intervertir Vec et l'intégrale (linéarité de Vec) et écrire que :

$$\text{biais}(A) = \text{Vec}\Big(\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau_i} \alpha_X(t)^T \sum_{j=0}^{t-1} UM^j \begin{pmatrix} \text{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} M^{j^T} U^T \alpha_X(t) dt\Big) =$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_{t=0}^{\tau_i} \sum_{j=0}^{t-1} \text{Vec}\Big(\alpha_X(t)^T UM^j \begin{pmatrix} \text{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} M^{j^T} U^T \alpha_X(t)\Big) dt =$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_{t=0}^{\tau_i} \sum_{j=0}^{t-1} \alpha_X(t)^T \otimes \alpha_X(t)^T \, U \otimes U \, M^j \otimes M^j \, dt \, \text{Vec}\Big(\begin{pmatrix} \text{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}\Big) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{t=0}^{\tau_i} \alpha_X(t)^T \otimes \alpha_X(t)^T\, U \otimes U \sum_{j=0}^{t-1} \left[M \otimes M\right]^j dt\, \mathrm{Vec}\Bigg(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \Bigg) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\int_{t=0}^{\tau_i} \alpha_X(t)^T \otimes \alpha_X(t)^T\, U \otimes U\, (Id - \left[M \otimes M\right]^t)(Id - M \otimes M)^{-1}\, dt\, \mathrm{Vec}\Bigg(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \Bigg)$$

$$= \int_{t=0}^{\tau_i} \alpha_X(t)^T \otimes \alpha_X(t)^T\, U \otimes U\, (Id - \left[M \otimes M\right]^t)(Id - M \otimes M)^{-1}\, dt\, \mathrm{Vec}\Bigg(\begin{pmatrix} \mathrm{diag}(\mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} X_i)) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \Bigg)$$

En pratique on ne connaît pas les paramètres du modèle autrement dit les coefficients du vecteur β_0^i et des matrices B_j, C_j, il nous faut les estimer. Si on note $\widehat{\beta}_0^i, \widehat{B}_j, \widehat{C}_j$ les estimateurs correspondant et $\widehat{M}$ l'estimateur de la matrice M on obtient, pour le biais, l'estimateur suivant :

$$\widehat{\mathrm{biais}}(A) = \int_{t=0}^{\tau_i} \alpha_X(t)^T \otimes \alpha_X(t)^T\, U \otimes U\, (Id - \left[\widehat{M} \otimes \widehat{M}\right]^t)(Id - \widehat{M} \otimes \widehat{M})^{-1}\, dt\, \mathrm{Vec}\Bigg(\mathrm{diag} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbb{E}(X_i)} \\ \vdots \\ 0 \end{pmatrix} \Bigg)$$

Avec :

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbb{E}(X_i)} = \frac{1}{n}\sum_{i=1}^{n} (Id_d - \sum_{j=1}^{p} \widehat{B}_j - \sum_{j=1}^{q} \widehat{C}_j)^{-1} \widehat{\beta}_0^i = (Id_d - \sum_{j=1}^{p} \widehat{B}_j - \sum_{j=1}^{q} \widehat{C}_j)^{-1} \frac{1}{n}\sum_{i=1}^{n} \widehat{\beta}_0^i$$

Reste maintenant à traiter l'estimation des paramètres d'un modèle VINGARCH(p,q).

3.3 Estimation des Paramètres

Nous allons dans cette partie, écrire le maximum de vraisemblance de nos modèles VINGARCH(p,0) en considérant n individus.
Il serait possible de faire un travail similaire dans le cas général (p,q) mais cela serait plus fastidieux (il faudrait faire une récurrence sur les dérivées partielles) et demanderait de maximiser une vraisemblance conditionnelle par rapports aux λ_0,λ_1,..,$\lambda_{max_{(p,q)}-1}$.
Pour ce faire il faudrait commencer par estimer ces λ_t pour $t < max_{(p,q)}$ par exemple en prenant leur moyenne, cela engendrerait un biais et ajouterai un problème supplémentaire quand à la convergence de cet estimateur.
On se restreint donc à un modèle plus simple VINGARCH(p,0), pour un individu $i \in [\![1,n]\!]$ et pour $t \geq p$ on peut écrire que :

$$\lambda_i(t) = \beta_0^i + \begin{pmatrix} B_1 & B_2 & \dots & B_p \end{pmatrix} \begin{pmatrix} X_i(t-1) \\ X_i(t-2) \\ \vdots \\ X_i(t-p) \end{pmatrix}$$

On a ici $d(n+dp)$ paramètres à estimer d provenant des vecteurs β_0^i et d^2 pour chaque matrice B_k. On note θ le vecteur total des paramètres dans l'ordre suivant : on commence par β_0^1 suivi de β_0^2 jusqu'à β_0^i viennent ensuite les paramètre de la matrice B_1 colonnes après colonnes (en fait Vec(B_1)) suivi de Vec(B_2) jusqu'à Vec(B_p).
Ainsi l'ensemble ouvert (Remarque 3) des paramètres du modèle s'écrit :

$$\Theta = \left\{\theta \in \mathbb{R}_+^{d(n+dp)} / \ \ \forall x \in \mathbb{C}, det\left(x^p Id_d - \sum_{i=1}^{p} x^{p-i} B_i\right) = 0 \implies \|x\| < 1\right\}$$

Si on suppose indépendance entre individus et si on note l le logarithme de la vraisemblance on obtient pour $\theta \in \Theta$ que :

$$l(\theta) = \sum_{i=1}^{n}\sum_{t=p}^{\tau_i}\sum_{j=1}^{d} log\Big(\mathbb{P}(X_i^j(t) = x_i^j(t) \mid F_{t-1}^i)\Big) = \sum_{i=1}^{n}\sum_{t=p}^{\tau_i}\sum_{j=1}^{d} x_i^j(t) log(\lambda_i^j(t)) - \lambda_i^j(t) - log(x_i^j(t)!)$$

La dépendance en θ étant claire on se permet d'écrire que $X_i^j(t,\theta) = X_i^j(t)$ et que $\lambda_i^j(t,\theta) = \lambda_i^j(t)$.

En oubliant le terme $log(x_i^j(t)!)$ qui ne dépend pas de θ, et en utilisant une écriture vectorielle on obtient que :

$$l(\theta) = \sum_{i=1}^{n}\sum_{t=p}^{\tau_i} x_i(t)^T log(\lambda_i(t)) - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^T \lambda_i(t)$$

En dérivant on obtient le gradient de l noté S qui vaut :

$$S(\theta) = \sum_{i=1}^{n}\sum_{t=p}^{\tau_i}\sum_{j=1}^{d} \frac{\partial \lambda_i^j(t)}{\partial \theta}\frac{1}{\lambda_i^j(t)}\Big(x_i^j(t) - \lambda_i^j(t)\Big) = \sum_{i=1}^{n}\sum_{t=p}^{\tau_i} \frac{\partial \lambda_i(t)}{\partial \theta}\mathrm{diag}(\lambda_i(t))^{-1}\Big(x_i(t) - \lambda_i(t)\Big)$$

Avec $\frac{\partial \lambda_i(t)}{\partial \theta}$ la matrice de taille $d(n+dp)\times d$ ayant pour colonnes les gradients des $\lambda_i^j(t)$. On peux remarquer l'égalité suivante :

$$\lambda_i(t) = \beta_0^i + Id_d \otimes \begin{pmatrix} x_i(t-1) \\ x_i(t-2) \\ \vdots \\ x_i(t-p) \end{pmatrix}^T \mathrm{Vec}\Big(B_1 \quad B_2 \quad \dots \quad B_p\Big)$$

En remarquant que $\mathrm{Vec}\Big(B_1 \quad B_2 \quad \dots \quad B_p\Big) = \begin{pmatrix} \mathrm{Vec}(B_1) \\ \mathrm{Vec}(B_2) \\ \dots \\ \mathrm{Vec}(B_p) \end{pmatrix}$, et en notant δ_{ij} le symbole de Kronecker, les dérives partielles de $\lambda_i(t)$ valent :

$$\frac{\partial \lambda_i(t)}{\partial \beta_0^{\tilde{i}}} = \delta_{i\tilde{i}} Id_d$$

$$\frac{\partial \lambda_i(t)}{\partial \mathrm{Vec}\Big(B_1 \quad B_2 \quad \dots \quad B_p\Big)} = \begin{pmatrix} x_i(t-1) \\ x_i(t-2) \\ \vdots \\ x_i(t-p) \end{pmatrix} \otimes Id_d$$

On remarque à l'aide de ces équation que les dérivées secondes de $\lambda_i(t)$ par rapport à θ valent 0, on peux donc calculer la matrice hessienne H qui vaut :

$$H(\theta) = \sum_{i=1}^{n}\sum_{t=p}^{\tau_i}\sum_{j=1}^{d} \frac{\partial \lambda_i^j(t)}{\partial \theta}\Big[\frac{-x_i^j(t)}{\lambda_i^j(t)}\Big]\frac{\partial \lambda_i^j(t)}{\partial \theta}^T = -\sum_{i=1}^{n}\sum_{t=p}^{\tau_i} \frac{\partial \lambda_i(t)}{\partial \theta}\mathrm{diag}\Big(\frac{x_i(t)}{\lambda_i(t)}\Big)\frac{\partial \lambda_i(t)}{\partial \theta}^T$$

On remarque ici que la matrice hessienne -H est bien positive par somme de matrice positive car s'écrivant sous la forme UU^T avec ici :

$$U = \frac{\partial \lambda_i(t)}{\partial \theta}\mathrm{diag}\Big(\sqrt{\frac{x_i(t)}{\lambda_i(t)}}\Big)$$

Le maximum de vraisemblance est donc une fonction globalement concave il suffit donc de trouver un point d'annulation du gradient S sur Θ pour trouver un maximum global et donc estimer nos paramètres.
On note $(\widehat{\beta_0},\widehat{B_1},...,\widehat{B_p})$ les estimateurs de $(\beta_0,B_1,...,B_p)$.

Conclusion

Nous sommes donc bien arrivé à étendre l'article [3] au cas de processus à espace d'état discret. Avec les notations utilisées dans ce rapport on obtient que le biais discret vaut :

$$\int_{t=0}^{\tau_i} \alpha_X(t)^T \otimes \alpha_X(t)^T U \otimes U \, (Id - \left[\widehat{M} \otimes \widehat{M}\right]^t)(Id - \widehat{M} \otimes \widehat{M})^{-1} \, dt \, \mathrm{Vec}\Big(\mathrm{diag}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \widehat{\mathbb{E}(X_i)} \\ \vdots \\ 0 \\ \end{pmatrix}\Big)$$

Ainsi on peux comme dans le cas continue estimer les coefficients de la régression initiale et obtenir un estimateur non biaisé de l'ATT pour tout $t \geq S$:

$$\widehat{\mathrm{ATT}}(t) = \widehat{\alpha_D}(t) + \widehat{\alpha_X}(t)\Big[\mathbb{E}_S(X^{1|S}(t) - \widetilde{X}^{0|S}(t))\Big] = \mathbb{E}_S\Big[\widehat{d^S}(t)\Big]$$

Quelques ouvertures se présentent à nous :

La première, qui est plutôt une amélioration de ce travail, serait de considérer une notion plus forte d'ergodicité que l'ergodicité pour la moyenne. On aimerai avoir des résultats du type géometriquement ergodique ou simplement ergodique au sens d'une chaîne de Markov. De plus les conditions (ii) et (iii) de la proposition 1 sont difficilement vérifiables en pratique.
Pour répondre à ce premiers point, je me suis beaucoup inspirer du cadre et des idées de preuve de la thèse de Liu Heng "Some Models for Time Series of Counts" [4] qui -si je ne me trompe pas- ne traite pas le cas exposé ici c'est à dire le modèle VINGARCH(p,q). Par manque de temps je n'ai pas pu rédiger cette partie mais l'idée est la suivante :
Premièrement il me semble que la condition (i) de la proposition 1 est en fait suffisante pour garantir la stationnarité et l'ergodicité pour la moyenne. En effet en vertu du modèle VINGARCH(p,q) on peux montrer que le vecteur :

$$(\lambda_t, \ldots, \lambda_{t-q+1}, X_{t-1}, \ldots, X_{t-p+1})$$

est une chaîne de Markov. Ainsi en m'inspirant de la proposition 4.2.1 de la thèse [4] et en adaptant quelques parties de la preuve on peux conclure que notre chaîne

de Markov admet une unique distribution stationaire.
Pour être plus précis il me semble que l'on peux en fait montrer à l'aide de la supposition (i) de la proposition 1 que notre chaîne de Markov est une e-chain, bornée en probabilité en moyenne et qui admet un état récurrent. Ainsi par le théorème 18.8.4 de Meyn et Tweedie [5] on peux montrer que notre chaine de Markov admet une unique distribution stationnaire. Ainsi si on suit le schéma de preuve de la proposition 1 on peux montrer l'ergodicité pour la moyenne.
L'ergodicité (au sens de Markov) ou l'ergodicité géométrique est plus complexe à mettre en place, sans rentrer dans trop de détails il me semble qu'il faudrait avoir une condition qui impose l'existence d'une inégalité d'une somme des norme des matrices considérées dans notre modèle pondérées par certains coefficients.

Une seconde ouverture possible et importante en pratique est celle de calculer l'ATT dans un cadre ou l'on mélange des processus continue et discret non indépendant. Avec les modèles développés ici il est possible d'allier un processus vectoriel discret et un processus vectoriel continue seulement si on les considèrent indépendant. En effet lors de l'écriture du biais, le calcul de la covariance du processus $\epsilon(t)$ s'écrit facilement par indépendance comme une matrice constitué de deux blocs de dimension d_X, d_Y ou d_X est la dimension du processus continue et d_Y est la dimension du processus discret.
Dans le cas ou il existe une dépendance entre les processus X et Y, il faudrait développer un modèle mixte discret et continue. On a envie de définir le modèle suivant :

$$\begin{pmatrix} X_t \\ \lambda_t \end{pmatrix} = \begin{pmatrix} \beta_0^X \\ \beta_0^Y \end{pmatrix} + \sum_{i=0}^{p} B_i \begin{pmatrix} X_{t-i} \\ Y_{t-i} \end{pmatrix} + \begin{pmatrix} \omega(t) \\ 0 \end{pmatrix}$$

Avec ω l'erreur du modèle, Malheureusement ce modèle simple est mal défini si X prend des valeurs négatives ou si le modèle lui fait prendre des valeurs négatives.
Il faut plus travailler pour obtenir un modèle convenable.

Enfin comme dit dans l'introduction, la mise en place d'un programme sur R permettant d'estimer les paramètre d'un modèle VINGARCH(p,0) est en cours de réalisation. Nous simulerons ensuite des processus discret $(X^{0|S}(t))_{t\geq 0}$ et $(X^{1|S}(t))_{t\geq 0}$ dont nous calculerons l'ATT. Nous nous placerons ensuite dans un contexte clinique ou l'on observe uniquement $(X^{0|S}(t))_{t<S}$ et $(X^{1|S}(t))_{t\geq S}$ et Calculerons l'ATT avec la méthode décrite dans ce stage. Enfin nous ferons un tableau comparatif des résultats obtenus.

Bibliographie

[1] Richard A. Davis, William T. M. Dunsmuir, and Sarah B. Streett. Observation-driven models for poisson counts. *Biometrika*, 90 :777–790, 2003.

[2] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued garch process. *Journal of Time Series Analysis*, 27(6) :923–942, 2006.

[3] J. M. Gran, R. Hoff, K. Røysland, B. Ledergerber, J. Young, and O. O. Aalen. Estimating the treatment effect on the treated under time-dependent confounding in an application to the Swiss HIV Cohort Study. *arXiv :1604.01597 [stat]*, October 2016. arXiv : 1604.01597.

[4] Heng Liu. *Some Models for Time Series of Counts*. PhD thesis, Columbia University, 2012.

[5] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, USA, 2nd edition, 2009.

[6] Tina Hviid Rydberg and Neil Shephard. A modelling framework for the prices and times of trades made on the new york stock exchange. *Capital Markets eJournal*, 1999.