UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÈLES DE CHAÎNES DE MARKOV CACHÉES ET DE CHAÎNES DE MARKOV COUPLES APPLIQUÉS AU NOMBRE DE DÉFAUTS EN RISQUE DE CRÉDIT

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

JEAN-FRANÇOIS FOREST-DÉSAULNIERS

JUILLET 2018

REMERCIEMENTS

D'entrée de jeu, je désire remercier les professeurs et chercheurs en actuariat, en finance et en mathématiques du Département de mathématiques de l'Université du Québec à Montréal, Jean-Philippe Boucher et Mathieu Boudreault. Leur soutien financier et psychologique tout au long de ma maîtrise fut grandement apprécié. Je tiens particulièrement à remercier Jean-Philippe Boucher pour son dévouement et sa patience à mon égard. De plus, j'aimerais souligner l'aide de Mohamed El Yazid dans la validation de mon code pour modéliser les chaînes de Markov couples (normale bivariée).

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

RÉSUMÉ

La modélisation du risque de crédit est un sujet de plus en plus important pour les institutions financières. Depuis la récente crise économique en 2008, plusieurs traités internationaux et lois nationales exigent des institutions financières une très grande rigueur dans la modélisation et le calcul du risque de crédit, et ce, afin de couvrir les risques de défaut. Le sujet de ce mémoire est la modélisation du nombre de défauts en risque de crédit à l'aide de modèles de chaînes de Markov cachées (*Hidden Markov models* - HMM) et de modèles de chaînes de Markov couples (*Pairwise Markov Chain* - PMC). Les principes importants des modèles HMM et PMC sont d'abord définis. Par la suite, ces modèles sont appliqués à une série temporelle de données empiriques du nombre de défauts de grandes compagnies américaines. Pour appliquer les modèles HMM et PMC, l'algorithme d'Espérance-Maximisation (EM - *Expectation Maximization*) est utilisé afin de calibrer les paramètres de ces modèles (Estimateur par maximum de vraisemblance - EMV). À l'aide des EMV, la modélisation et la prédiction de l'évolution du risque de défaut sont possibles. L'analyse des résultats montre que l'utilisation d'un modèle HMM capte mieux les mouvements non stationnaires de la série de défauts à l'étude que les modèles PMC. En effet, les modèles PMC - principalement utilisés en reconnaissance d'image - nécessitent un très grand nombre de paramètres, ce qui détériore significativement la qualité de ce modèle étant donné le peu de données disponibles.

Mots-clés
Chaîne de Markov caché (HMM - Hidden Markov Models), Chaîne de Markov couple (PMC - Pairwise Markov Chain), Série temporelle, Risque de crédit

INTRODUCTION

0.1 Fondement historique

Le risque de crédit se traduit par le risque qu'une contrepartie ayant contracté un prêt soit dans l'incapacité de rembourser cette dette à la compagnie émettrice. Il s'agit donc d'un danger financier très important pour les institutions financières. En effet, si une banque contracte une hypothèque avec un client et que celui-ci ne rembourse jamais cette dette, la banque subira une perte. Si tous les clients de la banque ne sont plus capables de rembourser leurs dettes, alors la banque aura de fortes chances de déclarer faillite. Il est donc primordial pour les institutions financières de modéliser le risque de crédit, puis d'instaurer des règles quant aux montants accordés à chacun de ses clients.

En ce sens, plusieurs agences règlementaires (nationales et internationales) surveillent de près le risque de crédit. Cette surveillance s'effectue pour toutes les institutions qui ont un apport financier important, en fonction de leur pays d'occupation. Ainsi, comme au Québec avec l'Autorité des Marchés Financiers (AMF), il existe des agences règlementaires pour un territoire donné. Cesdites agences veillent à ce que les institutions financières maintiennent suffisamment de capitaux afin de surmonter des périodes de crise financière et ainsi éviter la faillite. L'une des agences les plus reconnues est la *Bank for International Settlements* (BIS) qui s'assure du respect des exigences du *Basel Committee on Banking Supervision* (BCBS) (voir BIS, 2016). En effet, le BCBS, comité fondé en 1974, a mis sur pied une série d'accords internationaux ayant pour objectif d'assurer la solvabilité des grandes institutions financières. Ces accords portent le nom de d'Accords Bâle

(ou simplement Bâle), plus connus sous le nom anglais de *Basel Accords* (ou *Basel*). Or, Bâle fut amélioré plusieurs fois depuis sa version originale afin de suivre l'évolution des produits disponibles sur le marché et de traquer l'amélioration des modèles utilisés dans le calcul des risques économiques. La version actuelle (en date de 2018) de Bâle est la III (Committee *et al.*, 2010).

0.2 Risque de crédit

En plus du risque de crédit, les institutions financières font face à deux autres grands types de risques ; le risque de marché et le risque opérationnel (Dowd et Rowe, 2004). Le risque de marché concerne l'incertitude actuel de l'économie mondiale (globale). Il est difficile de modéliser ce risque, puisque l'analyse se fait sur de très courtes périodes de temps. Quant au risque opérationnel, il s'agit du risque de pertes dû à des problèmes de processus, de systèmes et de personnel, ou encore d'évènements externes autres que le défaut de paiement. Il est par définition difficile de modéliser ces risques de par leur rareté et leur caractère inattendu. Dès lors, la rareté des données constitue un tout autre obstacle à la modélisation de ce risque.

Bien qu'il soit important de modéliser tous les types de risques financiers, ce mémoire traite principalement du risque de crédit. En effet, ce risque est d'une importance capitale pour les institutions financières puisqu'un trop grand nombre de défauts peut mener à la faillite de celles-ci. La création d'agences gouvernementales qui veillent à ce que les institutions financières se prémunissent contre ce risque est d'ailleurs un autre indice du rôle central qu'il occupe dans le monde financier.

Le capital économique du risque de crédit (EC - *Credit risk economic capital*) correspond à la portion de capital qu'une institution financière doit conserver pour faire face au risque de crédit. En prenant des hypothèses sur les probabilités

de défaut de ses clients, l'institution financière peut générer une distribution de perte afin d'estimer le capital économique nécessaire pour couvrir les défauts de paiement.

Pour illustrer ce phénomène, prenons l'exemple d'un consommateur qui achète des biens dans un magasin avec une carte de crédit émise par une institution financière. Si cette personne fait défaut sur le paiement de sa carte de crédit, c'est alors à l'institution financière de payer les achats réalisés au magasin. Si beaucoup d'individus font défaut de paiement sur leur carte de crédit en même temps, l'institution financière sera à court de liquidité et ne pourra pas payer ses créanciers (dans notre exemple, il s'agit du magasin). Cependant, si l'institution financière se couvre contre le risque de crédit grâce à des réserves (soit le capital économique de risque de crédit), elle pourra payer ses créanciers et ainsi éviter une potentielle faillite.

Le capital économique est utilisé comme protection en excédent de la moyenne des pertes. L'EC peut aussi être vu comme les pertes non anticipées. Il n'y a pas de consensus autour de la définition de l'EC, mais son objectif est toujours de prévenir l'insolvabilité de l'institution financière. La définition de l'EC qui est utilisée dans ce mémoire est celle de Moody's. Selon cette définition, l'EC est la différence entre la VaR (*Value-at-Risk*, valeur au risque) et l'espérance des pertes (EL - *Expected loss*, perte anticipée)(Chorafas, 2004), soit

$$EC = VaR - EL.$$

La VaR est la valeur maximale de perte possible qu'une institution financière peut subir avec une certaine probabilité (aussi appelé seuil). Par exemple, une VaR au seuil de 99.99% serait le montant en perte tel qu'au maximum 99.99% des pertes ne dépassent pas ce seuil. La VaR ne doit pas être vue comme la simple addition de tous les risques de manière indépendante. En effet, le calcul de la VaR prend

en considération la corrélation entre chacun des risques.

La figure 0.1 illustre que le défaut de compagnies actives sur les marchés semble positivement corrélé aux évènements économiques importants, comme les crises économiques et les récessions. Or, il est donc primordial pour une institution financière de modéliser ce risque, afin d'éviter de grandes pertes.

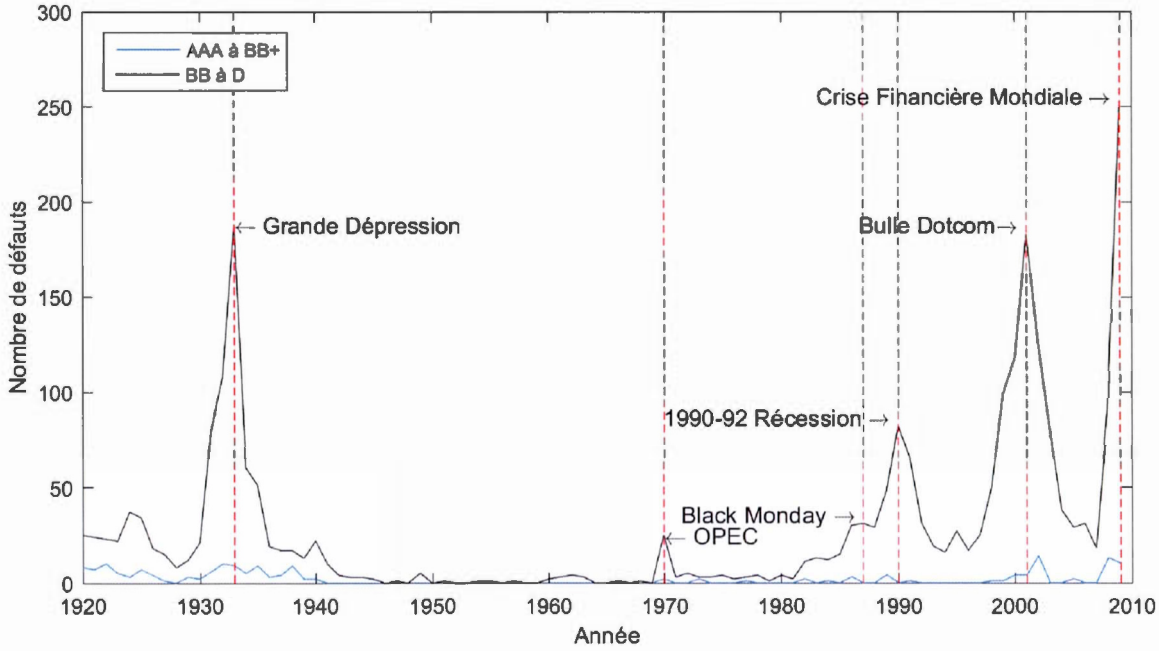
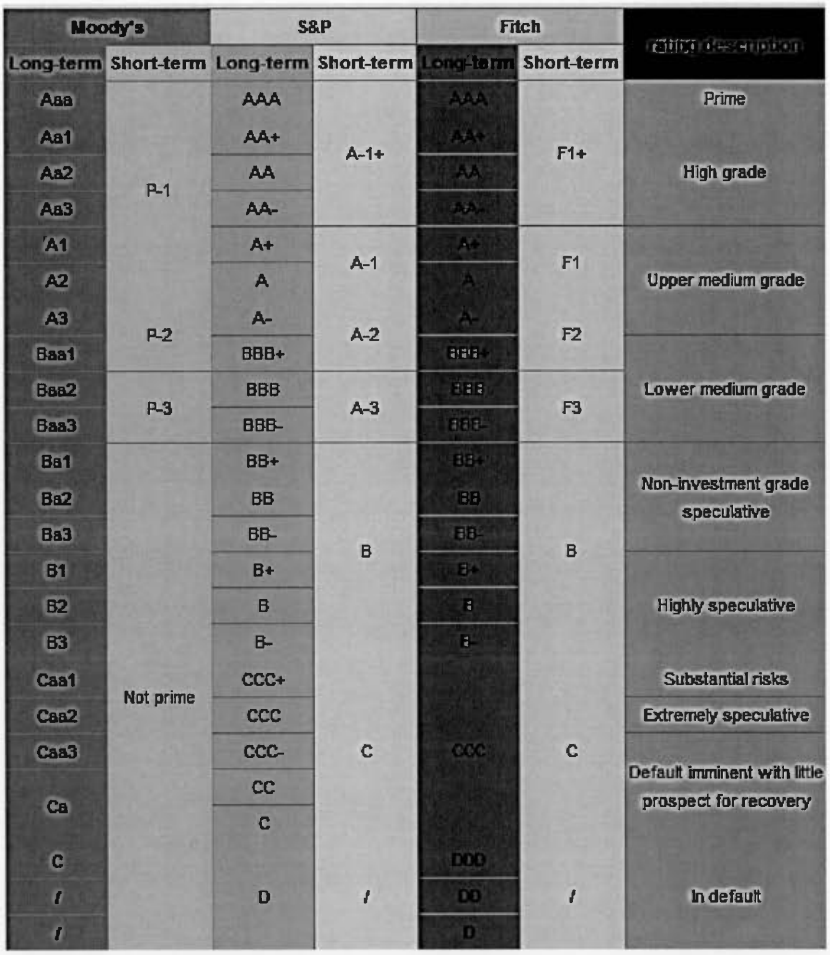


Figure 0.1 Évolution des défauts et évènements économiques importants

0.3 Cote de crédit

La figure 0.1 utilise des lettres pour désigner la cote de crédit d'une compagnie. Cette cote est évaluée par de grandes agences spécialisées dans ce domaine, telles *Standard & Poor's* et *Moody's Investors Service*. Dans la figure 0.1, la charte utilisée est celle fournie par *Standard & Poor's*. Les chartes de cotes sont très utiles, car elles permettent de déterminer le risque de défaut d'une compagnie simplement

en regardant son indice. À la figure 0.2, un comparatif des chartes de cotes des plus grandes agences américaines est illustré. De manière générale, une cote de A ou B représente un investissement ayant peu de risque de crédit. En ce qui concerne les cotes de crédit plus basses, elles représentent un investissement spéculatif, aussi appelé investissement risqué et dangereux. Les investissements à risque sont fortement affectés par les crises financières en comparaison aux investissements peu risqués (voir figure 0.1). L'utilisation des cotes de crédit semble donc une bonne méthode pour différencier les compagnies en bonne santé financière de celles ayant une plus grande probabilité de défaut. Cependant, le modèle de cote de crédit n'est pas infaillible. En guise d'exemple, la banque américaine *Lehman Brothers* était cotée A2 par *Moody's Investors Service* jusqu'à 5 jours avant qu'elle ne déclare faillite le 15 septembre 2008, lors de la crise financière globale (Moody's, 2008). Cela démontre que les modèles de risque de crédit ne sont pas encore parfaits, et c'est pourquoi il est primordial de poursuivre le développement de ceux-ci.



Moody's		S&P		Fitch		rating description
Long-term	Short-term	Long-term	Short-term	Long-term	Short-term	
Aaa	P-1	AAA	A-1+	AAA	F1+	Prime
Aa1		AA+		AA+		High grade
Aa2		AA		AA		
Aa3		AA-		AA-		
A1	P-2	A+	A-1	A+	F1	Upper medium grade
A2		A		A		
A3		A-	A-2	A-	F2	
Baa1		BBB+		BBB+		Lower medium grade
Baa2	P-3	BBB	A-3	BBB	F3	
Baa3		BBB-		BBB-		
Ba1	Not prime	BB+	B	BB+	B	Non-investment grade speculative
Ba2		BB		BB		
Ba3		BB-		BB-		
B1		B+		B+		Highly speculative
B2		B		B		
B3		B-		B-		
Caa1		CCC+				Substantial risks
Caa2		CCC				Extremely speculative
Caa3		CCC-	C	CCC	C	Default imminent with little prospect for recovery
Ca		CC				
		C				
C				DDD		In default
/		D	/	DD	/	
/				D		

Figure 0.2 Tableau comparatif des grandes agences de crédit (Wikipedia, 2014)

0.4 Modélisation du risque de crédit

La modélisation du risque de crédit est un sujet qui fait l'objet d'études académiques depuis les années 1970. L'une des premières modélisations pour ce risque fut introduite par Merton (1974). Merton compare la probabilité de faire défaut (suffisance des fonds propres) à l'évaluation d'une option d'achat (*call*) sur l'actif de la compagnie (sous-jacent) et utilise la valeur de la dette comme prix d'exercice. La méthode utilisée par Merton pour trouver la fréquence de défaut espérée (*Expected default frequency* - EDF) est similaire à celle de Black et Scholes (1973). Par la suite, le modèle de Black et Cox (1976) reprit cette idée, en ajoutant la

possibilité que le défaut puisse arriver à n'importe quel moment avant l'échéance (cette approche s'appelle typiquement "modèle de premier passage"). Plus récemment, plusieurs auteurs ont écrit sur la modélisation du risque de crédit et plusieurs compagnies ont commencé à utiliser ces modèles afin de mitiger leurs risques.

De nos jours, un modèle fortement utilisé dans l'industrie est celui développé par *KMV Corporation* puis racheté par *Moody's*. Le modèle KMV fait le calcul de l'EDF en utilisant la distance au défaut (*distance to default* - DD). La DD est un indice de défaut se calculant ainsi

$$DD = \frac{V_0 - L}{V_0 \sigma_V},$$

où V_0 est la valeur de l'actif au temps initial, L est la somme des dettes de la compagnie à rembourser dans la prochaine année et σ_V est la volatilité des rendements de l'actif.

D'autres modèles sont aussi utilisés dans l'industrie. RiskMetrics (voir Reuters, 2017) modélise le risque de crédit à l'aide de matrices de transition sur les cotes de crédit.

Creditrisk+, développé par *Credit Suisse Financial* en 1997 (voir CSFBI, 2017), modélise le risque de crédit à l'aide de modèle statistiques appliquées aux données historiques. Contrairement au modèle KMV, *Creditrisk+* n'attribue pas la structure de capital d'une compagnie à une cause de défaut.

De plus en plus de publications traitent de la modélisation du risque de crédit. Giampieri *et al.* (2005) utilisent les modèles de chaînes de Markov cachées avec une distribution binomiale pour modéliser les cycles financiers de plusieurs secteurs d'industrie (détail, énergie, média, transports). Plus récemment, les modèles de réseaux de neurones ont aussi été utilisés pour modéliser le risque de crédit (voir

Hamdy et Hussein, 2016). Le risque de crédit est donc un domaine en constante évolution et l'intérêt des institutions financières pour des modèles plus performants ne cesse d'augmenter.

0.5 Contenu du mémoire

L'importance de la modélisation du risque de crédit n'est plus à prouver et le présent mémoire s'inscrit dans la florissante recherche faite sur le sujet. Plus précisément, ce mémoire traite de la modélisation du nombre de défauts en risque de crédit. Pour se faire, les modèles de chaînes de Markov cachées ainsi que les chaînes de Markov couples sont appliqués sur une série mensuelle du nombre de défauts de grandes compagnies américaines. Le défaut de la compagnie est observé lorsque celle-ci rempli un document de Chapitre 7 ou 11 du rapport annuel *10-K* du SEC (*Securities and Exchange Commission*)(respectivement liquidation et réorganisation comme définie par le *United States Bankruptcy Code*). Cette série va de 1980 à 2017 et provient du Département de droit de l'Université de Californie, Los Angeles (UCLA). Le nom de cette base de données est *UCLA-LoPucki Bankruptcy Research Database* (BRD) (voir UCLA, 2017). Le chapitre 1 traite des concepts de base en mathématiques et plus spécifiquement en probabilités et séries temporelles. Les chapitres 2 et 3, quant à eux, développent les caractéristiques et formules essentielles pour la modélisation de séries temporelles à l'aide des chaînes de Markov cachées et des chaînes de Markov couples. Finalement, le chapitre 4 applique les modèles aux données empiriques. Une fois appliqué, il sera possible de déterminer si les modèles de chaînes de Markov cachées ainsi que les chaînes de Markov couples sont de bons mandataires pour déterminer le risque de défaut.

CHAPITRE I

CONCEPTS GÉNÉRAUX

1.1 Données de comptage

Les données de comptage sont utilisées dans plusieurs domaines tels que l'actuariat (Denuit *et al.*, 2007), la finance (Rose, 1990) et la modélisation de la démographie en économétrie (Winkelmann, 1995). Ce mémoire utilise plusieurs modèles de comptage pour modéliser le nombre de défauts, et ainsi le risque de crédit.

Les données de comptage prennent des valeurs entières dans l'intervalle $[0, \infty)$ ($\mathbb{N}_0$). L'utilisation de modèles sur de telles données permet de comprendre la distribution du nombre d'occurrences d'un évènement.

1.1.1 Distribution de Poisson

La distribution de Poisson est une distribution discrète proposée par Poisson (1837) qui permet de calculer la probabilité qu'un certain nombre d'évènements se réalisent sur une certaine période de temps. La fonction de masse de la loi de Poisson s'exprime comme suit :

$$\mathbb{P}(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y = 0, 1, ... \text{ et } \lambda > 0 \tag{1.1}$$

où λ est le seul paramètre de la distribution de Poisson. L'espérance de Y est égale à λ ($\mathbb{E}[Y] = \lambda$) et la variance de Y est égale à λ ($Var[Y] = \lambda$). Tout au long de

ce mémoire, la lettre Y est utilisée pour représenter le nombre d'occurrences d'un évènement quelconque et y est utilisée pour représenter une observation. L'indice de temps t est ajouté à y afin de définir la période de temps durant laquelle l'observation a lieu, soit y_t. La variable aléatoire (v.a.) représentant le nombre d'occurrences au temps t est Y_t.

1.1.2 Distribution de Poisson avec régresseurs

Les régresseurs sont des phénomènes observables qui ne sont pas corrélés à la variable à l'étude. Cependant, ils sont corrélés avec celle-ci. Par exemple, le nombre d'heures d'étude (régresseurs, variable indépendante) peut expliquer la note obtenue à un examen (variable dépendante). Les régresseurs sont donc possiblement prédictifs du phénomène à l'étude.

Le taux de chômage, le revenu familial, l'indice des prix à la consommation (IPC) sont tous des exemples de régresseurs se rattachant à la finance.

Les régresseurs peuvent être fixes ou évoluer dans le temps. Par exemple, l'âge d'une personne évoluera dans le temps, tandis que sa grandeur sera stable si cette personne est d'âge adulte. La valeur du régresseur i au temps t est $x_{i,t}$ et le paramètre se rattachant à cette variable est β_i.

Il est possible de rendre l'unique paramètre λ de la distribution de Poisson dépendant de plusieurs régresseurs variant dans le temps (pour plus de détails sur ce sujet, voir Cameron et Trivedi, 2013). En effet, si

$$\lambda_t = \exp(\boldsymbol{x}_t'\boldsymbol{\beta}), \tag{1.2}$$

où $\boldsymbol{\beta}$ est un vecteur des paramètres $(\beta_0, ..., \beta_p)$ et $\boldsymbol{x}_t'$ est un vecteur transposé de la valeur de chacun des p régresseurs au temps t, alors la valeur de λ_t varie dans le temps par rapport à l'évolution de la valeur des régresseurs.

D'autres distributions de comptage peuvent être utilisées et la majorité des distributions peut être modifiée afin d'admettre des régresseurs. Toutefois, le paramètre λ, de ces distributions, ne sera pas nécessairement lié de la même façon avec le score $\boldsymbol{x}_t'\boldsymbol{\beta}$.

1.1.3 Estimateur par maximum de vraisemblance (EMV)

L'estimateur par maximum de vraisemblance (EMV) fut introduit par Fisher (1912, 1922). Cette technique permet de retrouver les paramètres d'une distribution qui maximisent la vraisemblance d'obtenir un certain échantillon d'observations (y_t).

L'EMV maximise donc la fonction de vraisemblance. La fonction de vraisemblance est la fonction de masse de probabilité de Y évaluée en y, conditionnellement à la valeur du vecteur $\boldsymbol{\theta}$, où $\boldsymbol{\theta}$ représentant un vecteur de paramètres. Pour une distribution de comptage, la fonction de vraisemblance s'écrit :

$$\begin{aligned} L(\boldsymbol{\theta}; y_1, ..., y_T) &= L_T \\ &= \mathbb{P}(Y_1 = y_1, Y_2 = y_2, ..., Y_T = y_T|\boldsymbol{\theta}) \qquad (1.3) \\ &= \prod_{t=1}^{T} \mathbb{P}(Y_t = y_t|\boldsymbol{\theta}). \qquad (1.4) \end{aligned}$$

Le passage de l'équation (1.3) à (1.4) se fait en supposant que $y_1, \ldots, y_T$ sont des réalisations de variables aléatoires indépendantes et identiquement distribuées (iid) $Y_1, \ldots, Y_T$, où T la taille de l'échantillon. Comme $0 \leq \mathbb{P}(Y_t = y_t) \leq 1$, la valeur de L_T tendra vers 0 plus T sera grand. Généralement, puisque la dérivée d'un produit est plus difficile à effectuer que la dérivée d'une somme, il est préférable d'utiliser le logarithme de la vraisemblance (qui sera nommé log-vraisemblance), soit :

$$l(\boldsymbol{\theta}; y_1, ..., y_T) = l_T = \log L_T.$$

Par la suite, le vecteur de paramètres $\boldsymbol{\theta}$ est estimé en maximisant la fonction de log-vraisemblance par rapport au paramètre $\boldsymbol{\theta}$. Cet estimateur est dénoté $\hat{\boldsymbol{\theta}}$.

Pour trouver la valeur des $\hat{\boldsymbol{\theta}}$, il est possible d'utiliser une large variété d'algorithmes numériques, dont la méthode de Newton-Raphson (pour des détails au sujet de cet algorithme, voir Ypma, 1995).

Soulignons que la maximisation par l'algorithme de Newton-Raphson n'est pas toujours nécessaire. En effet, si la fonction ne possède qu'un seul paramètre, la résolution de la première dérivée de l_T par rapport à ce paramètre donne la valeur de $\hat{\boldsymbol{\theta}}$. La résolution analytique à plusieurs dimensions est faisable, mais cette tâche est complexe.

La maximisation de la fonction de vraisemblance est généralement utilisée pour trouver la valeur des paramètres. L'EMV est choisi pour ses caractéristiques intéressantes d'estimation. Quatre des caractéristiques principales de l'EMV sont (Woodcock, 2014) :

1. **Convergence :** L'EMV converge en probabilité vers la vraie valeur du paramètre estimé, soit :

$$\operatorname*{plim}_{n\to\infty} \hat{\theta}_{i,n} = \theta_i,$$

où $\hat{\theta}_{i,n}$ est le $i$ème paramètre de $\hat{\boldsymbol{\theta}}$ et l'indice n représente le nombre d'observations pour estimer le paramètre θ_i.

2. **Asymptotiquement normale :** plus le nombre d'observations augmente, plus la distribution du paramètre estimé tend vers une loi normale de moyenne $\hat{\theta}_i$ et de variance égale à l'inverse de la matrice d'information de Fisher du paramètre θ_i ($\mathcal{I}_{\theta_i}^{-1}$), soit :

$$\lim_{n\to\infty} \hat{\theta}_{i,n} \sim N(\theta_i, \mathcal{I}_{\theta_i}^{-1}),$$

où $N(a, b)$ est une loi normale de moyenne a et de variance b. De plus,

$$\mathcal{I}_{\theta_i} = -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}, Y_1, ..., Y_T)}{\partial(\theta_i)^2}\right].$$

3. **Invariance :** Si $\hat{\theta}_{i,n}$ est l'EMV de θ_i, alors peu importe la fonction $v = g(.)$, l'EMV de v est :

$$\hat{v} = g(\hat{\theta}_{i,n}).$$

4. **Asymptotiquement efficiente :** La variance de $\hat{\theta}_i$ atteint la borne inférieure de Cramér–Rao avec un nombre d'observations tendant vers l'infini, soit :

$$\lim_{n\to\infty} Var(\hat{\theta}_{i,n}) \geq \mathcal{I}_{\theta_i}^{-1}.$$

Puisque $\lim_{n\to\infty} Var(\hat{\theta}_{i,n}) = \mathcal{I}_{\theta_i}^{-1}$, alors l'EMV est un estimateur pleinement efficient.

1.1.4 EMV - Poisson et Poisson avec régresseurs

La fonction de la log-vraisemblance d'une distribution de Poisson s'écrit

$$\begin{aligned} L_T &= \prod_{i=1}^{T} \mathbb{P}(Y_i = y_i) \\ l_T &= \sum_{i=1}^{T} \log\left(\mathbb{P}(Y_i = y_i)\right) \\ &= \sum_{i=1}^{T} \log\left(\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}\right) \\ &= \sum_{i=1}^{T} \left[-\lambda + y_i \log(\lambda) - \log(y_i!)\right], \end{aligned}$$

où $y_1, ..., y_T \in \mathbb{N}_0$ sont des observations. Pour trouver la valeur de l'EMV de λ, il ne reste plus qu'à trouver la valeur de λ qui maximise l_t. La valeur du paramètre peut être déterminée en dérivant l_t par rapport au paramètre recherché. Il s'ensuit

que :

$$\frac{\partial l_T}{\partial \lambda} = \frac{\partial}{\partial \lambda} \sum_{i=1}^{T} [-\lambda + y_i \log(\lambda) - \log(y_i!)]$$

$$\Leftrightarrow 0 = \sum_{i=1}^{T} \left[-1 + \frac{y_i}{\lambda}\right] \tag{1.5}$$

$$\Leftrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{T} y_i}{T}. \tag{1.6}$$

À l'équation (1.5), la valeur de zéro provient du fait que le maximum du logarithme de la fonction de vraisemblance est recherché.

L'EMV des paramètres d'une Poisson avec régresseurs peut aussi être trouvé. L'équation 1.7 représente la fonction de log-vraisemblance d'une loi de Poisson où $\lambda_t = \exp(\boldsymbol{x}_t'\boldsymbol{\beta})$. L'équation 1.8 représente, qu'en a-t-elle, la première dérivée par rapport à $\boldsymbol{\beta}$ de l'équation 1.7 (pour des détails au sujet des équations 1.7 et 1.8, voir Cameron et Trivedi, 2013).

$$L_T = \prod_{t=1}^{T} \mathbb{P}(Y_t = y_t)$$

$$\log(L_T) = \sum_{t=1}^{T} -\lambda_t + y_t \log(\lambda_t) - \log(y_t!)$$

$$l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} -\exp(\boldsymbol{x}_t'\boldsymbol{\beta}) + y_t \boldsymbol{x}_t'\boldsymbol{\beta} - \log(y_t!); \tag{1.7}$$

$$\frac{\partial l_T}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{t=1}^{T} -\exp(\boldsymbol{x}_t'\boldsymbol{\beta}) + y_t \boldsymbol{x}_t'\boldsymbol{\beta} - \log(y_t!)$$

$$= \sum_{t=1}^{T} \boldsymbol{x}_t'(y_t - \hat{\lambda}_t); \tag{1.8}$$

où $\hat{\lambda}_t = \exp(\boldsymbol{x}_t'\hat{\boldsymbol{\beta}})$. Pour trouver les paramètres de la loi de Poisson avec régresseurs, la technique élaborée dans la section 1.1.3 peut être utilisée, ou bien un algorithme de maximisation sur l_T, tel `fmincon` de *Matlab*.

Les algorithmes de maximisation (tel Newton-Raphson) sont itératifs. Ces algorithmes requièrent donc des valeurs initiales pour les paramètres à maximiser, soit $\hat{\boldsymbol{\theta}}^{(0)}$. La notation suivante est introduite, $\hat{\boldsymbol{\theta}}^{(n)}$, soit la valeur des $\hat{\boldsymbol{\theta}}$ à la $(n+1)$-ième itération d'un algorithme de maximisation itératif. Donc les $\hat{\boldsymbol{\theta}}^{(n)}$ ont toujours un retard de 1 par rapport à l'itération de l'algorithme de maximisation. Par exemple, les valeurs de $\hat{\boldsymbol{\theta}}^{(0)}$ sont utilisées dans la première itération, soit $n+1=0+1=1$. Il s'en suit que $\hat{\theta}_i^{(n)}$, représente l'estimation de θ_i à la $(n+1)$-ième itération. Les valeurs de $\hat{\boldsymbol{\theta}}^{(0)}$ (paramètres initiaux) sont définies par l'utilisateur. L'utilisateur doit donc choisir celles-ci avec soin.

Pour la Poisson, il est préférable de mettre tous les $\beta_i^{(0)} (i \in [0,p])$ initiaux de l'algorithme à zéro à l'exception de $\beta_0^{(0)} = \log\left(\frac{\sum_{i=1}^n y_i}{n}\right)$, soit

$$\hat{\boldsymbol{\theta}}^{(0)} = \left[\log\left(\frac{\sum_{i=1}^n y_i}{n}\right), 0, \ldots, 0\right].$$

1.1.5 Distribution conditionnelle

Empiriquement, il est connu que les séries de comptages admettent plus souvent une surdispersion (Cameron et Trivedi, 2013), c'est-à-dire que la variance empirique du phénomène à l'étude est supérieure à son espérance.

Une technique simple pour générer de la surdispersion consiste en l'ajout d'une variable aléatoire F modifiant le paramètre λ. La distribution de Y est conditionnellement indépendant à la v.a. F. Ce modèle s'appelle le modèle avec hétérogénéité. Par exemple, une distribution Y est supposée équidispersée, soit :

$$\mathbb{E}[Y] = Var[Y] = \lambda.$$

L'ajout d'une variable aléatoire F à cette distribution, a donc l'effet suivant sur

l'espérance et la variance de cette distribution :

$$\begin{aligned}\mathbb{E}[Y] =& \mathbb{E}[\mathbb{E}[Y|F]] \\ =& \mathbb{E}[\lambda F] \\ =& \lambda \mathbb{E}[F];\end{aligned}$$

$$\begin{aligned}Var[Y] =& Var[\mathbb{E}[Y|F]] + \mathbb{E}[Var[Y|F]] \\ =& Var[\lambda F] + \mathbb{E}[\lambda F] \\ =& \lambda^2 Var[F] + \lambda \mathbb{E}[F] \ .\end{aligned}$$

Si la variance de $F \neq 0$, la variance de Y est donc plus grande que l'espérance de Y soit :

$$\begin{aligned}&\lambda^2 Var[F] + \lambda \mathbb{E}[F] > \lambda \mathbb{E}[F] \\ &\Leftrightarrow Var[Y] > \mathbb{E}[Y].\end{aligned}$$

1.2 Série temporelle de comptage

Une série temporelle est une suite d'observations d'un phénomène dans le temps. Les observations seront faites à des intervalles de temps discrets, de mêmes durées et peuvent leurs valeur peuvent dépendre en fonction de la période d'occurrence. En effet, la valeur de l'observation au temps t (y_t) peut varier par rapport à la période d'occurrence (t) et de la valeur des observations passées (soit $y_1, ..., y_{t-1}$). L'objectif de la modélisation des séries temporelles est de prédire les valeurs futures de la série temporelle. L'évolution de la valeur de l'indice financier *S&P 500* est un exemple classique de série temporelle.

Jung et Tremayne (2011) ont défini plusieurs modèles de séries de comptage, soit les modèles à régression statique, les modèles autorégressifs à moyenne conditionnelle, les modèles autorégressifs naturels (INAR) et les modèles linéaires généralisés autorégressifs à moyenne mobile (GLARMA).

1.2.1 Propriété de dépendance de Markov

Deux propriétés fort utiles que l'on peut admettre pour une série temporelle sont celles des dépendances de Markov (voir Paroli et Spezia, 2000). Ces deux propriétés sont :

— **Indépendance conditionnelle :** Il a été mentionné à la section 1.2 qu'une série temporelle peut utiliser les valeurs obtenues antérieurement (en $s = 0, ..., t-1$) pour prédire la valeur de la période actuelle (en t). La propriété d'indépendance conditionnelle prévoit que toute l'information sur le passé est inutile et que simplement la valeur actuelle est importante pour déterminer la valeur future (voir Durrett, 2010), soit :

$$\mathbb{P}(Y_{t+1} = y_{t+1}|Y_t, Y_{t-1}, ..., Y_1) = \mathbb{P}(Y_{t+1} = y_{t+1}|Y_t).$$

— **Dépendance contemporaine** [1] **:** La distribution de Y_t conditionnellement aux variables $(X_1, ..., X_t)$, ne dépend que de la valeur présente (contemporaine) de la variable X_t, soit :

$$\mathbb{P}(Y_t = y_t|X_t, X_{t-1}, ..., X_1) = \mathbb{P}(Y_t = y_t|X_t).$$

La v.a. X représente un quelconque phénomène dont la v.a. Y dépend.

1.2.2 Fonction de vraisemblance des séries temporelles

L'équation (1.3) est la forme générale de la fonction de vraisemblance. En présence de séries temporelles, cette équation tient toujours. Cependant il est, en général, plus simple d'utiliser l'équation qui suit pour effectuer le calcul du maximum de vraisemblance, soit :

---

1. Ceci est la traduction libre de *contemporary dependence*.

$$
\begin{aligned}
L_T &= \mathbb{P}(Y_1 = y_1, Y_2 = y_2, ..., Y_T = y_T) \\
&= \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}) \qquad (1.9) \\
&= \mathbb{P}(Y_1 = y_1) \times \mathbb{P}(Y_2 = y_2 | Y_1 = y_1) \times \mathbb{P}(Y_3 = y_3 | Y_1 = y_1, Y_2 = y_2) \qquad (1.10) \\
&\quad \times ... \times \mathbb{P}(Y_T = y_T | \boldsymbol{Y}^{(T-1)} = \boldsymbol{y}^{(T-1)}) \,,
\end{aligned}
$$

où $\boldsymbol{Y}^{(T)}$ représente toute l'information du temps 1 à T, soit $\boldsymbol{Y}^{(T)} = Y_1, Y_2, ..., Y_T$.

Il est trivial de voir que si Y_t ne dépend d'aucune valeur Y_{t_1} (où $t_1 < T$), alors l'équation (1.9) est égale à l'équation (1.4).

1.3 Multiplicateur de Lagrange

La méthode des multiplicateurs de Lagrange est une technique introduite par Lagrange (1788), utilisée pour retrouver le maximum de fonctions qui comportent des contraintes d'égalité. Par exemple, une fonction de masse a toujours la contrainte que $\int_\Omega f_x(x)dx = 1$.

La méthode des multiplicateurs de Lagrange modifie la fonction à maximiser avec des paramètres λ_i qui représentent chacune des contraintes. Cette nouvelle fonction à maximiser est appelée Lagrangien. L'exemple suivant de Lagrangien s'applique sur la fonction $f(x_1, x_2, ..., x_p)$:

$$
\begin{aligned}
&W(x_1, x_2, ..., x_p; \lambda_1, \lambda_2, ..., \lambda_p) \\
&= f(x_1, x_2, ..., x_p) + \lambda_1(c_1 - g_1(x_1, x_2, ..., x_p)) \\
&+ \lambda_2(c_2 - g_2(x_1, x_2, ..., x_p)) + ... + \lambda_p(c_p - g_p(x_1, x_2, ..., x_p)),
\end{aligned}
$$

où les c_i et g_i représentent des conditions à respecter, soit

$$
c_i = g_i(x_1, x_2, ..., x_p), \text{ pour } i = 1, 2, ..., p \,.
$$

1.4 Critère AIC et BIC

Le critère d'information d'Akaike (AIC - Akaike, 1974) et le critère d'information bayésien (BIC - Schwarz *et al.*, 1978) sont deux mesures utilisées pour la sélection du modèle. L'utilisation des EMV dans la fonction de la log-vraisemblance maximisera cette fonction. La valeur maximisée donne une mesure de comparaison pour des modèles ayant le même nombre de paramètres. Lorsque les modèles n'ont pas le même nombre de paramètres, la comparaison de la valeur maximisée des log-vraisemblances à l'aide des EMV est non représentative. C'est ce qui explique l'utilité des mesures AIC et BIC. En effet, ces deux mesures prennent en considération le nombre de paramètres, ainsi que le nombre d'observations (BIC seulement). Le calcul de ces critères d'information utilise les équations suivantes :

$$AIC = 2k - 2l_T$$
$$BIC = k\log(n) - 2l_T,$$

où k est le nombre de paramètres à estimer, n le nombre d'observations, et l_T est la valeur maximisée de la log-vraisemblance à l'aide des EMV telle que définie plus tôt.

1.5 Propriété des distributions de mélanges

La fonction de densité d'une distribution de mélange est créée en additionnant le produit des proportions ω_i aux fonctions de probabilités $f_i(y)$. Ces distributions ont les propriétés suivantes, et ce, peu importe les distributions qui sont mélangées entre elles (pour plus d'informations sur les distributions de mélange, voir

Frühwirth-Schnatter, 2006) :

$$f(y) = \sum_i \omega_i f_i(y)$$

$$F(y) = \sum_i \omega_i F_i(y)$$

$$\mathbb{E}[Y^k] = \sum_i \omega_i \mathbb{E}_i[Y^k]$$

$$Var[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2,$$

où $f(y)$ est une fonction de densité associée à la distribution de mélanges, $F(y)$ est une fonction de répartition associée à la distribution de mélanges et $\mathbb{E}[Y^k]$ est le moment k.

CHAPITRE II

CHAÎNES DE MARKOV CACHÉES

2.1 Notions de base

Les modèles de chaînes de Markov cachées (aussi appelés HMM pour *Hidden Markov Model*) ont été introduits par Baum, Eagon, Petrie, Soules et Weiss (1966, 1970, 1972). Ces modèles sont constitués d'une série temporelle observée (ayant comme notation $\boldsymbol{y}$) ainsi que d'une série temporelle inobservée (ayant comme notation $\boldsymbol{c}$). La série temporelle observée est le phénomène à l'étude. Dans ce mémoire, cette série sera le nombre de défauts (voir l'application au chapitre 4). Quant à la série temporelle cachée, il s'agit d'une série fictive qui est générée par l'algorithme qui construit le modèle HMM. Cette série représente un changement de régime (état).

Tout au long de ce mémoire, la v.a. C est utilisée pour représenter un l'état inobservé quelconque et c est utilisée pour représenter une réalisation de C. L'indice de temps t est ajouté à c afin de définir la période de temps durant laquelle la réalisation a lieu, soit c_t. La v.a. de l'état inobservé au temps t est dénoté par C_t. La série temporelle inobservée prend une valeur i (pour $i = 1, ..., m$) à chaque période de temps t. Cette valeur représente l'état caché au temps t. Par exemple, l'évolution de la valeur d'un titre financier peut être utilisée pour déterminer qu'il existe deux états cachés ($m = 2$). Une analyse pourrait démontrer que lorsque les

marchés sont plutôt stables, il s'agit de l'état $i = 1$ et lorsque ceux-ci sont très volatiles, il s'agit de l'état $i = 2$. Le fait de diviser un phénomène en m états permet de dire qu'il y a m distributions conditionnelles pour la série temporelle observée. Toujours avec l'exemple des marchés financiers, l'état $i = 1$ pourrait avoir un rendement ainsi qu'une volatilité plus faible, alors que l'état $i = 2$ pourrait avoir un plus grand rendement, mais aussi une plus grande volatilité.

Afin de mieux comprendre ce qu'est un HMM, un schéma de transition très simple montrant le lien entre les périodes de temps et les réalisations observées et inobservées (cachées) est illustré à la figure 2.1.

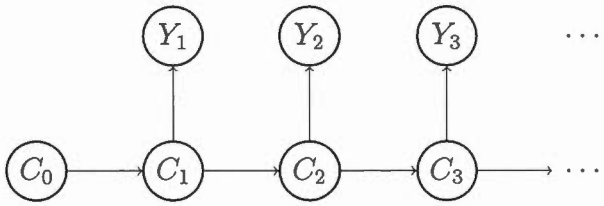


Figure 2.1 Schéma de transition modèle HMM simple

Le modèle HMM illustré à la figure 2.1 a comme hypothèse l'indépendance conditionnelle ainsi que la dépendance contemporaine des chaînes de Markov, tel qu'énoncé à la section 1.2.1. De plus, Y_t et $\mathbf{Y}^{(t-1)}$ sont indépendants sachant C_t (indépendance conditionnelle des Y_t sachant C_t), soit :

$$\mathbb{P}(Y_t|C_t, \mathbf{Y}^{(t-1)}) = \mathbb{P}(Y_t|C_t), \tag{2.1}$$

Dans ce mémoire, le terme ***HMM simple*** désignera le modèle de HMM illustré à la figure 2.1.

La série cachée change ou non d'état à chaque période de temps. C'est pourquoi les modèles HMM utilisent une matrice de transition $\mathbf{\Gamma}$, également inobservée, pour représenter toutes les probabilités de transitions. En utilisant les propriétés

d'indépendance conditionnelle (voir la section 1.2.1), il s'ensuit que :

$$\mathbb{P}(C_t = c_j | C_1 = c_1, ..., C_{t-1} = c_i) = \mathbb{P}(C_t = c_j | C_{t-1} = c_i) = \gamma_{ij}. \tag{2.2}$$

Le paramètre γ_{ij} est la probabilité de passer de l'état i au temps $t-1$ à l'état j au temps t, sachant l'état i au temps $t-1$. La matrice $\mathbf{\Gamma}$ est définie comme :

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{bmatrix}. \tag{2.3}$$

Tel qu'énoncé plus tôt, le paramètre m représente le nombre d'états cachés possibles du modèle en question (m est donc préalablement défini). Puisque cette matrice est cachée, le vecteur de réalisations observées (soit $\boldsymbol{y}_t$) doit être utilisé afin d'estimer les valeurs de $\mathbf{\Gamma}$. Sachant que chaque ligne de la matrice $\mathbf{\Gamma}$ représente des probabilités conditionnelles à l'état de départ, il s'ensuit que la somme des probabilités de chacune des lignes est égale à 1. Il est donc possible de redéfinir cette matrice en omettant les γ_{im} (pour $i = 1, ..., m$), soit :

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & 1 - \sum_{j=1}^{m-1} \gamma_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & 1 - \sum_{j=1}^{m-1} \gamma_{mj} \end{bmatrix}. \tag{2.4}$$

Pour simplifier la notation, le paramètre $\boldsymbol{\eta}(t)$ est introduit :

$$\boldsymbol{\eta}(t) = \begin{bmatrix} \mathbb{P}(C_t = 1) & ... & \mathbb{P}(C_t = m) \end{bmatrix}, \qquad \text{pour } t \in \mathbb{N}. \tag{2.5}$$

L'équation (2.5) représente le vecteur des probabilités d'être à l'état i au temps t (pour $i = 1, ..., m$). Par conséquent :

$$\boldsymbol{\eta}(0) = \boldsymbol{\delta}$$

$$\eta_i(t) = \mathbb{P}(C_t = i)$$

$$\boldsymbol{\eta}(t+1) = \boldsymbol{\eta}(t)\mathbf{\Gamma}$$

où $\boldsymbol{\delta}$ est le vecteur des probabilités d'être à l'état i (pour $i = 1, ..., m$) au temps $t = 0$. Ce vecteur peut être choisi au départ, ou peut aussi se calculer en supposant que ses valeurs sont égales au vecteur stationnaire de $\boldsymbol{\Gamma}$, soit (voir Zucchini et MacDonald, 2009, p.68) :

$$\boldsymbol{\delta} = \boldsymbol{J}_{1,m}(\boldsymbol{I}_m - \boldsymbol{\Gamma} + \boldsymbol{J}_{m,m})^{-1}, \tag{2.6}$$

où $\boldsymbol{I}_m$ est une matrice identité de $m \times m$ et $\boldsymbol{J}_{a,b}$ est une matrice de dimension $a \times b$ dont tous les éléments sont égaux à 1. En d'autres mots, dans cette situation, $\boldsymbol{\delta}$ est le vecteur de probabilités de se retrouver à l'état i lorsque $t \to \infty$. Le vecteur $\boldsymbol{\delta}$ sera aussi nommé le vecteur stationnaire de la matrice $\boldsymbol{\Gamma}$

2.2 Fonction de vraisemblance et *Complete-data log-likelihood* pour le modèle HMM simple

Les modèles de chaînes de Markov cachées sont des séries temporelles où la distribution de Y_t dépend de $\boldsymbol{Y}^{(t-1)}$. Le calcul de la fonction de vraisemblance se fait donc à partir de l'équation (1.9), soit $L_T = \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})$.

Toutefois, la distribution de Y_t dépend aussi des $\boldsymbol{C}^{(T)}$, où $\boldsymbol{C}^{(T)}$ représente toute l'information sur les états inobservés du temps 0 à T. Il s'ensuit que (voir Zucchini et MacDonald, 2009, p.37) :

$$\begin{aligned} L_T &= \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}) \\ &= \sum_{c_0, c_1, c_2, \ldots, c_T = 1}^{m} \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, \boldsymbol{C}^{(T)} = \boldsymbol{c}^{(T)}), \end{aligned} \tag{2.7}$$

De l'équation ci-dessus et en admettant que les états inobservés sont connus, la fonction de vraisemblance des données complète (*Complete-data likelihood* - CDL)

est définie comme suit :

$$
\begin{aligned}
&\mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)}) \\
&= \mathbb{P}(Y_T, \boldsymbol{Y}^{(T-1)}, \boldsymbol{C}^{(T)}) \\
&= \mathbb{P}(Y_T|\boldsymbol{Y}^{(T-1)}, \boldsymbol{C}^{(T)})\mathbb{P}(\boldsymbol{Y}^{(T-1)}, \boldsymbol{C}^{(T)}) \\
&\dots \\
&= \mathbb{P}(Y_T|\boldsymbol{Y}^{(T-1)}, \boldsymbol{C}^{(T)})\mathbb{P}(Y_{T-1}|\boldsymbol{Y}^{(T-2)}, \boldsymbol{C}^{(T)}) \dots \mathbb{P}(Y_1|\boldsymbol{C}^{(T)})\mathbb{P}(\boldsymbol{C}^{(T)}) \\
&= \prod_{t=1}^{T} \mathbb{P}(Y_t|\boldsymbol{Y}^{(t-1)}, \boldsymbol{C}^{(T)})\mathbb{P}(\boldsymbol{C}^{(T)}) \\
&= \prod_{t=1}^{T} \mathbb{P}(Y_t|\boldsymbol{Y}^{(t-1)}, \boldsymbol{C}^{(T)})\mathbb{P}(C_T|\boldsymbol{C}^{(T-1)})\mathbb{P}(\boldsymbol{C}^{(T-1)}) \\
&\dots \\
&= \mathbb{P}(C_0) \prod_{t=1}^{T} \mathbb{P}(Y_t|\boldsymbol{Y}^{(t-1)}, \boldsymbol{C}^{(T)})\mathbb{P}(C_t|\boldsymbol{C}^{(t-1)}).
\end{aligned} \tag{2.8}
$$

Avec les deux propriétés de dépendance de Markov (voir 1.2.1), ainsi que l'hypothèse d'indépendance conditionnelle des Y_t sachant C_t (voir l'équation (2.1)), il est possible de dire que :

$$
\mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)}) = \mathbb{P}(C_0) \prod_{t=1}^{T} \mathbb{P}(C_t|C_{t-1})\mathbb{P}(Y_t|C_t). \tag{2.9}
$$

La fonction de vraisemblance de l'équation (2.7) peut donc être réécrite à l'aide de l'équation (2.9), soit :

$$
L_T = \sum_{c_0, c_1, c_2, \dots, c_T = 1}^{m} \mathbb{P}(C_0) \prod_{t=1}^{T} \mathbb{P}(C_t|C_{t-1})\mathbb{P}(Y_t|C_T). \tag{2.10}
$$

Cette expression peut être réécrite avec une notation matricielle :

$$
L_T = \boldsymbol{\delta \Gamma P}(y_1) \dots \boldsymbol{\Gamma P}(y_T)\mathbf{1}', \tag{2.11}
$$

où

$$\boldsymbol{P}(y_t) = \begin{bmatrix} p_1(y_t) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & p_m(y_t) \end{bmatrix},$$

et $p_i(y_t)$ représente la probabilité observée y_t au temps t sachant l'état i au temps t, soit :

$$p_i(y_t) = \mathbb{P}(Y_t = y_t | C_t = i).$$

Le logarithme de la fonction de vraisemblance des données complètes (*complete-data log-likelihood* - CDLL) est un concept primordial pour maximiser la fonction de la log-vraisemblance dans un modèle HMM. L'équation (2.9) peut se réécrire ainsi :

$$\log \mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)}) = \log \delta_{C_0} + \sum_{t=1}^{T} \log \gamma_{C_{t-1}, C_t} + \sum_{t=1}^{T} \log p_{C_t}(y_t). \tag{2.12}$$

où $\delta_{C_0} = \mathbb{P}(C_0 = c_0)$. De nouveaux paramètres sont introduits au modèle afin d'exprimer la probabilité de se retrouver dans un état à une période de temps donnée :

— $\zeta_t(i) = \begin{cases} 1, & \text{si } C_t = i \\ 0, & \text{sinon.} \end{cases}$

— $\psi_t(i,j) = \begin{cases} 1, & \text{si } C_{t-1} = i \text{ et } C_t = j \\ 0, & \text{sinon.} \end{cases}$

Finalement, l'équation (2.12) peut ainsi être réécrite sous la forme suivante :

$$\begin{aligned} \log \mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)}) = &\sum_{i=1}^{m} \zeta_t(i) \log \delta_i + \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\sum_{t=1}^{T} \psi_t(i,j) \right) \log \gamma_{ij} \\ &+ \sum_{i=1}^{m} \sum_{t=1}^{T} \zeta_t(i) \log p_i(y_t). \end{aligned} \tag{2.13}$$

2.3 Paramétrisation du modèle HHM simple

Pour résumer, pour un modèle HMM simple, les paramètres sont les suivants :

1. Un total de $m \times (m-1)$ valeurs de γ_{ij} $(i, j = 1, 2, ..., m)$(définies à l'équation (2.2)). La soustraction de m paramètres, provient du fait que la matrice $\boldsymbol{\Gamma}$ est de dimension $m \times m$ et la somme de chaque ligne devra être égale à 1. Il en résulte que $\gamma_{im} = 1 - \sum_{j=1}^{m-1} \gamma_{ij}$ (voir la matrice (2.4)).

2. Aucun ou $m-1$ paramètre(s) pour $\boldsymbol{\delta}$. La soustraction de 1 paramètre (soit -1 de $m-1$), provient du fait que la somme du vecteur $\boldsymbol{\delta}$ est de 1, donc il est possible d'exprimer δ_m ainsi,

$$\delta_m = 1 - \sum_{i=1}^{m-1} \delta_i.$$

   Dans ce mémoire, le vecteur $\boldsymbol{\delta}$ utilisé sera formé à partir de la matrice $\boldsymbol{\Gamma}$, comme indiqué à l'équation (2.6).

3. Un nombre indéterminé de paramètres pour les distributions de $\boldsymbol{Y}|\boldsymbol{C}$. En effet, ce nombre de paramètres varie en fonction des distributions de $Y_t|C_t = i$ $(i = 1, 2, ..., m)$. Par exemple si les distributions de $Y_t|C_t = i$ sont des lois de Poisson, un total de m paramètres serait à évaluer. Le m vient du fait qu'il y aura toujours m distributions de $\boldsymbol{Y}|C_i (i = 1, 2, ..., m)$, car celles-ci représentent les m états possibles. Dans cet exemple, il y a donc m paramètres λ, soit λ_i où i $(i = 1, 2, ..., m)$ représente l'état inobservé.

À noter que le vecteur de tous les paramètres des modèles HMM est noté $\boldsymbol{\theta}$.

Il est possible de réécrire le maximum de la fonction de vraisemblance avec une autre paramétrisation. Les notions de probabilités *Forward* et *Backward*[1], ainsi que plusieurs autres probabilités, indispensables sous cette paramétrisation, sont introduites pour les modèles HMM. Les notations et équations (2.14)-(2.23) sont définies :

---

1. Certaines publications françaises utilisent les termes passe-avant et passe-arrière, cependant les termes *Forward* et *Backward* sont usuellement utilisés pour désigner ces probabilités.

— La probabilité d'obtenir $Y_1 = y_1, Y_2 = y_2, ..., Y_t = y_t$ et $C_t = i$, soit une probabilité *Forward*, est notée :

$$\alpha_t(i) = \mathbb{P}(\boldsymbol{Y}^{(t)} = \boldsymbol{y}^{(t)}, C_t = i). \tag{2.14}$$

Le vecteur de toutes les valeurs $\alpha_t(i)$ pour $i = 1, ..., m$, soit $\boldsymbol{\alpha}_t$, peut s'obtenir à partir de la formule récursive suivante :

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}). \tag{2.15}$$

Par hypothèse, $\boldsymbol{\alpha}_0 = \boldsymbol{\delta}$ et donc le vecteur $\boldsymbol{\alpha}_t$ peut aussi s'écrire ainsi :

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta} \boldsymbol{\Gamma} \boldsymbol{P}(y_1) \boldsymbol{\Gamma} \boldsymbol{P}(y_2) ... \boldsymbol{\Gamma} \boldsymbol{P}(y_t). \tag{2.16}$$

La démonstration de ces trois équations se trouve à la proposition 4 de l'annexe A.

— La probabilité d'obtenir $Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, ..., Y_T = y_T$ sachant que $C_t = i$, soit une probabilité *Backward*, est notée :

$$\beta_t(i) = \mathbb{P}(\boldsymbol{Y}_{t+1}^T = \boldsymbol{y}_{t+1}^T | C_t = i), \tag{2.17}$$

où

$$\boldsymbol{Y}_a^b = (Y_a, Y_{a+1}, ..., Y_b). \tag{2.18}$$

Il ne faut pas confondre les coefficients $\beta_i (i = 1, 2, ..., p)$ d'un modèle de Poisson avec régresseurs avec les probabilités *Backward*. Le vecteur de toutes les valeurs $\beta_t(i)$ pour $i = 1, ..., m$, soit $\boldsymbol{\beta}_t$, peut s'obtenir à partir de la formule récursive suivante :

$$\boldsymbol{\beta}_t' = \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \boldsymbol{\beta}_{t+1}'. \tag{2.19}$$

Par hypothèse, $\boldsymbol{\beta}_T$ est un vecteur de taille m composé de 1 uniquement et donc le vecteur $\boldsymbol{\beta}_t$ peut aussi s'écrire ainsi :

$$\boldsymbol{\beta}_t' = \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) ... \boldsymbol{\Gamma} \boldsymbol{P}(y_T) 1'. \tag{2.20}$$

La démonstration de ces trois équations se trouve à la proposition 5 de l'annexe A.

— La probabilité d'obtenir $\boldsymbol{Y}^{(T)}$ et d'être dans l'état i en t :

$$\alpha_t(i)\beta_t(i) = \mathbb{P}\left(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, C_t = i\right). \tag{2.21}$$

La démonstration de cette équation se trouve à la proposition 6 de l'annexe A.

— La probabilité d'être à l'état i au temps t, sachant la valeur de $\boldsymbol{Y}^{(T)}$, soit les probabilités marginales *a posteriori* :

$$\hat{\zeta}_t(i) = \mathbb{P}\left(C_t = i|\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}\right) = \frac{\alpha_t(i)\beta_t(i)}{L_T}. \tag{2.22}$$

La démonstration de cette équation se trouve à la proposition 7 de l'annexe A.

— La probabilité d'être à l'état i au temps $t-1$ et à l'état j au temps t, sachant la valeur de $\boldsymbol{Y}^{(T)}$, soit les probabilités marginales jointes *a posteriori* :

$$\hat{\psi}_t(i,j) = \mathbb{P}\left(C_{t-1} = i, C_t = j|\boldsymbol{Y}^{(T)}\right) = \frac{\alpha_{t-1}(i)\gamma_{ij}p_j(y_t)\beta_t(j)}{L_T}. \tag{2.23}$$

La démonstration de cette équation se trouve à la proposition 8 de l'annexe A.

— La valeur de la fonction de vraisemblance :

$$\boldsymbol{\alpha}_t\boldsymbol{\beta}_t' = \mathbb{P}\left(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}\right) = L_T. \tag{2.24}$$

La démonstration de cette équation se trouve à la proposition 9 de l'annexe A.

Les formules récursives de cette paramétrisation viennent diminuer la difficulté du calcul du maximum de vraisemblance.

2.4 Estimation du modèle HHM simple

Pour utiliser un modèle HMM afin de faire des prévision à partir de celui-ci, les paramètres du modèle sont nécessaires. Ceux-ci sont donc estimés par les EMV

(voir la sous-section 1.1.3). Pour estimer la valeur des EMV d'un modèle HMM, l'algorithme itératif Espérance-Maximisation (EM) est appliqué sur celui-ci. Pour débuter l'algorithme EM, des valeurs initiales de paramètres sont nécessaires.

Puisque qu'une des étapes de l'algorithme EM requiert la réestimation de nouvelles valeurs au vecteur de paramètre $\boldsymbol{\theta}$ à chaque itération, celui-ci est réécrit comme, $\hat{\boldsymbol{\theta}}^{(n)}$. Le vecteur $\hat{\boldsymbol{\theta}}^{(n)}$ représente les valeurs que l'algorithme itératif utilise à la $(n+1)$-ième itération. Les paramètres initiaux représentent donc le vecteur $\hat{\boldsymbol{\theta}}^{(0)}$ utilisé à la première itération de l'algorithme.

2.4.1 Paramètres initiaux de l'algorithme EM, utilisés dans l'estimation des EMV pour le modèle HMM simple

Pour un modèle HMM simple, le vecteur $\hat{\boldsymbol{\theta}}^{(0)}$ est construit ainsi :

- Pour la matrice de transition $\boldsymbol{\Gamma}$, les probabilités γ_{ii} (pour $i = 1, ..., m$) seront égales à 0.9 et les γ_{ij} (pour $i \neq j$ et $i, j = 1, ..., m$) à $\frac{0.1}{m-1}$. L'attribution d'un poids plus élevé aux paramètres γ_{ii} vient de l'hypothèse qu'un processus aura une plus grande probabilité de conserver son état actuel (stationnarité du processus) (voir Tsay, 2005). Les transitions γ_{ii} ont donc, en général, des probabilités supérieures aux transitions γ_{ij}. La matrice initiale de $\boldsymbol{\Gamma}$ ressemblera à la matrice suivante :

$$\begin{bmatrix} 0.9 & \frac{0.1}{m-1} & \frac{0.1}{m-1} & \cdots & \frac{0.1}{m-1} \\ \frac{0.1}{m-1} & 0.9 & \frac{0.1}{m-1} & \cdots & \frac{0.1}{m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{0.1}{m-1} & \frac{0.1}{m-1} & \frac{0.1}{m-1} & \cdots & 0.9 \end{bmatrix}. \tag{2.25}$$

- Avec l'équation (2.6), le vecteur $\boldsymbol{\delta}$ est déterminé.
- Pour ce qui est des paramètres des distributions $Y_t|C_t = i$ $(i = 1, 2, ..., m)$, les valeurs initiales peuvent être estimées à partir de la méthode des moments.

2.4.2 Algorithme Espérance-Maximisation

L'algorithme Espérance-Maximisation (EM) appliqué à un modèle HMM est une méthode itérative développée par Baum *et al.* (1970). Celle-ci est utilisée lorsque certaines données sont manquantes pour calculer les EMV. Dans le cas des HMM, les données manquantes sont les états possibles de la série inobservée $\boldsymbol{C}$. L'algorithme EM s'effectue en deux étapes. Premièrement, le calcul de l'espérance conditionnelle de la log vraisemblance des données manquantes en utilisant les paramètres $\hat{\boldsymbol{\theta}}^{(n-1)}$ est effectué. Cette étape sera dénoté étape E, soit pour représenter le E de l'algorithme EM. Par la suite, la maximisation de l'espérance, obtenue à l'étape E, est effectuée afin de trouver les nouveaux paramètres $\hat{\boldsymbol{\theta}}^{(n)}$. Cette étape sera dénoté étape M, soit pour représenter le M de l'algorithme EM. Les paramètres $\hat{\boldsymbol{\theta}}^{(n)}$, obtenues à la fin de l'étape M, sont donc les paramètres maximisant l'espérance de l'étape E. Ces 2 étapes sont répétées jusqu'à ce qu'il y ait convergence des EMV.

Pour trouver les EMV du modèle HMM, la maximisation de l_T par rapport à un élément de $\boldsymbol{\theta}$ est requis. Par conséquent, le calcul suivant est à effectuer (à partir de l'équation (2.7)) :

$$\underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log \left(\sum_{\boldsymbol{c}^{(T)}} \mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right) \right).$$

L'utilisation de l'algorithme EM requiert seulement la maximisation du CDLL (voir équation (2.12)) pour trouver les EMV. Il n'est pas trivial de voir que l'algorithme EM maximise la fonction de vraisemblance en utilisant le CDLL. C'est pourquoi la dérivation de l'algorithme EM est démontrée. En partant de l'équation

(2.7), il est possible de dire que :

$$\begin{aligned}
\log \mathbb{P}(\boldsymbol{Y}^{(T)}|\boldsymbol{\theta}) &= l_T(\boldsymbol{\theta}) \\
&= \log \sum_{\boldsymbol{c}^{(T)}} \mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right) \\
&= \log \sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \frac{\mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right)}{q(\boldsymbol{C}^{(T)})} \\
&\geq \sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \frac{\mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right)}{q(\boldsymbol{C}^{(T)})}. \\
&=: \mathcal{L}(q, \boldsymbol{\theta})
\end{aligned} \tag{2.26}$$

L'inégalité ci-dessus vient de l'utilisation de l'inégalité de Jensen (1906) qui démontre que :

$$\zeta(\mathbb{E}[X]) \leq \mathbb{E}[\zeta(X)],$$

où ζ est une fonction convexe quelconque. L'expression $\mathcal{L}(q, \boldsymbol{\theta})$ est donc une borne inférieure de $\log \mathbb{P}(\boldsymbol{Y}^{(T)}|\boldsymbol{\theta})$. Il est à noter que la fonction q est une fonction de masse quelconque de $\boldsymbol{C}^{(T)}$. L'expression de $\mathcal{L}(q, \boldsymbol{\theta})$ est donc divisée en 2 parties.

$$\mathcal{L}(q, \boldsymbol{\theta}) = \underbrace{\sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right)}_{\mathbb{E}_{q(\boldsymbol{C}^{(T)})}\,[\text{CDLL}]} - \underbrace{\sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log q(\boldsymbol{C}^{(T)})}_{\boldsymbol{\theta}\text{ absent}}. \tag{2.27}$$

Pour obtenir les EMV, la fonction de vraisemblance est maximisée en dérivant par rapport à chaque élément du vecteur $\boldsymbol{\theta}$. Dans l'équation (2.27), la deuxième partie de l'équation ne dépend pas de $\boldsymbol{\theta}$. La première partie de l'équation est le CDLL (voir la section 2.2). La maximisation de la vraisemblance revient donc à maximiser l'espérance du CDLL.

Bien que l'équation obtenue soit beaucoup plus simple à maximiser que l'équation (2.7), la distribution $q(\boldsymbol{C}^{(T)})$ qui maximise L_T n'est pas connue. Comme mentionné ci-dessus, l'algorithme EM se divise en 2 parties. À l'étape E de l'algorithme EM, la

valeur $q(\boldsymbol{C}^{(T)})$ qui maximise $\mathcal{L}(q, \boldsymbol{\theta})$ est donc déduite. Ainsi à partir de l'équation (2.26) :

$$\begin{aligned}
&\sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \frac{\mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right)}{q(\boldsymbol{C}^{(T)})} \\
&= \sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \frac{\mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}, \boldsymbol{Y}^{(T)}\right) \mathbb{P}\left(\boldsymbol{Y}^{(T)} \middle| \boldsymbol{\theta}\right)}{q(\boldsymbol{C}^{(T)})} \\
&= \sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \frac{\mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}, \boldsymbol{Y}^{(T)}\right)}{q(\boldsymbol{C}^{(T)})} + \sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \mathbb{P}\left(\boldsymbol{Y}^{(T)} \middle| \boldsymbol{\theta}\right) \\
&= \underbrace{-KL\left(q(\boldsymbol{C}^{(T)}) || \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}, \boldsymbol{Y}^{(T)}\right)\right)}_{\text{Kullback-Leibler } (KL(q||p))} + \underbrace{\log \mathbb{P}\left(\boldsymbol{Y}^{(T)} \middle| \boldsymbol{\theta}\right)}_{q(\boldsymbol{C}^{(T)}) \text{ absent}}, \qquad (2.28)
\end{aligned}$$

où $KL(q||p)$ est la divergence de Kullback-Leibler (voir Byrne et Eggermont, 2011). L'équation (2.28) représente une deuxième méthode pour exprimer $\mathcal{L}(q, \boldsymbol{\theta})$. Si cette équation est maximisée par rapport à $q(\boldsymbol{C}^{(T)})$, alors

$$q(\boldsymbol{C}^{(T)}) = \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right).$$

Le paramètre $\hat{\boldsymbol{\theta}}^{(n-1)}$ représente le vecteur des paramètres obtenus à la fin de l'itération $n-1$ de l'algorithme EM. Si l'équation (2.28) est dérivée par rapport à $q(\boldsymbol{C}^{(T)})$, la seconde partie devient 0 et la première partie est maximisée lorsque $q(\boldsymbol{C}^{(T)}) = \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right)$. Ceci découle des propriétés de KL (voir Byrne et Eggermont, 2011). En effet, le résultat de $KL(q||p)$ sera toujours plus grand que 0, sauf lorsque q est égale à p. Dans ce cas particulier, la valeur de KL est 0. Le maximum de $KL(q||p)$ (dérivé par rapport à q) est obtenu en utilisant les multiplicateurs de Lagrange (voir 1.3). Le Lagrangien à résoudre est le suivant :

$$W(q, \lambda_1) = KL(p||q) + \lambda_1(1 - \sum_i q_i) = \sum_i q_i \log \frac{q_i}{p_i} + \lambda_1(1 - \sum_i q_i).$$

Si $W(q, \lambda_1)$ est dérivé par rapport à q_i et λ_i, alors :

$$\left.\begin{array}{l} \frac{\partial W}{\partial q_i} = \log q_i - \log p_i + 1 - \lambda_1 = 0 \Rightarrow q_i = p_i \exp(\lambda_1 - 1) \\ \frac{\partial W}{\partial \lambda_1} = 1 - \sum_i q_i = 0 \Rightarrow \sum_i q_i = 1 \end{array}\right\} \Rightarrow q_i = p_i.$$

Lorsque l'équation (2.28) est maximisée par rapport à $q(\boldsymbol{C}^{(T)})$, il s'ensuit que :

$$q(\boldsymbol{C}^{(T)}) = \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right).$$

Si $q(\boldsymbol{C}^{(T)}) = \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right)$, alors à partir de l'équation (2.28), l'égalité suivante est obtenue :

$$\begin{aligned}
\mathcal{L}(q, \theta) &= -KL\left(q(\boldsymbol{C}^{(T)}) || \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right)\right) + \log \mathbb{P}\left(\boldsymbol{Y}^{(T)} \middle| \boldsymbol{\theta}\right) \\
&= -KL\left(\mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right) || \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right)\right) \\
&\quad + \log \mathbb{P}\left(\boldsymbol{Y}^{(T)} \middle| \boldsymbol{\theta}\right) \\
&= \log \mathbb{P}\left(\boldsymbol{Y}^{(T)} \middle| \boldsymbol{\theta}\right) = l_T.
\end{aligned}$$

L'étape E, de l'algorithme EM, assure donc que la nouvelle log-vraisemblance est au moins égale à l_T lorsque $q(\boldsymbol{C}^{(T)}) = \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right)$. Maintenant que les $q(\boldsymbol{C}^{(T)})$ sont connus, ces valeurs sont incorporées dans l'équation (2.27) pour obtenir les nouvelles valeurs de $\boldsymbol{\theta}$ (soit $\hat{\boldsymbol{\theta}}^{(n)}$). L'étape M de l'algorithme EM est effectuée en dérivant $\mathbb{E}_{q(\boldsymbol{C}^{(T)})}(CDLL)$ par rapport à un élément du vecteur $\boldsymbol{\theta}$, soit

$$\begin{aligned}
&\underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{\boldsymbol{c}^{(T)}} q(\boldsymbol{C}^{(T)}) \log \mathbb{P}\left(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)} \middle| \boldsymbol{\theta}\right) \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathbb{E}_{q(\boldsymbol{C}^{(T)})}(CDLL).
\end{aligned}$$

Ceci termine une itération de l'algorithme EM, ce qui engendre l'inégalité suivante :

$$l_T(\hat{\boldsymbol{\theta}}^{(n-1)}) \underset{\text{Étape E}}{=} \mathcal{L}(q^{(n)}, \hat{\boldsymbol{\theta}}^{(n-1)}) \underset{\text{Étape M}}{\leq} \mathcal{L}(q^{(n)}, \hat{\boldsymbol{\theta}}^{(n)}) \underset{\text{Jensen}}{\leq} l_T(\hat{\boldsymbol{\theta}}^{(n)}).$$

Il ne reste plus qu'à itérer entre les étapes E et M jusqu'à convergence du vecteur $\hat{\boldsymbol{\theta}}$.

Pour résumer l'algorithme EM se construit en deux étapes répétées jusqu'à la convergence :

1) **Calcul de l'Espérance :** Comme mentionné ci-dessus, l'étape d'Espérance de l'algorithme EM consiste à trouver $\mathbb{E}_{q(\boldsymbol{C}^{(T)})}[CDLL]$, où $q(\boldsymbol{C}^{(T)}) = \mathbb{P}\left(\boldsymbol{C}^{(T)} \middle| \hat{\boldsymbol{\theta}}^{(n-1)}, \boldsymbol{Y}^{(T)}\right)$, à partir de la valeur des paramètres trouvée par la maximisation de l'itération précédente ($\hat{\boldsymbol{\theta}}^{(n-1)}$). Lorsqu'il s'agit de la première itération, les valeurs sont telles que définies à la sous-section 2.4.

2) **Maximisation :** L'étape de Maximisation de l'algorithme EM utilise $\mathbb{E}_{q(\boldsymbol{C}^{(T)})}[CDLL]$ et maximise celle-ci par rapport à $\boldsymbol{\theta}$ pour trouver de nouveau $\hat{\boldsymbol{\theta}}^{(n)}$. Ici, c'est le CDLL qui est maximisé, et non la fonction de la log-vraisemblance, car l'algorithme EM admet l'hypothèse que la distribution des C_t $(t = 1, ..., T)$, conditionnelle aux $\boldsymbol{Y}^{(T)}$ et aux paramètres $\hat{\boldsymbol{\theta}}^{(n-1)}$, soit $(C_t|\boldsymbol{Y}^{(T)}, \hat{\boldsymbol{\theta}}^{(n-1)})$, est connue.

L'algorithme EM passe par les étapes d'Espérance (1) et de Maximisation (2), et ce, jusqu'à convergence des paramètres (EMV). La convergence de la log-vraisemblance peut être utilisée pour déterminer la convergence de l'algorithme.

2.5 Application

L'algorithme EM est utilisé pour trouver les EMV d'un modèle HMM.

Les étapes du calcul de l'espérance et de la maximisation de l'algoritme EM appliqué au modèle HMM sont donc définies ainsi :

Calcul de l'Espérance : Les paramètres ($\hat{\boldsymbol{\theta}}^{(n-1)}$) sont utilisés pour trouver la valeur des probabilités *Forward* et *Backward* à la n^e itération (voir 2.3). Pour la première itération, le vecteur $\hat{\boldsymbol{\theta}}^{(0)}$ est constitué des paramètres initiaux. Avec les probabilités *Forward* et *Backward*, les valeurs de $\hat{\zeta}_t(i)$ et $\hat{\psi}_t(i,j)$ sont déterminées (voir équations (2.22) et (2.23)). Ces valeurs sont essentielles, car $\mathbb{E}_{q(\boldsymbol{C}^{(T)})}[CDLL]$

doit être résolu. À partir de l'équation (2.13), il s'ensuit que :

$$\begin{aligned}
\mathbb{E}_{q(\boldsymbol{C}^{(T)})}[CDLL] &= \mathbb{E}_{q(\boldsymbol{C}^{(T)})}\left[\sum_{i=1}^{m}\zeta_0(i)\log\delta_i + \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\sum_{t=1}^{T}\psi_t(i,j)\right)\log\gamma_{ij}\right. \\
&\qquad \left. + \sum_{i=1}^{m}\sum_{t=1}^{T}\zeta_t(i)\log p_i(y_t)\right] \\
&= \sum_{i=1}^{m}\hat{\zeta}_0(i)\log\delta_i + \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\sum_{t=1}^{T}\hat{\psi}_t(i,j)\right)\log\gamma_{ij} \\
&\qquad + \sum_{i=1}^{m}\sum_{t=1}^{T}\hat{\zeta}_t(i)\log p_i(y_t) \\
&= \text{partie 1} + \text{partie 2} + \text{partie 3}. \qquad (2.29)
\end{aligned}$$

Maximisation : Si $\boldsymbol{\delta}$ est un vecteur de paramètres à déterminer par maximisation de la vraisemblance (et non déterminé par la stationnarité de $\boldsymbol{\Gamma}$), alors le CDLL se divise en 3 parties indépendantes les unes des autres (équation (2.29)). En effet, la partie 1 ne dépend que des δ_i $(i = 1, 2, ..., m)$, la partie 2 de $\boldsymbol{\Gamma}$ et la partie 3 des $p_i(y_t)$ (soit tous les paramètres des distributions $Y_t|C_t = i$ $(i = 1, 2, ..., m)$). Écrit sous cette forme, il est simple de maximiser chacune des 3 parties en fonction des paramètres dont elles dépendent. Les trois maximisations suivantes sont effectuées :

1. Maximiser $\sum_{i=1}^{m}\hat{\zeta}_0(i)\log\delta_i$ (partie 1) par rapport à δ_i. Soit $\boldsymbol{\delta} = \frac{\hat{\boldsymbol{\zeta}}_0}{\sum_{j=1}^{m}\hat{\zeta}_0(i)}$.
2. Maximiser $\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\sum_{t=1}^{T}\hat{\psi}_t(i,j)\right)\log\gamma_{ij}$ (partie 2) par rapport aux γ_{ij}. Soit $\gamma_{ij} = f_{ij}/\sum_{k=1}^{m}f_{ij}$, où $f_{ij} = \sum_{t=1}^{T}\hat{\psi}_t(i,j)$.
3. Finalement, maximiser $\sum_{i=1}^{m}\sum_{t=1}^{T}\hat{\zeta}_t(i)\log p_i(y_t)$ (partie 3) par rapport à (aux) paramètre(s) des distributions $p_i(y_t)$. Les distributions $p_i(y_t)$ varient en fonction des hypothèses choisies.

Dans ce mémoire, il est admis que $\boldsymbol{\delta}$ est le vecteur stationnaire de $\boldsymbol{\Gamma}$ tel qu'indiqué à la section 2.3, donc la partie 1 et la partie 2, de l'équation (2.29), doivent

être regroupées. Puisque la maximisation des deux parties conjointement est complexe, un algorithme de maximisation peut être utilisé pour trouver les EMV (`fminsearch` de *Matlab* par exemple).

2.5.1 HMM simple - Poisson avec régresseurs

Pour différencier les paramètres λ_t de chacune des distributions $Y_t|C_t = i$ $(i = 1, 2, ..., m)$, l'indice i est ajouté au paramètre λ_t, soit $\lambda_{t,i}$.

Si $Y_t|C_t = i$ $(i = 1, 2, ..., m)$ suit une loi de Poisson avec régresseurs et que le paramètre $\lambda_{t,i}$ est égal à :

$$\lambda_{t,i} = \exp(\boldsymbol{x}_t'\boldsymbol{\beta})F_i = \lambda_t F_i,$$

alors le paramètre F_i peut être vu comme un nouveau régresseur qui sera différent pour chacune des distributions de $Y_t|C_t = i$ $(i = 1, 2, ..., m)$. Ce paramètre amène donc une différence dans la valeur du paramètre $\lambda_{t,i}$ et représente ainsi plusieurs états d'un modèle HMM. Les p paramètres de $\boldsymbol{\beta}$ (régresseurs) sont les mêmes, peu importe l'état de la série inobservée.

La partie 1 et la partie 2 de l'équation (2.29) seront toujours les mêmes, dans un modèle HMM simple, peu importe le modèle utilisé pour les distributions conditionnelles des $\boldsymbol{Y}^{(T)}$. Il ne reste donc plus qu'à définir la partie 3 lorsque les distributions conditionnelles des $\boldsymbol{Y}^{(T)}$ suivent une loi de Poisson avec régresseurs (qui sera nommé HMM simple - Poisson avec régresseurs), soit :

$$\begin{aligned} CDLL_3 &= \sum_{i=1}^{m}\sum_{t=1}^{T} \hat{\zeta}_t(i) \log p_i(y_t) \\ &= \sum_{i=1}^{m}\sum_{t=1}^{T} \hat{\zeta}_t(i) \log \left[\frac{e^{-\lambda_t F_i}(\lambda_t F_i)^{y_t}}{y_t!}\right] \\ &= \sum_{i=1}^{m}\sum_{t=1}^{T} \hat{\zeta}_t(i) \left[-\lambda_t F_i + y_t \log(\lambda_t F_i) - \log(y_t!)\right]. \end{aligned}$$

La log-vraisemblance partielle de la 3^e partie (équation (2.29)) du CDLL est notée $CDLL_3$. L'équation ci-dessus est donc à maximiser par rapport à $\boldsymbol{\beta}$ et F_i. Il n'y a pas de condition sur les paramètres β_i, mais les valeurs de F_i doivent toutes être positives. Une simple modification de F_i pour $F_i = \exp(F_i')$ permet d'utiliser une fonction de maximisation sans condition qui, en général, converge beaucoup plus vite.

Pour initialiser l'algorithme EM pour le modèle HMM simple - Poisson avec régresseurs, les valeurs des paramètres suivants sont utilisées :

i) Le vecteur $\boldsymbol{\delta}$ et la matrice $\boldsymbol{\Gamma}$ seront tels que définis dans la section 2.4.

ii) Puisque $\lambda_{t,i} = \exp(\boldsymbol{x}_t'\boldsymbol{\beta})F_i$, il y a donc p régresseurs à définir et m paramètres F_i $(i = 1, 2, ..., m)$. Pour la valeur initiale des régresseurs, une analyse préliminaire peut être effectuée. En effet, la valeur initiale de ces paramètres peut être obtenue en modélisant la série de réalisations par rapport à un modèle de distribution Poisson avec régresseurs (voir section 1.1.4).

iii) Pour les paramètres F_i, il n'est pas toujours évident de trouver des valeurs initiales qui seront idéales. Puisque les paramètres initiaux des régresseurs représentent en quelque sorte la moyenne, soit $\hat{\lambda}_t = \exp(\boldsymbol{x}_t'\hat{\boldsymbol{\beta}})$, si $m = 3$ (nombre d'états), les valeurs initiales des F_i peuvent, par exemple, être les suivantes :

$$\begin{bmatrix} 0.5 & 1 & 1.5 \end{bmatrix}.$$

Ces trois valeurs représentent, dans cet exemple, un état avec un résultat plus petit, égal et plus grand que la moyenne. Soit trois distributions de Poisson avec une moyenne de $0.5\hat{\lambda}_t$, $\hat{\lambda}_t$ et $1.5\hat{\lambda}_t$.

2.5.2 Représentation graphique du modèle HMM simple

Une fois les EMV trouvés, ces paramètres peuvent être utilisés pour faire une représentation graphique du HMM. Il est intéressant de comparer la série temporelle

de données à la série temporelle obtenue grâce au modèle HMM. Pour obtenir la valeur moyenne du HMM à un temps t (pour $t = 1, ..., T$) la technique du filtrage est utilisée.

Pour l'évaluation au temps t, seulement les réalisations $y_1, y_2, ...y_{t-1}(\boldsymbol{y}^{(t-1)})$ sont utilisées. La valeur de $\mathbb{E}[Y_t|\boldsymbol{Y}^{(t-1)}]$ est donc recherchée. La distribution conditionnelle est la suivante :

$$\begin{aligned}\mathbb{P}(Y_t = y|\boldsymbol{Y}^{(t-1)}) &= \frac{\mathbb{P}(Y_t = y, \boldsymbol{Y}^{(t-1)})}{\mathbb{P}(\boldsymbol{Y}^{(t-1)})} \\ &= \frac{\boldsymbol{\delta\Gamma P}(y_1)...\boldsymbol{\Gamma P}(y)\mathbf{1'}}{\boldsymbol{\delta\Gamma P}(y_1)...\boldsymbol{\Gamma P}(y_{t-1})\mathbf{1'}} \\ &= \frac{\boldsymbol{\alpha}_{t-1}\boldsymbol{\Gamma P}(y)\mathbf{1'}}{\boldsymbol{\alpha}_{t-1}\mathbf{1'}}.\end{aligned}$$

Si $\boldsymbol{k} = \boldsymbol{\alpha}_{t-1}/\boldsymbol{\alpha}_{t-1}\mathbf{1'}$ et que $\boldsymbol{l} = \boldsymbol{k\Gamma}$ alors :

$$\mathbb{P}(Y_t = y|\boldsymbol{Y}^{(t-1)}) = \sum_i^m l_i p_i(y),$$

où l_i est l'élément i du vecteur $\boldsymbol{l}$. Avec la probabilité conditionnelle, il s'ensuit que :

$$\mathbb{P}(Y_t \leq y|\boldsymbol{Y}^{(t-1)}) = \sum_{j=0}^{y}\sum_i^m l_i p_i(j) \tag{2.30}$$

$$\mathbb{E}[Y_t|\boldsymbol{Y}^{(t-1)}] = \sum_j j \cdot \sum_i^m l_i p_i(j) \tag{2.31}$$

$$Var[Y_t|\boldsymbol{Y}^{(t-1)}] = \sum_j j^2 \cdot \sum_i^m l_i p_i(j) - \mathbb{E}[Y_t|\boldsymbol{Y}^{(t-1)}]^2.$$

2.6 Chaînes de Markov cachées du second ordre

Il est possible d'utiliser exactement tout le contenu de ce chapitre afin de créer des chaînes de Markov cachées du second ordre. Une chaîne de Markov cachée du second ordre admet que la valeur observée au temps t dépend des états inobservés

au temps t et $t-1$. La seule différence dans la construction de ce modèle, en comparaison au modèle HMM simple, est la matrice $\mathbf{\Gamma}$. Une chaîne de Markov cachée du second ordre admet que :

$$\mathbb{P}(C_t = k | C_{t-1} = j, C_{t-2} = i) \neq \mathbb{P}(C_t = k | C_{t-1} = j)$$

La matrice $\mathbf{\Gamma}$ pour ce modèle est donc la suivante :

$$\mathbf{\Gamma} = \begin{array}{c} \\ (1,1) \\ (1,2) \\ (1,3) \\ \vdots \\ (m,m) \end{array} \begin{array}{c} \begin{array}{cccccccccccc} (1,1) & (1,2) & \dots & (1,m) & (2,1) & (2,2) & \dots & (2,m) & (3,1) & \dots & (m,m-1) & (m,m) \end{array} \\ \begin{bmatrix} \gamma_{111} & \gamma_{112} & \dots & \gamma_{11m} & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & \gamma_{121} & \gamma_{122} & \dots & \gamma_{12m} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \gamma_{131} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \gamma_{mm(m-1)} & \gamma_{mmm} \end{bmatrix} \end{array}$$

où γ_{ijk} représente la probabilité de transition à l'état k au temps t sachant qu'à la période $t-2$ et $t-1$ le processus était à l'état i et j respectivement. Dans cette matrice, il doit toujours y avoir un lien entre les couples d'états de départ et d'arrivée. Par exemple de la cellule (1,1) seulement les transitions aux cellules $(1,x)$ (pour $x = 1, ..., m$) sont possibles. Autre exemple, si l'état actuel est (10,8), la transition à (8,2) est possible et représente donc la probabilité de transition à l'état 2 au temps t sachant qu'à la période $t-2$ et $t-1$ le processus était à l'état 10 et 8 respectivement, soit $\gamma_{10,8,2}$.

Les figures 2.2 et 2.3 illustrent le schéma de transition de modèles HMM du second ordre :

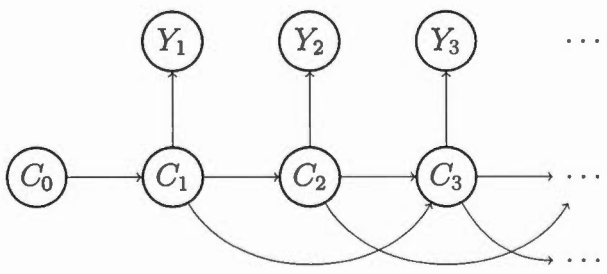


Figure 2.2 Schéma de transition modèle HMM second ordre (distribution conditionnelle des valeurs observées : $Y_t | C_t = i$ $(i = 1, 2, ..., m)$)

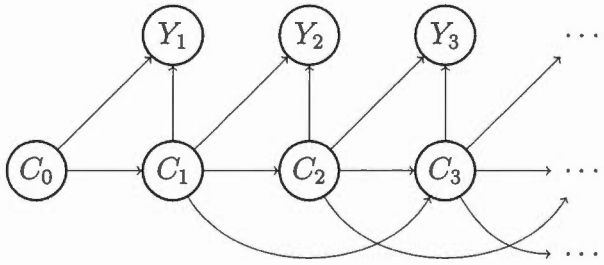


Figure 2.3 Schéma de transition modèle HMM second ordre (distribution conditionnelle des valeurs observées : $Y_t|C_{t-1} = i, C_t = j$ $(i, j = 1, 2, ..., m)$)

Ces deux figures sont des modèles HMM du second ordre tel que défini ci-dessus, mais la seconde figure admet de plus que la distribution des valeurs observées varie par rapport à l'état actuel et l'état précédent, soit $Y_t|C_{t-1} = i, C_t = j$ $(i, j = 1, 2, ..., m)$.

La modélisation du second modèle ci-dessus est semblable à celle d'un modèle HMM du second ordre simple. En effet, il y a m^2 états qui sont reliés à $m^2 n$ valeurs de paramètres de la distribution conditionnelle des valeurs observées (où n est le nombre de paramètres d'une distribution conditionnelle des valeurs observées). La matrice $\mathbf{\Gamma}$ est donc modifiée comme mentionnée ci-dessus et la matrice $\boldsymbol{P}$ est aussi modifiée, soit :

$$\boldsymbol{P}(y_t) = \begin{bmatrix} p_1(y_t) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & p_{m^2}(y_t) \end{bmatrix}. \tag{2.32}$$

Les modèles du présent chapitre seront appliqués sur des données financières au chapitre 4.

CHAPITRE III

CHAÎNES DE MARKOV COUPLES

3.1 Notions de base

Les chaînes de Markov couples (*Pairwise Markov Chain* - PMC) sont des modèles plus généraux que les modèles HMM et ont été présentés pour la première fois par Pieczynski (2003). La figure 3.1 illustre le schéma de transition d'un modèle PMC général (voir Rafi *et al.*, 2011), soit :

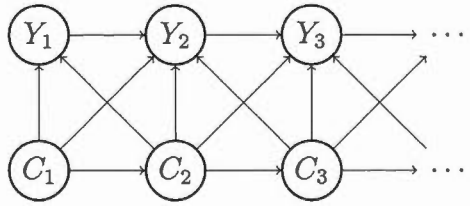


Figure 3.1 Schéma de transition modèle PMC général

Dans la figure 3.1, il est possible de constater qu'il existe un lien entre Y_t et Y_{t-1}, mais aussi un lien entre Y_t, C_t et C_{t-1}. Le modèle PMC voit la distribution de Y_t et C_t comme un couple qui sera dénoté $\boldsymbol{Z_t} = (Y_t, C_t)$. Malgré la structure quelque peu différente de celle d'un HMM, le modèle PMC utilise les mêmes concepts de probabilité *Forward* et *Backward* ainsi que plusieurs autres caractéristiques se rattachant aux modèles HMM.

La fonction de vraisemblance de ce modèle est la suivante :

$$L_T = \sum_{c^{(T)}} \mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)\cdots\mathbb{P}(\boldsymbol{Z}_T|\boldsymbol{Z}_{T-1}). \tag{3.1}$$

La fonction de masse de probabilité $\boldsymbol{Z}_t$ conditionnelle à $\boldsymbol{Z}_{t-1}$ est la suivante :

$$\begin{aligned}\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1}) &= \frac{\mathbb{P}(\boldsymbol{Z}_{t-1}, \boldsymbol{Z}_t)}{\mathbb{P}(\boldsymbol{Z}_{t-1})} \\ &= \frac{\mathbb{P}(C_{t-1}, Y_{t-1}; C_t, Y_t)}{\mathbb{P}(C_{t-1}, Y_{t-1})} \\ &= \frac{\mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t})\mathbb{P}(C_{t-1}, C_t)}{\mathbb{P}(C_{t-1}, Y_{t-1})},\end{aligned} \tag{3.2}$$

où

$$\boldsymbol{Y}_{t-1:t} = Y_{t-1}, Y_t,$$

et

$$\boldsymbol{C}_{t-1:t} = C_{t-1}, C_t.$$

Si les états $\boldsymbol{C}^{(T)}$ sont connus, alors le CDL est égal à (avec les équations (3.1) et (3.2)) :

$$\begin{aligned}CDL &= \mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)\cdots\mathbb{P}(\boldsymbol{Z}_T|\boldsymbol{Z}_{T-1}) \\ &= \prod_{t=2}^{T}\mathbb{P}(\boldsymbol{C}_{t-1:t}) \times \frac{1}{\prod_{t=2}^{T-1}\mathbb{P}(\boldsymbol{Z}_t)} \times \prod_{t=2}^{T}\mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t}).\end{aligned} \tag{3.3}$$

Puisque les modèles PMC peuvent admettre une dépendance entre Y_t et Y_{t-1} et que la fonction de masse de probabilité $\boldsymbol{Z}_t$ conditionnelle à $\boldsymbol{Z}_{t-1}$ nécessite la fonction de masse de probabilité $\boldsymbol{Z}_{t-1}$, alors il est nécessaire de définir cette fonction de masse de probabilité. Sachant que :

$$\mathbb{P}(\boldsymbol{Z}_{t-1}, \boldsymbol{Z}_t) = \mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t})\mathbb{P}(C_{t-1}, C_t), \tag{3.4}$$

alors :

$$\begin{aligned}\mathbb{P}(\boldsymbol{Z}_{t-1}) &= \sum_{\boldsymbol{Z}_t}\mathbb{P}(\boldsymbol{Z}_{t-1}, \boldsymbol{Z}_t) \\ &= \sum_{j\in\Omega}\mathbb{P}(C_{t-1}, C_t = j)\sum_{m} f_{i,j}(l, m),\end{aligned}$$

où $f_{i,j}(l,m) = \mathbb{P}(Y_{t-1} = l, Y_t = m | C_{t-1} = i, C_t = j)$. À la section 2.4, l'algorithme EM a été développé de manière générale afin que celui-ci s'applique sur quelconque modèle ayant des paramètres cachés. Les étapes de l'algorithme EM à appliquer sont donc les mêmes pour trouver les EMV du modèle PMC. C'est pour cette raison que la section estimation n'est pas présente dans de chapitre. Pour l'étape E de l'algorithme EM, les probabilités *Forward* et *Backward* sont à déterminer afin de trouver les paramètres $\hat{\zeta}_t(i)$ et $\hat{\psi}_t(i,j)$.

3.2 Log-vraisemblance des données complètes du modèle PMC

À partir de la définition de L_T (équation (3.1)), la fonction de la log-vraisemblance des données complètes (CDLL) est définie par :

$$\begin{aligned} CDLL &= \log\left(\mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)\cdots\mathbb{P}(\boldsymbol{Z}_T|\boldsymbol{Z}_{T-1})\right) \\ &= \log\left(\prod_{t=2}^{T}\mathbb{P}(\boldsymbol{C}_{t-1:t}) \times \frac{1}{\prod_{t=2}^{T-1}\mathbb{P}(\boldsymbol{Z}_t)} \times \prod_{t=2}^{T}\mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t})\right) \\ &= \sum_{t=2}^{T}\log\mathbb{P}(\boldsymbol{C}_{t-1:t}) - \sum_{t=2}^{T-1}\log\mathbb{P}(\boldsymbol{Z}_t) + \sum_{t=2}^{T}\log\mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t}) \\ &= \sum_{t=2}^{T}\log c_{ij} - \sum_{t=2}^{T-1}\log\mathbb{P}(C_t, Y_t) + \sum_{t=2}^{T}\log\mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t}), \end{aligned} \tag{3.5}$$

où $\mathbb{P}(\boldsymbol{C}_{t-1:t}) = c_{ij}$. En utilisant la même logique que celle appliquée à la section 2.2, les paramètres suivants sont énoncés :

— $\zeta_t(i) = \begin{cases} 1, & \text{si } C_t = i \\ 0, & \text{sinon.} \end{cases}$

— $\psi_t(i,j) = \begin{cases} 1, & \text{si } C_{t-1} = i \text{ et } C_t = j \\ 0, & \text{sinon.} \end{cases}$

Avec les paramètres ci-dessus, l'équation (3.5) peut être réécrite ainsi :

$$CDLL = \sum_{t=2}^{T} \sum_{i,j \in \Omega^2} \psi_t(i,j) \log c_{ij} - \sum_{t=2}^{T-1} \sum_{i \in \Omega} \zeta_t(i) \log \mathbb{P}(C_t, Y_t) \qquad (3.6)$$
$$+ \sum_{t=2}^{T} \sum_{i,j \in \Omega^2} \psi_t(i,j) \log \mathbb{P}(Y_{t-1}, Y_t | C_{t-1}, C_t).$$

3.3 Paramétrisation du modèle PMC

Il est possible réécrire le maximum de la fonction de vraisemblance avec une autre paramétrisation. Les notions de probabilités *Forward* et *Backward*, ainsi que plusieurs autres notions de probabilités indispensables sous cette paramétrisation, sont introduites pour les modèles PMC aux équations (3.7)-(3.12). Dans certaines situations, ces équations sont identiques aux équations du modèle HMM (voir la section 2.3).

— La probabilité d'obtenir $Y_1 = y_1, Y_2 = y_2, ..., Y_t = y_t$ et $C_t = i$, soit une probabilité *Forward* est notée (voir Pieczynski, 2003) :

$$\alpha_1(i) = \mathbb{P}(C_1 = i, Y_1) \qquad (3.7)$$
$$\alpha_{t+1}(j) = \sum_{i \in \Omega} \alpha_t(i) \mathbb{P}(\boldsymbol{Z}_{t+1} | \boldsymbol{Z}_t) \qquad (3.8)$$
$$= \mathbb{P}(C_{t+1} = j, \boldsymbol{Y}^{(t+1)}).$$

La démonstration de cette équation se trouve à la proposition 14 de l'annexe B.

— La probabilité d'obtenir $Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, ..., Y_T = y_T$ conditionnelle à $(C_t = i, Y_t)$, soit une probabilité *Backward*, est notée (voir Pieczynski,

2003) :

$$\begin{aligned}\beta_T(i) &= 1 \\ \beta_t(i) &= \sum_{j\in\Omega} \beta_{t+1}(j)\mathbb{P}(\boldsymbol{Z}_{t+1}|\boldsymbol{Z}_t) \\ &= \mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(i)).\end{aligned} \tag{3.9}$$

où $\boldsymbol{Z}_t(i) = (Y_t, C_t = i)$. La démonstration de cette équation se trouve à la proposition 15 de l'annexe B.

— La probabilité d'obtenir $\boldsymbol{Y}^{(T)}$ et d'être dans l'état i en t :

$$\alpha_t(i)\beta_t(i) = \mathbb{P}\left(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, C_t = i\right). \tag{3.10}$$

La démonstration de cette équation se trouve à la proposition 16 de l'annexe B.

— La probabilité d'être à l'état i au temps t, sachant la valeur de $\boldsymbol{Y}^{(T)}$, soit les probabilités marginales *a posteriori* (voir Lanchantin, 2006) :

$$\hat{\zeta}_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_i \alpha_t(i)\beta_t(i)}. \tag{3.11}$$

La démonstration de cette équation se trouve à la proposition 17 de l'annexe B.

— La probabilité d'être à l'état i au temps $t-1$ et à l'état j au temps t, sachant la valeur de $\boldsymbol{Y}^{(T)}$, soit les probabilités marginales jointes *a posteriori* (voir Lanchantin, 2006) :

$$\hat{\psi}_t(i,j) = \frac{\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}{\sum_{i,j\in\Omega^2} \alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}. \tag{3.12}$$

La démonstration de cette équation se trouve à la proposition 18 de l'annexe B.

— La valeur de la fonction de vraisemblance :

$$\boldsymbol{\alpha}_t\boldsymbol{\beta}_t' = \mathbb{P}\left(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}\right) = L_T. \tag{3.13}$$

La démonstration de cette équation se trouve à la proposition 19 de l'annexe B.

3.4 Applications

3.4.1 Application d'un modèle PMC

Lanchantin (2006) trouve une forme analytique de la maximisation du CDLL du modèle PMC à l'aide des multiplicateurs de Lagrange. Grâce à ce résultat, l'algorithme EM peut s'appliquer. À l'étape E de l'algorithme EM, la fonction à maximiser (l'espérance du CDLL) est déterminée. À partir de l'équation (3.6), l'espérance du CDLL est déterminée, soit

$$
\begin{aligned}
&\mathbb{E}_{q(c^{(T)})}[CDLL] \\
&= \mathbb{E}_{q(c^{(T)})}\left[\sum_{t=2}^{T}\sum_{i,j\in\Omega^2}\psi_t^{(n)}(i,j)\log c_{ij} - \sum_{t=2}^{T-1}\sum_{i\in\Omega}\zeta_t^{(n)}(i)\log\mathbb{P}(C_t,Y_t)\right. \\
&\left.+\sum_{t=2}^{T}\sum_{i,j\in\Omega^2}\psi_t^{(n)}(i,j)\log\mathbb{P}(Y_{t-1},Y_t|C_{t-1},C_t)\right] \\
&= \sum_{t=2}^{T}\sum_{i,j\in\Omega^2}\hat{\psi}_t^{(n)}(i,j)\log c_{ij} - \sum_{t=2}^{T-1}\sum_{i\in\Omega}\hat{\zeta}_t^{(n)}(i)\log\mathbb{P}(C_t,Y_t) \\
&+\sum_{t=2}^{T}\sum_{i,j\in\Omega^2}\hat{\psi}_t^{(n)}(i,j)\log\mathbb{P}(Y_{t-1},Y_t|C_{t-1},C_t) \\
&= \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}),
\end{aligned}
$$

où l'exposant (n) représente la $(n+1)$-ième itération de l'algorithme EM. Les contraintes utilisées pour le Lagrangien sont les suivantes :

i) $\sum_{(i,l)\in\Omega\times\mathbb{R}} c_{ij}\mathbb{P}(Y_t = l, Y_{t+1} = m|C_t = i, C_{t+1} = j) = b_j(m)$,

ii) $\sum_{(l,m)\in\mathbb{R}^2}\mathbb{P}(Y_t = l, Y_{t+1} = m|C_t = i, C_{t+1} = j) = 1$,

iii) $\sum_{(i,j)\in\Omega^2} c_{ij} = 1$.

Le Lagrangien est donc le suivant :

$$L(c_{ij}, f_{ij}, \lambda_1, \lambda_{j,m}, \lambda_3) = \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(n)}) - \lambda_1 \left(\sum_{(i,j)\in\Omega^2} c_{ij} - 1 \right) \tag{3.14}$$
$$- \sum_{(j,m)\in\Omega\times\mathbb{R}} \lambda_{j,m} \left(\sum_{(i,l)\in\Omega\times\mathbb{R}} c_{ij} f_{i,j}(l,m) - b_j(m) \right)$$
$$- \lambda_3 \left(\sum_{l,m} f_{i,j}(l,m) - 1 \right).$$

En dérivant par rapport à tous les termes $(c_{ij}, f_{ij}, \lambda_1, \lambda_{j,m}, \lambda_3)$ (et en utilisant les contraintes (i), (ii) et (iii)), les résultats suivants sont obtenus :

$$\begin{aligned}
\sum_{(i,j)\in\Omega^2} c_{ij} = 1 &= \sum_{(i,j)\in\Omega^2} \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_1 + \sum_{l,m\in\mathbf{R}^2} \lambda_{j,m} f_{i,j}(l,m)} \\
&= \sum_{(i,j)\in\Omega^2} \frac{T-1}{\lambda_1 + \sum_{l,m\in\mathbf{R}^2} \lambda_{j,m} f_{i,j}(l,m)}, \\
\sum_{(l,m)\in\mathbb{R}^2} f_{ij}(l,m) = 1 &= \sum_{(l,m)\in\mathbf{R}^2} \frac{\sum_{t=2:\boldsymbol{y}_{t-1:t}=(l,m)}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_3} \\
&= \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_3}.
\end{aligned}$$

De ces résultats, les formules de réestimation des paramètres suivantes sont déterminées et l'étape M de l'algorithme EM est terminée, soit :

$$c_{ij}^{(n+1)} = \frac{1}{T-1} \sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j) \tag{3.15}$$

$$f_{i,j}^{(n+1)}(l,m) = \frac{\sum_{t=2:(y_{t-1},y_t)=(l,m)}^{T} \hat{\psi}_t^{(n)}(i,j)}{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}, \tag{3.16}$$

où l'exposant $(n+1)$ représente les valeurs au début de la $(n+2)$-ième itération de l'algorithme EM, soit le vecteur $\hat{\boldsymbol{\theta}}^{(n+1)}$. Voir la proposition 10 de l'annexe B pour la démonstration de ces résultats.

3.4.2 Modèle PMC indépendant

Lanchantin (2006) définit le modèle PMC indépendant (PMC-IN). Ce modèle est un modèle PMC avec de l'indépendance entre les réalisations, soit :

$$\begin{aligned} \mathbb{P}(\boldsymbol{Y}_{t-1:t}|C_{t-1}, C_t) &= \mathbb{P}(Y_{t-1}|C_{t-1}, C_t)\mathbb{P}(Y_t|C_{t-1}, C_t) \\ f_{i,j}(l, m) &= f_i(l)f_j(m). \end{aligned} \tag{3.17}$$

L'équation (3.17) montre qu'il y a indépendance entre les réalisations et que Y_t est dépendant du passé $(t-1)$, présent (t) et future $(t+1)$ des réalisations de la série inobservée. La figure 3.2 illustre le schéma de transition d'un modèle PMC-IN.

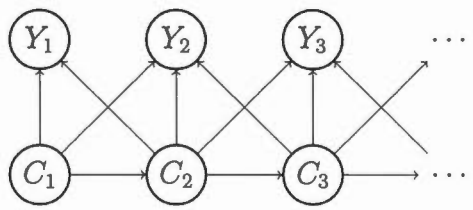


Figure 3.2 Schéma de transition modèle PMC-IN

De l'équation (3.17) et (3.2), il s'en suit que :

$$\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1}) = \frac{\mathbb{P}(Y_{t-1}|C_{t-1}, C_t)\mathbb{P}(Y_t|C_{t-1}, C_t)\mathbb{P}(C_{t-1}, C_t)}{\mathbb{P}(Y_{t-1}, C_{t-1})}.$$

En utilisant cette équation dans le CDLL du modèle PMC (voir la section 3.2),

alors :

$$\begin{aligned}
CDLL &= \log\left(\mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)\cdots\mathbb{P}(\boldsymbol{Z}_T|\boldsymbol{Z}_{T-1})\right) \\
&= \log\left(\mathbb{P}(\boldsymbol{Z}_1) \times \frac{\mathbb{P}(\boldsymbol{Z}_1,\boldsymbol{Z}_2)}{\mathbb{P}(\boldsymbol{Z}_1)}\cdots\frac{\mathbb{P}(\boldsymbol{Z}_{T-1},\boldsymbol{Z}_T)}{\mathbb{P}(\boldsymbol{Z}_{T-1})}\right) \\
&= \log\left(\prod_{t=2}^{T}\mathbb{P}(C_{t-1},C_t) \times \frac{1}{\prod_{t=2}^{T-1}\mathbb{P}(\boldsymbol{Z}_t)} \times \prod_{t=2}^{T}\mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t})\right) \\
&= \log\left(\prod_{t=2}^{T}\mathbb{P}(C_{t-1},C_t) \times \frac{1}{\prod_{t=2}^{T-1}\mathbb{P}(\boldsymbol{Z}_t)} \times \prod_{t=2}^{T}\mathbb{P}(Y_{t-1}|\boldsymbol{C}_{t-1:t})\right. \\
&\qquad \left.\times \prod_{t=2}^{T}\mathbb{P}(Y_t|\boldsymbol{C}_{t-1:t})\right) \\
&= \sum_{t=2}^{T}\log\mathbb{P}(C_{t-1},C_t) - \sum_{t=2}^{T-1}\log\mathbb{P}(\boldsymbol{Z}_t) + \sum_{t=2}^{T}\log\mathbb{P}(Y_{t-1}|\boldsymbol{C}_{t-1:t}) \\
&\qquad + \sum_{t=2}^{T}\log\mathbb{P}(Y_t|\boldsymbol{C}_{t-1:t}).
\end{aligned}$$

Dans le modèle PMC-IN, la probabilité de $\boldsymbol{Z}_{t-1}$ se calcule ainsi :

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Z}_{t-1}) &= \sum_{\boldsymbol{Z}_t}\mathbb{P}(\boldsymbol{Z}_{t-1},\boldsymbol{Z}_t) \\
&= \sum_{j\in\Omega}\mathbb{P}(C_{t-1},C_t=j)\sum_{m} f_{i,j}(l,m) \\
&= \sum_{j\in\Omega} c_{ij}\sum_{m} f_i(l)f_j(m) \\
&= \sum_{j\in\Omega} c_{ij} f_i(l).
\end{aligned}$$

Si les distributions $Y_{t-1}|C_{t-1},C_t$ et $Y_t|C_{t-1},C_t$, d'un modèle PMC-IN, sont toutes deux des distributions de Poisson avec régresseurs, alors

$$\mathbb{P}(\boldsymbol{Y}_{t-1:t}=y_{t-1:t}|C_{t-1}=i,C_t=j) = \frac{e^{-\lambda_t F^1_{ij}}\left(\lambda_t F^1_{ij}\right)^{y_{t-1}}}{y_{t-1}!}\frac{e^{-\lambda_t F^2_{ij}}\left(\lambda_t F^2_{ij}\right)^{y_t}}{y_t!}.$$

L'indice ij de F représente l'état caché au temps $t-1$ et t. Les exposants 1 et 2 de F représentent les distributions de $Y_{t-1}|C_{t-1},C_t$ et $Y_t|C_{t-1},C_t$ respectivement.

3.4.3 Représentation graphique du modèle PMC

Comme à la section 2.5.2, la technique de filtrage est appliquée afin de représenter graphiquement les modèles PMC. La valeur de $\mathbb{P}(Y_t = y|\boldsymbol{Y}^{(t-1)})$ est déterminée, soit :

$$\begin{aligned}
\mathbb{P}(Y_t = y|\boldsymbol{Y}^{(t-1)}) &= \frac{\mathbb{P}(Y_t = y, \boldsymbol{Y}^{(t-1)})}{\mathbb{P}(\boldsymbol{Y}^{(t-1)})} \\
&= \frac{\mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)...\mathbb{P}(\boldsymbol{Z}_{t-1}|\boldsymbol{Z}_{t-2})\mathbb{P}(Y_t = y, C_t|\boldsymbol{Z}_{t-1})}{\mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)...\mathbb{P}(\boldsymbol{Z}_{t-1}|\boldsymbol{Z}_{t-2})} \\
&= \frac{\alpha_{t-1}(i)\frac{\mathbb{P}(C_{t-1}, C_t, Y_{t-1}, Y_t = y)}{\mathbb{P}(C_{t-1}, Y_{t-1})}}{\sum_i \alpha_{t-1}(i)} \\
&= \frac{\alpha_{t-1}(i)\frac{\mathbb{P}(Y_t = y|C_{t-1}, C_t, Y_{t-1})\mathbb{P}(C_t|\boldsymbol{Z}_{t-1})\mathbb{P}(\boldsymbol{Z}_{t-1})}{\mathbb{P}(\boldsymbol{Z}_{t-1})}}{\sum_i \alpha_{t-1}(i)} \\
&= \frac{\alpha_{t-1}(i)\mathbb{P}(Y_t = y|C_{t-1}, C_t, Y_{t-1})\mathbb{P}(C_t|\boldsymbol{Z}_{t-1})}{\sum_i \alpha_{t-1}(i)} \\
&= \frac{\alpha_{t-1}(i)\mathbb{P}(Y_t = y|C_t, \boldsymbol{Z}_{t-1})\mathbb{P}(C_t|\boldsymbol{Z}_{t-1})}{\sum_i \alpha_{t-1}(i)},
\end{aligned}$$

où

$$\mathbb{P}(C_t|\boldsymbol{Z}_{t-1}) = \frac{c_{ij} f_{ij}(y_{t-1})}{\sum_j c_{ij} f_{ij}(y_{t-1})},$$

et

$$\mathbb{P}(Y_t = y|C_t, \boldsymbol{Z}_{t-1}) = \frac{\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})}{\mathbb{P}(C_t|\boldsymbol{Z}_{t-1})} = \frac{f_{ij}(y_{t-1}, Y_t = y)}{f_{ij}(y_{t-1})}.$$

Avec les paramètres

$$k_i = \frac{\alpha_{t-1}(i)}{\sum_i \alpha_{t-1}(i)}$$

et

$$l_i(j) = k_i\mathbb{P}(C_t|\boldsymbol{Z}_{t-1}) = k_i\mathbb{P}(C_t = j|Y_{t-1} = y_{t-1}, C_{t-1} = i),$$

alors la probabilité de Y_t conditionnelle à $\boldsymbol{Y}^{(t-1)}$ est déterminée, soit :

$$\begin{aligned}
\mathbb{P}(Y_t = y|\boldsymbol{Y}^{(t-1)}) &= \sum_{i,j} l_i(j)\mathbb{P}(Y_t = y|C_t, \boldsymbol{Z}_{t-1}) \\
&= \sum_{i,j} l_i(j)\mathbb{P}(Y_t = y|C_{t-1} = i, C_t = j, Y_{t-1} = y_{t-1}).
\end{aligned}$$

Avec la probabilité conditionnelle, il s'ensuit que :

$$\mathbb{P}(Y_t \leq y|\boldsymbol{Y}^{(t-1)}) = \sum_{k=0}^{y} \sum_{i,j} l_i(j)\mathbb{P}(Y_t = k|C_{t-1} = i, C_t = j, Y_{t-1} = y_{t-1}) \quad (3.18)$$

$$\mathbb{E}[Y_t|\boldsymbol{Y}^{(t-1)}] = \sum_{k=0}^{y} k \sum_{i,j} l_i(j)\mathbb{P}(Y_t = k|C_{t-1} = i, C_t = j, Y_{t-1} = y_{t-1}) \quad (3.19)$$

$$\begin{aligned} Var[Y_t|\boldsymbol{Y}^{(t-1)}] = &\sum_{k=0}^{y} k^2 \sum_{i,j} l_i(j)\mathbb{P}(Y_t = k|C_{t-1} = i, C_t = j, Y_{t-1} = y_{t-1}) \\ &- \mathbb{E}[Y_t|\boldsymbol{Y}^{(t-1)}]^2. \end{aligned}$$

3.5 Lien entre PMC et HMM

Afin de démontrer que le modèle PMC est plus général que le modèle HMM et aussi mieux expliquer celui-ci, le lien entre ces deux modèles sera démontré.

Comme il a été mentionné au début de la section 2.1, dans un modèle HMM simple, Y_t ne dépend que de C_t, et C_t ne dépend que de C_{t-1}. Si ces concepts sont repris dans l'équation (3.4) du modèle PMC, il est possible de dire que (voir Pieczynski, 2003) :

$$\begin{aligned} \mathbb{P}(\boldsymbol{Z}_{t-1}, \boldsymbol{Z}_t) &= \mathbb{P}(\boldsymbol{Y}_{t-1:t}|\boldsymbol{C}_{t-1:t})\mathbb{P}(C_{t-1}, C_t) \\ &= \mathbb{P}(Y_t|C_t)\mathbb{P}(Y_{t-1}|C_{t-1})\mathbb{P}(C_{t-1}, C_t). \end{aligned} \quad (3.20)$$

Si l'égalité (3.20) est remise dans la probabilité $\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})$, définie à l'équation (3.2), il s'en suit que :

$$\begin{aligned} \mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1}) &= \frac{\mathbb{P}(Y_t|C_t)\mathbb{P}(Y_{t-1}|C_{t-1})\mathbb{P}(C_{t-1}, C_t)}{\mathbb{P}(Y_{t-1}|C_{t-1})\mathbb{P}(C_{t-1})} \\ &= \frac{\mathbb{P}(Y_t|C_t)\mathbb{P}(Y_{t-1}|C_{t-1})\mathbb{P}(C_t|C_{t-1})\mathbb{P}(C_{t-1})}{\mathbb{P}(Y_{t-1}|C_{t-1})\mathbb{P}(C_{t-1})} \\ &= \mathbb{P}(Y_t|C_t)\mathbb{P}(C_t|C_{t-1}). \end{aligned} \quad (3.21)$$

À partir de l'équation du CDL du modèle PMC, équation (3.3), et de l'équation (3.21), alors le CDL d'un modèle HMM simple est retrouvé (voir équation (2.9)),

soit :

$$\begin{aligned} CDL &= \mathbb{P}(\boldsymbol{Z}_1)\mathbb{P}(\boldsymbol{Z}_2|\boldsymbol{Z}_1)\cdots\mathbb{P}(\boldsymbol{Z}_T|\boldsymbol{Z}_{T-1}) \\ &= \mathbb{P}(C_1,Y_1)\mathbb{P}(C_2|C_1)\mathbb{P}(Y_2|C_2)\mathbb{P}(C_3|C_2)\mathbb{P}(Y_3|C_3) \\ &\quad \cdots\mathbb{P}(C_T|C_{t-1})\mathbb{P}(Y_T|C_T) \\ &= \mathbb{P}(C_1)\mathbb{P}(Y_1|C_1)\mathbb{P}(C_2|C_1)\mathbb{P}(Y_2|C_2)\mathbb{P}(C_3|C_2)\mathbb{P}(Y_3|C_3) \\ &\quad \cdots\mathbb{P}(C_T|C_{t-1})\mathbb{P}(Y_T|C_T) \\ &= \mathbb{P}(C_0)\prod_{t=1}^{T}\mathbb{P}(Y_t|C_t)\mathbb{P}(C_t|C_{t-1}). \end{aligned}$$

Cette dernière égalité démontre que le modèle PMC est plus général que le modèle HMM.

3.6 Reconnaissance d'image à l'aide du modèle PMC

La majorité des publications qui traitent du modèle PMC utilise celui-ci en admettant que :

$$\boldsymbol{Y}_{t-1:t}|C_{t-1}=i,C_t=j\sim\mathcal{N}(\boldsymbol{y}_{t-1:t},\mu_{ij},\Sigma_{ij}).$$

Soit, la distribution de l'état connu sachant l'état caché est une normale bivariée de moyenne μ_{ij} et une matrice de variance-covariance Σ_{ij}. Avec les équations (3.15) et (3.16), les équations des paramètres re-estimés sont déterminés, soit :

$$\begin{aligned} c_{ij}^{(n+1)} &= \frac{1}{T-1}\sum_{t=2}^{T}\psi_t^{(n)}(i,j), \\ \mu_{ij}^{(n+1)} &= \frac{\sum_{t=2}^{T}\boldsymbol{Y}_{t-1:t}\psi_t^{(n)}(i,j)}{\sum_{t=2}^{T}\psi_t^{(n)}(i,j)}, \\ \Sigma_{ij}^{(n+1)} &= \frac{\sum_{t=2}^{T}(\boldsymbol{Y}_{t-1:t}-\mu_{ij}^{(n+1)})(\boldsymbol{Y}_{t-1:t}-\mu_{ij}^{(n+1)})'\psi_t^{(n)}(i,j)}{\sum_{t=2}^{T}\psi_t^{(n)}(i,j)}. \end{aligned}$$

Tous les éléments indispensables afin d'appliquer le modèle PMC à la reconnaissance (segmentation) d'image sont déterminés. Le modèle PMC a été développé dans le but de déchiffrer des états cachés dans une image. Par exemple, il arrive

parfois que la photocopie d'une page de document ne soit pas parfaite et que les informations contenues sur l'endos de la feuille soient visibles sur le recto de celle-ci. Il est donc possible, à l'aide d'un modèle PMC, de décoder la photocopie afin d'en faire ressortir les éléments appartenant au recto et au verso. Les modèles HMM peuvent aussi être utiles pour cette tâche. Dans la sous-section suivante, l'algorithme d'Hilbert-Peano, utilisé dans la décomposition d'image, est introduit.

3.6.1 Hilbert-Peano

Il est possible de lire une image comme la combinaison de nombres (un nombre représente la couleur d'un pixel). En effet, pour une image de 126x126 pixels, il y a donc 15876 données. Pour appliqué le modèle PMC à une image, celle-ci doit-être transformée en vecteur. En effet, l'image en deux dimensions (2D, matrice) doit être transformée en une dimension (1D, vecteur). Cette transformation n'est pas triviale. En effet, un simple aller-retour (*row prime order scan*) des pixels horizontaux peut être exécuté, mais ne serait pas optimal. Pour une image de 4x8, par exemple (voir la figure 3.3), le pixel à la cellule 0 représenterait le premier élément du vecteur $(1, 32)$. Si un simple aller-retour est exécuté, alors le pixel à la cellule 15 serait l'élément $(1, 16)$ du vecteur. Ce n'est pas optimal, car il y a de grandes probabilités que ces 2 pixels appartiennent au même état. Un exemple de balayage aller-retour est illustré à la figure 3.3.

L'algorithme d'Hilbert-Peano (voir Peano, 1890; Griffiths, 1985; Hilbert, 1891) vient corriger ce problème et conserve la propriété de localité 2D. À noter que le balayage (scan) d'Hilbert-Peano fait partie de la famille des courbes de remplissage (space filling curves). La description exhaustive du balayage d'Hilbert-Peano n'est pas faite dans ce mémoire. Cependant, une représentation graphique de ce balayage est illustrée à la figure 3.4 (cette figure provient de Hilbert, 1891) : Le balayage passe donc par plusieurs pixels à proximité avant de se déplacer vers une

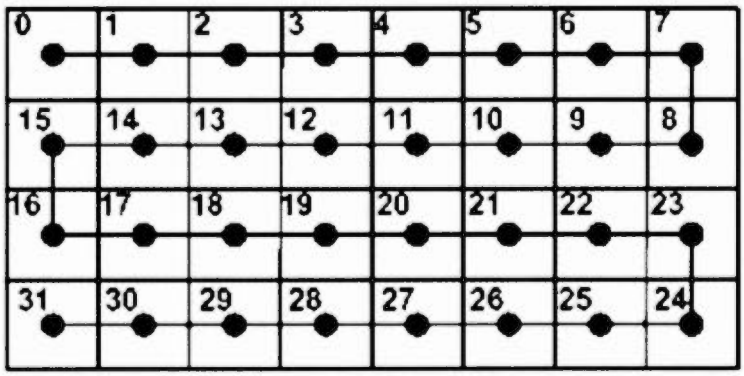


Figure 3.3 Exemple schéma balayage aller-retour. Ce schéma provient de Artyomov *et al.* (2005).

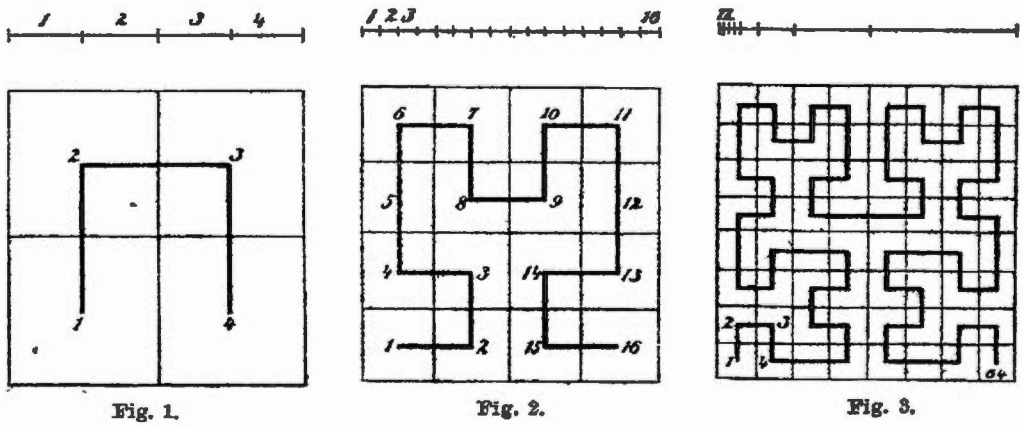


Figure 3.4 Balayage d'Hilbert-Peano

autre région. Ce balayage n'est pas parfait, mais toutefois mieux qu'un simple aller-retour.

3.6.2 Application PMC à la segmentation d'image

Le modèle PMC a été principalement créé pour la reconnaissance d'image. Dans cette section, le modèle PMC est appliqué à la reconnaissance d'images afin de répliquer les papiers de la littérature. Les modèles du présent chapitre seront appliqués sur des données financières au chapitre 4.

Pour appliquer le modèle PMC à la reconnaissance d'image, une image représentant un zèbre est utilisée. La photo originale est illustrée à la figure 3.5.

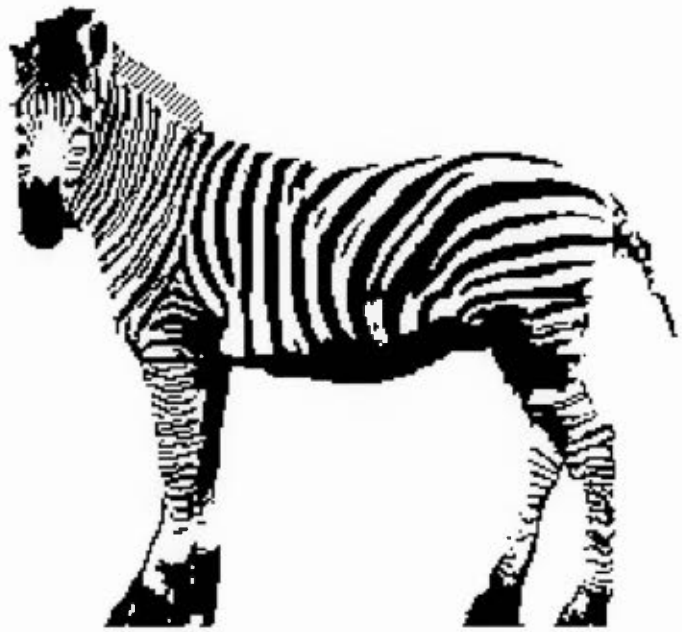


Figure 3.5 Image originale du zèbre

Pour vérifier l'efficacité de la reconnaissance des signaux noirs et blancs avec un modèle PMC, cette image a été volontairement embrouillée (*busy signal*). Autrement dit, elle a été embrouillée à l'aide d'un ordinateur. La figure 3.6 illustre l'image du zèbre embrouillée.

Le zèbre est toujours visible dans la figure 3.6, mais moins que dans la figure

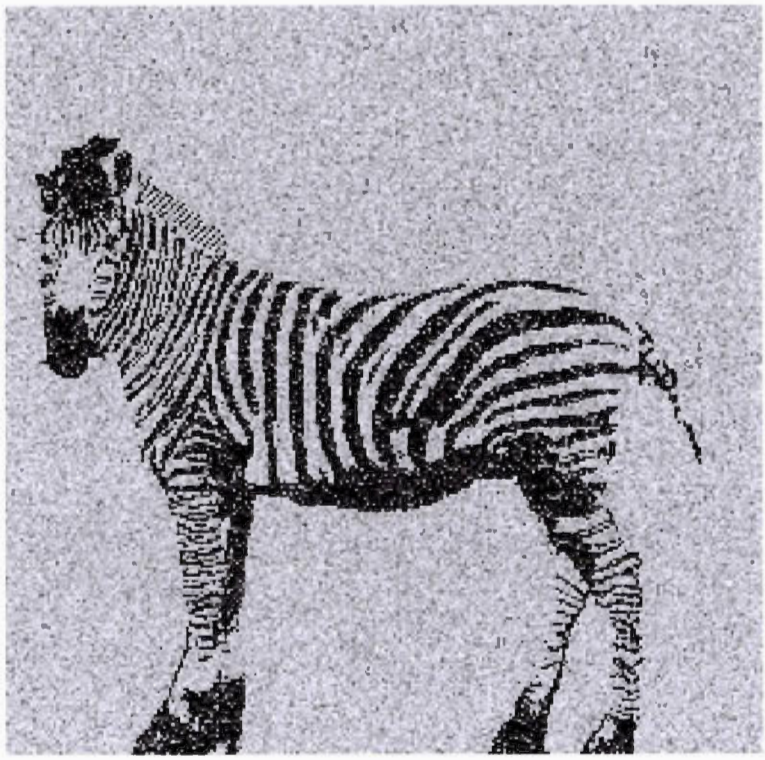


Figure 3.6 Image embrouillée du zèbre

3.5. Le balayage d'Hilbert-Peano est appliqué une première fois afin d'obtenir un vecteur de nombres représentant des couleurs. Par la suite, l'algorithme EM, avec un modèle PMC, est appliqué sur ce vecteur de données. Finalement, le vecteur de résultats est transformé en une matrice qui est transformée à son tour (grâce à *Matlab*) en une image possédant seulement 2 couleurs (soit noir et blanc). Lors de l'application de l'algorithme EM au modèle PMC, le nombre d'états a été délibérément mis à 2 états, car l'image originale ne possède que 2 couleurs (noir/blanc). Le résultat de l'application d'un modèle PMC sur l'image embrouillée est illustré

à la figure 3.7. La distribution de $\boldsymbol{Y}_{t:t+1}|\boldsymbol{C}_{t:t+1}$ du modèle PMC est, par hypothèse, une distribution normale bivariée. Les figures 3.7-3.9 représentent les états cachés et non l'application de l'algorithme de filtrage. En effet, la réplication des valeurs observées importe peu ici. La formule utilisée pour déterminer l'état caché au temps t est $\hat{C}_t = \max(\hat{\boldsymbol{\zeta}}_t)$, où $\hat{\boldsymbol{\zeta}}_t$ représente le vecteur des $\hat{\zeta}_t(i)$ (pour $i = 1, ..., m$). La valeur du maximum ($\hat{\zeta}_t(i)$) n'est pas le résultat final, il s'agit de l'état i. Dans ces figures, l'état 1 représente la couleur noire et l'état 2 la couleur blanche.

Figure 3.7 Image après PMC du zèbre

Le résultat de l'application d'un modèle PMC-IN sur l'image embrouillée est illus-

tré à la figure 3.8. Les distributions $\boldsymbol{Y}_t|\boldsymbol{C}_{t:t+1}$ et $\boldsymbol{Y}_{t+1}|\boldsymbol{C}_{t:t+1}$ du modèle PMC-IN sont des distributions normales.

Figure 3.8 Image après PMC-IN du zèbre

Finalement, afin de comparer le modèle PMC, le résultat de l'application d'un modèle HMM sur l'image embrouillée est illustré à la figure 3.9. Le modèle HMM admet une loi normale comme distribution de $Y|C$.

Pour comparer ces 3 résultats, les pixels, après l'application de chacun des modèles, sont comparés aux pixels de la figure 3.5. Le pourcentage de correspondance des

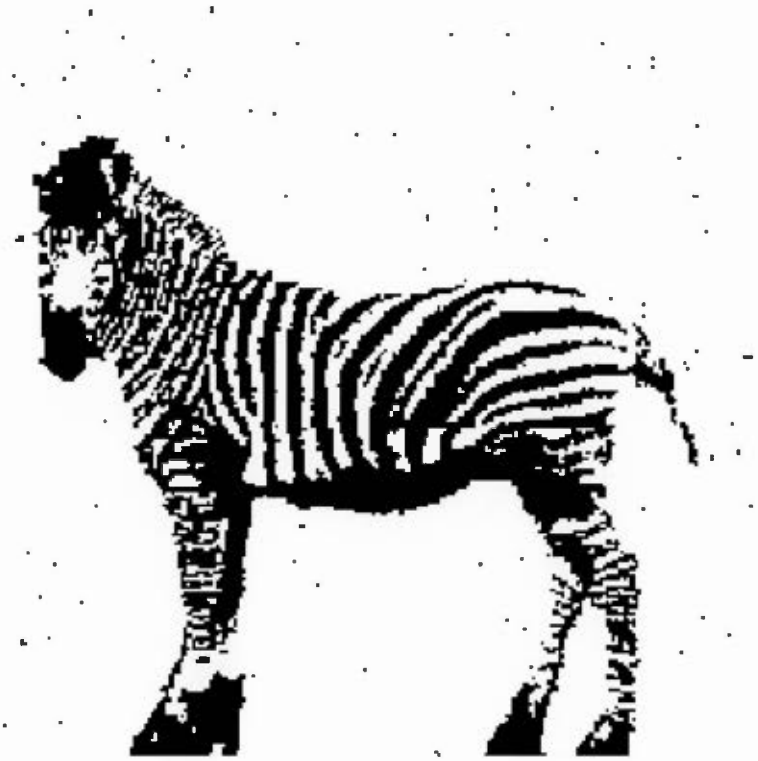


Figure 3.9 Image après HMM du zèbre

pixels est donné dans la table 3.1. Les trois modèles sont sensiblement égaux. Le modèle PMC-IN obtient des résultats légèrement supérieurs aux deux autres. Les résultats sont similaires à ceux publiés dans Lanchantin (2006).

Tableau 3.1

Résultats de l'application des modèles PMC, PMC-IN et HMM comparé à l'image originale du zèbre (voir figure 3.5)

Modèle	Pourcentage de correspondance (%)
PMC	96.10
PMC-IN	97.09
HMM	97.01

CHAPITRE IV

APPLICATION DE MODÈLES HMM ET PMC À DES DONNÉES FINANCIÈRES

Dans ce chapitre, les modèles de chaînes de Markov cachées et de chaînes de Markov couples développés aux chapitres 2 et 3 sont appliqués sur des données financières. Ces modèles sont appliqués à une série temporelle de nombre de défauts. La section 4.1 précise la provenance de ces données. À la section 4.3, les régresseurs utilisés lors des différentes analyses sont déterminés. Aux sections 4.2, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 et 4.10, les modèles élaborés aux chapitres précédents sont appliqués à ces données. Les EMV des paramètres de chaque modèle y sont de plus présentés. Le maximum de la fonction de vraisemblance, le AIC et le BIC de ces mêmes modèles se retrouvent à la section 4.11, et ce, afin de faire la comparaison de l'ajustement des modèles.

4.1 Base de données

La base de données utilisée dans ce mémoire provient du Département de droit de l'Université de Californie à Los Angeles (UCLA). Le nom de cette base de données est UCLA-LoPucki Bankruptcy Research Database (BRD) (UCLA, 2017). Celle-ci a été conçue dans le but de réunir toute l'information disponible sur les faillites (évènement de défaut) de grandes entreprises publiques américaines. La BRD contient plus de 1000 évènements de défauts recensés du 1er octobre 1979 à ce jour.

Pour les fins du présent mémoire, les données comptabilisés jusqu'en novembre 2017 sont utilisées.

Une compagnie doit remplir un rapport annuel *10-K* du SEC (*Securities and Exchange Commission*) pour être considérée dans cette base de données (voir SEC, 2017a). Le *10-K* est un rapport financier que les grandes compagnies publiques américaines doivent remplir annuellement. Une compagnie publique doit remplir, selon le *US Bankcruptcy Code*, le chapitre 7 ou 11 du rapport *10-K* afin de signaler tout évènement de défaut à ses actionnaires. Le chapitre 7 s'applique en cas de faillite, ce qui implique que le débiteur (dans ce mémoire, le débiteur est une compagnie publique) n'a plus d'obligation de paiement envers ses créanciers (dans ce mémoire, les créanciers sont les actionnaires). Le chapitre 11 concerne les cas de restructuration de la dette et la poursuite future des opérations du débiteur. Bien que la dette soit simplement restructurée dans le cas du formulaire de chapitre 11, le créancier souffrira souvent de cette restructuration. Pour obtenir des informations plus détaillées au sujet du chapitre 7 et 11, voir SEC (2017b).

Pour faire partie de la BRD, la compagnie doit, de plus, avoir une valeur d'actifs de plus de 100 millions en dollars de 1980. L'indice des prix à la consommation (IPC) est utilisé afin de convertir la valeur d'actifs en dollars de 1980. Le tableau C.1, en annexe, présente la valeur annuelle minimale des actifs requis pour faire partie de cette base de données.

La base BRD comprend un total de plus de 200 colonnes d'informations qui répertorient tous les enregistrements de compagnie ayant rempli le chapitre 7 ou 11 du rapport 10-K. Dans ce mémoire, seulement la colonne *DateFiled*, qui représente la date de l'évènement de défaut, est utilisée. Dans la base BRD, il y a 24 compagnies ayant rempli le chapitre 7 et 1071 pour le chapitre 11. Pour ce mémoire, les chapitres 7 et 11 sont considérés comme un évènement de défaut

quelconque. De plus, les évènements de défaut d'un même mois ont été regroupés afin d'obtenir une série temporelle mensuelle du nombre d'évènements de défauts (risque de crédit).

4.1.1 Régresseurs

Les régresseurs utilisés sont reliés à des séries temporelles financières. Ces séries sont disponibles dans les bases de données en ligne sur le site de la FRED (Federal Reserve Bank of St. Louis Economic Data - voir FRED, 2017a). Les bases de données canadiennes n'ont pas été utilisées, puisque la BRD comptabilise seulement les grandes compagnies publiques américaines.

Les indicateurs économiques sont les séries temporelles les plus utilisées pour modéliser des séries temporelles quelconques se rattachant au domaine financier. Ce style de série donne en général une bonne vue d'ensemble de l'économie à un moment voulu. Pour ce mémoire, les séries temporelles les plus classiques ont été prises, soit celles énoncées dans Lahiri et Moore (1992), Griliches (1990), Sullivan (2003), Quandl (2017), FRED (2017b) etc. De plus, les séries sélectionnées sont faciles à interpréter d'un point de vue économique. Les indicateurs utilisés dans ce mémoire sont présentés ci-dessous. Le code à utiliser pour les retrouver facilement sur le site de FRED, ainsi qu'une brève description de chacun d'entre eux, est fourni.

1. **Produit intérieur brut réel (PIB réel) :** (FRED : GNPC96) Le PIB réel représente la production économique qu'un pays peut produire, et ce, pour tous produits et services confondus. Le PIB réel, en comparaison au PIB, est un indice en dollar constant. Une augmentation du PIB est, en général, accompagnée d'une amélioration de l'économie, donc d'une diminution du taux de défauts.

Le PIB réel n'est pas disponible en valeur mensuelle, car celui-ci est seulement enregistré trimestriellement. Ce régresseur ne sera donc pas retenu. Toutefois, il a été démontré par Okun (1963) qu'il y a une forte corrélation entre le PIB réel et la différence du taux de chômage. Afin de vérifier cette hypothèse, le calcul de la corrélation entre les régresseurs trimestriels a été effectué sur la période à l'étude. Il est possible d'observer une corrélation de -65% entre le PIB réel et la différence du taux de chômage. Le chômage est donc utilisé comme régresseur mandataire du PIB réel.

2. **Taux de chômage :** (FRED : UNRATE_CHG) Le taux de chômage représente le pourcentage de la main-d'œuvre active qui n'a pas d'emploi. Une augmentation du taux de chômage est, en général, accompagnée d'une détérioration de l'économie, donc d'une augmentation du taux de défauts. La différence mensuelle du taux est utilisée.

3. **Indice S&P 500 :** (FRED : SP500) Le S&P 500 est un indice boursier qui représente l'état financier des 500 plus grandes compagnies américaines cotées en bourse. Une augmentation de l'indice S&P 500 est, en général, accompagné d'une amélioration de l'économie, donc d'une diminution du taux de défauts. Les log-rendements sont utilisés, soit $\log\left(\frac{S_{t_1}}{S_{t_2}}\right)$, où S_{t_1} représente la valeur de l'indice au temps t_1 et S_{t_2} représente la valeur de l'indice au temps t_2

4. **Écart de rendement entre les obligations corporatives (BAA) et celle du gouvernement fédéral (*Spread*) :** (FRED : BAA10Y) Le *Spread* représente l'écart du rendement des obligations qu'une compagnie cotée Baa offrira en comparaison au taux obtenu avec des obligations du Trésor. Les obligations du Trésor sont des obligations émises par le gouvernement et considérées sans risque. Cet indice économique représente l'incertitude du marché face aux investissements considérés risqués. Une augmentation du

Spread est, en général, accompagnée d'une détérioration de l'économie, donc d'une augmentation du taux de défauts.

5. **Indice des prix à la consommation (IPC) :** (FRED : CPIAUCSL_PCH) L'IPC représente l'évolution moyenne du prix des biens et services offerts. Une augmentation de l'IPC est, en général, accompagnée d'une amélioration de l'économie. Cependant, cet effet peut n'être que temporaire. Le pourcentage de différence mensuelle est utilisé.
6. **Taux directeur (*Fed fund rate*) :** (FRED : FEDFUNDS) Le taux directeur est le taux d'intérêt que les institutions financières doivent payer lorsqu'elles empruntent d'une autre institution financière. Une augmentation du taux directeur est, en général, accompagnée d'une détérioration de l'économie, donc d'une augmentation du taux de défauts.
7. **VIX :** (FRED : VIXCLS) Le VIX est un indice boursier qui représente la volatilité à court terme espérée sur les marchés. Puisque cet indice est seulement disponible depuis le début des années 1990, ce régresseur ne sera pas utilisé. La corrélation du VIX avec le log-rendement du S&P 500 est de plus de 60%, donc le S&P 500 est utilisé comme régresseur mandataire du VIX.

La matrice de corrélation est calculée pour les régresseurs mensuels, soit :

	Chômage	*Spread*	FEDFUNDS	IPC	S&P 500
Chômage	1	0.286	0.096	0.002	−0.015
Spread	0.286	1	−0.388	−0.311	−0.058
FEDFUNDS	0.096	−0.388	1	0.497	−0.006
IPC	0.002	−0.311	0.497	1	−0.031
S&P 500	−0.015	−0.058	−0.006	−0.031	1

Avec cette matrice, il est possible de voir que les régresseurs ne sont pas fortement corrélés entre eux.

4.1.2 Nombre de compagnies (exposition)

Le nombre de compagnies actives (exposition) vient affecter le nombre de défauts. En effet, il est normal de remarquer un plus grand nombre de défauts si le nombre de compagnies actives est plus élevé. Pour obtenir le nombre annuel de compagnies actives, les bases de données de Compustat (2017) ont été utilisées. En guise de rappel, seules les compagnies américaines ayant un actif de plus de 100,000,000$ en dollar de 1980 ont été comptabilisées. Afin de déterminer la valeur minimale d'actifs qu'elles devaient posséder annuellement, le tableau C.1 a été utilisé. Puisque le nombre de compagnies est présenté sur une base annuelle, la valeur obtenue par Compustat est considérée comme le nombre de compagnies à la mi-année. Le nombre de compagnies aux autres dates est obtenu à l'aide d'une spline polynomiale cubique (voir Judd, 1998). Les valeurs obtenues de Compustat se trouvent à l'annexe D.

La figure 4.1 illustre le nombre de compagnies annuelles répondant aux critères énoncés. La diminution du nombre de compagnies depuis la fin des années 1990 est causée par une augmentation des fusions-acquisitions et par la crainte des nouvelles compagnies à devenir publique (voir Schumpeter, 2017).

Pour prendre en considération l'exposition au temps t, le paramètre λ_t de la loi de Poisson est modifié, soit :

$$\lambda_t = \exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \log(E_t)), \tag{4.1}$$

où E_t est l'exposition au temps t. L'exposition est donc considérée comme un régresseur avec un paramètre de valeur un.

Puisque le nombre annuel de compagnies est élevé, le nombre de compagnies observées en 1979 sera considéré comme ayant une valeur de un et les autres

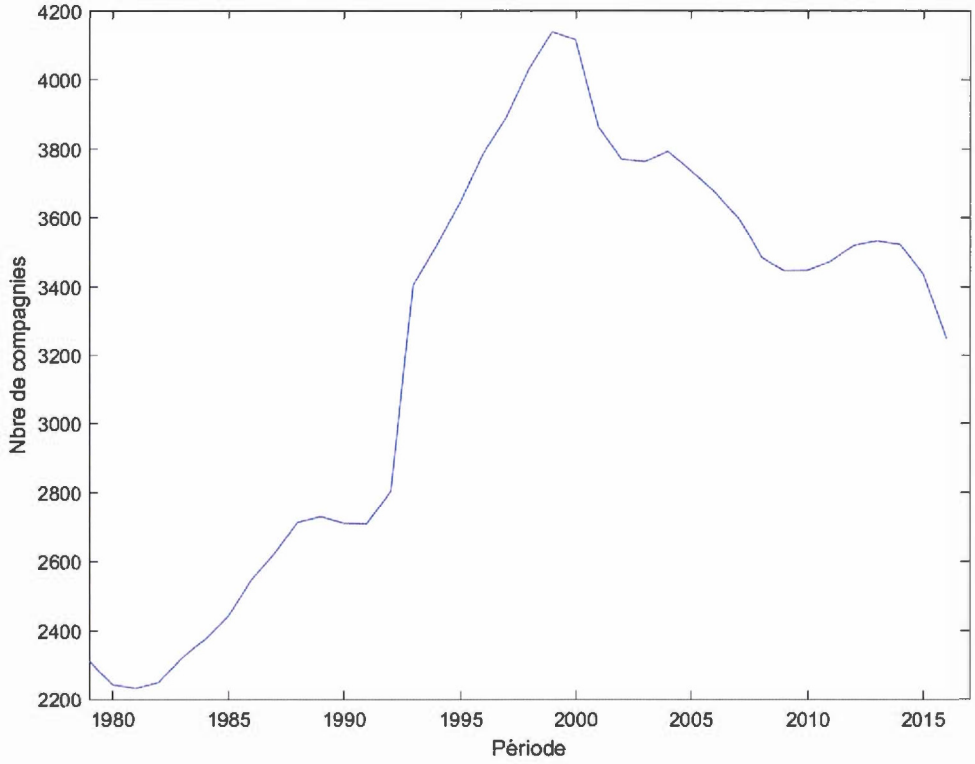


Figure 4.1 Évolutions temporelles du nombre de compagnies actives avec un actif $>$ 100,000,000$ en valeur de 1980.

valeurs seront un ratio de cette valeur, soit :

$$E_t = \frac{N_i}{N_{1979}},$$

où N_i est le nombre de compagnies observé à l'année i.

4.1.3 Données empiriques

La figure 4.2 représente l'évolution de la série temporelle de défauts qui est à l'étude dans ce mémoire :

Les statistiques descriptives de cette série temporelle sont fournies à titre indicatif :

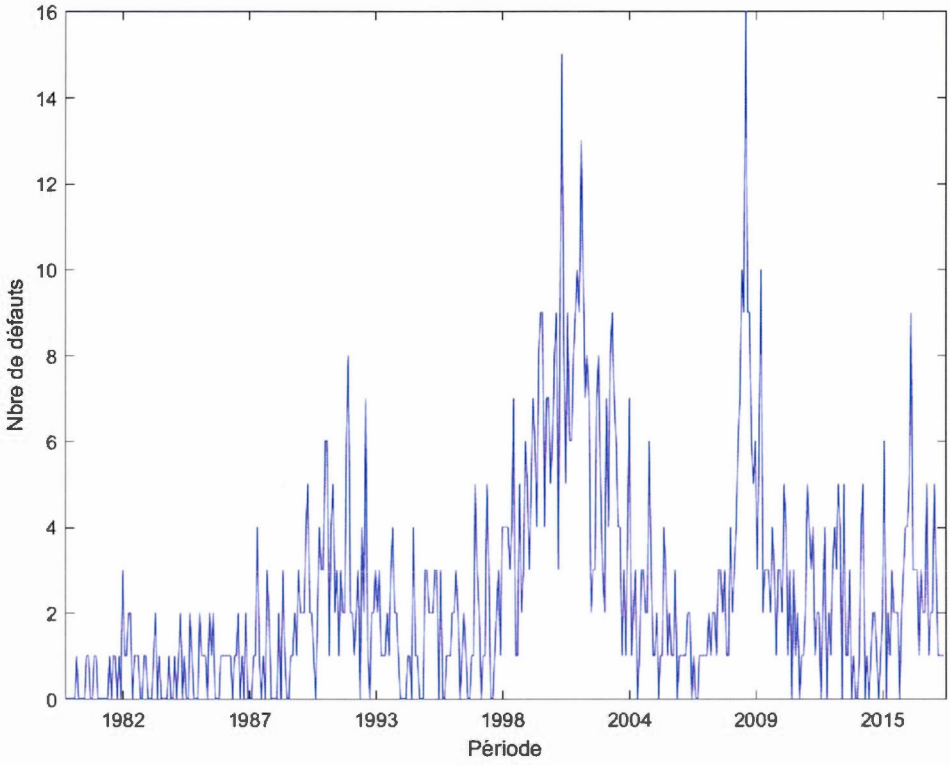


Figure 4.2 Évolutions temporelles des défauts

— $\bar{y} = 2.4026$

— $\hat{s}_Y = 2.5443$: Il semble y avoir une certaine surdispersion.

— Coefficient d'asymétrie = 1.7246 : Il est normal de voir une valeur positive pour le coefficient d'asymétrie(*skewness*), puisque les observations sont bornées à 0.

La distribution empirique est illustrée à la figure 4.3.

La distribution empirique a aussi été comparée à une distribution de Poisson de moyenne $\frac{\sum_{t=1}^{T} y_t}{T} = 2.4026$. La distribution empirique n'est pas bien ajustée par la distribution de Poisson. En effet, la distribution empirique a une très longue queue (grande distance de certaines données en comparaison à la moyenne) et présente

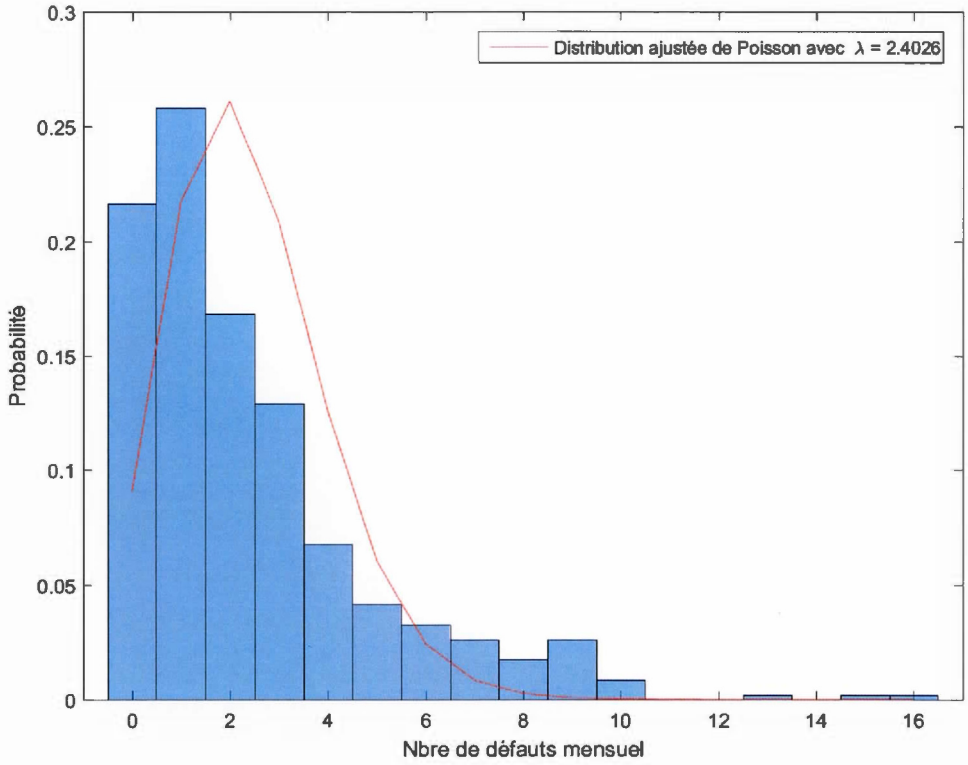


Figure 4.3 Distribution empirique des défauts

une plus grande probabilité d'observer zéro ou un défaut, en comparaison à une distribution de Poisson. De plus, un test du χ^2 d'adéquation (voir Pearson, 1900) sur la loi de Poisson rejette ce modèle à un seuil de 1%. La p-value de ce test est inférieure à 10^{-50}.

4.2 Application - Poisson

La figure 4.4 représente l'application d'une distribution de Poisson à la série de défauts. La figure 4.5 représente le même modèle, mais cette fois-ci en considérant l'exposition. Pour obtenir la valeur du paramètre λ_t, l'équation (1.6) a été utilisée. Pour trouver le paramètre du modèle de Poisson considérant l'exposition, l'équation (4.1) a été utilisée. Cependant, il n'y a pas de régresseur (celui-ci est

considéré constant), soit :

$$\begin{aligned} \lambda_t &= \exp(\boldsymbol{x}_t'\boldsymbol{\beta} + \log(E_t)) \\ &= \exp(\beta_0 + \log(E_t)). \end{aligned} \tag{4.2}$$

Pour le reste des applications de ce chapitre, l'exposition est toujours considérée. De plus, le paramètre β_0 est seulement utilisé dans ce modèle.

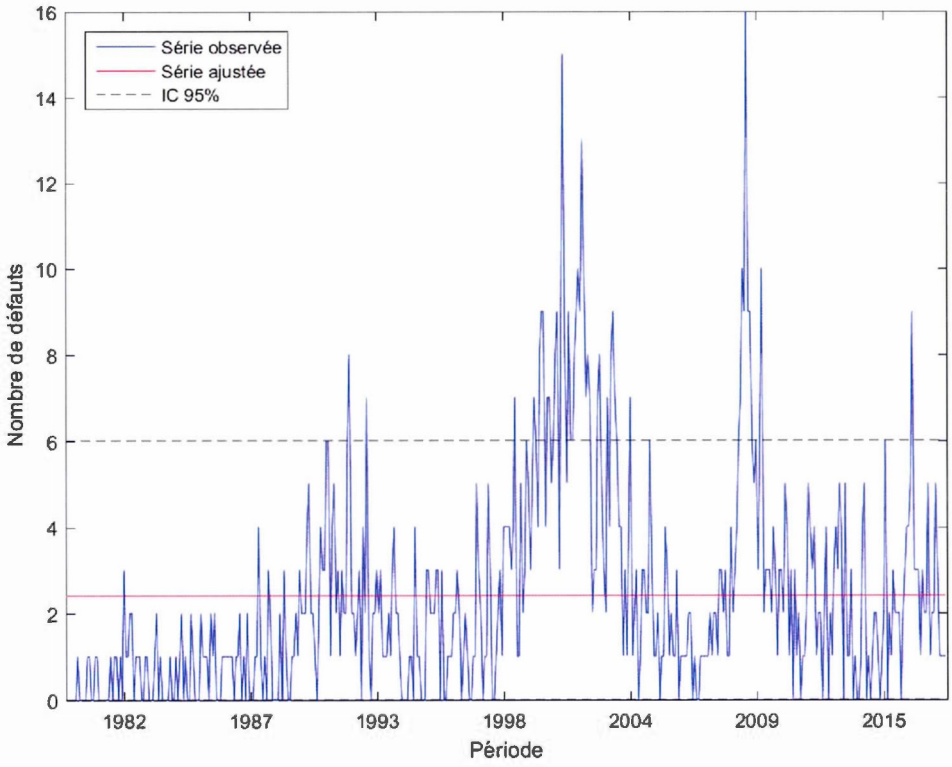


Figure 4.4 Application - Poisson

Dans les figures 4.4 et4.5, la ligne rouge représente simplement l'espérance au temps t, soit λ_t. Dans la première figure, cette ligne est constante, car l'exposition n'est pas prise en compte. La ligne pointillée représente un intervalle de confiance

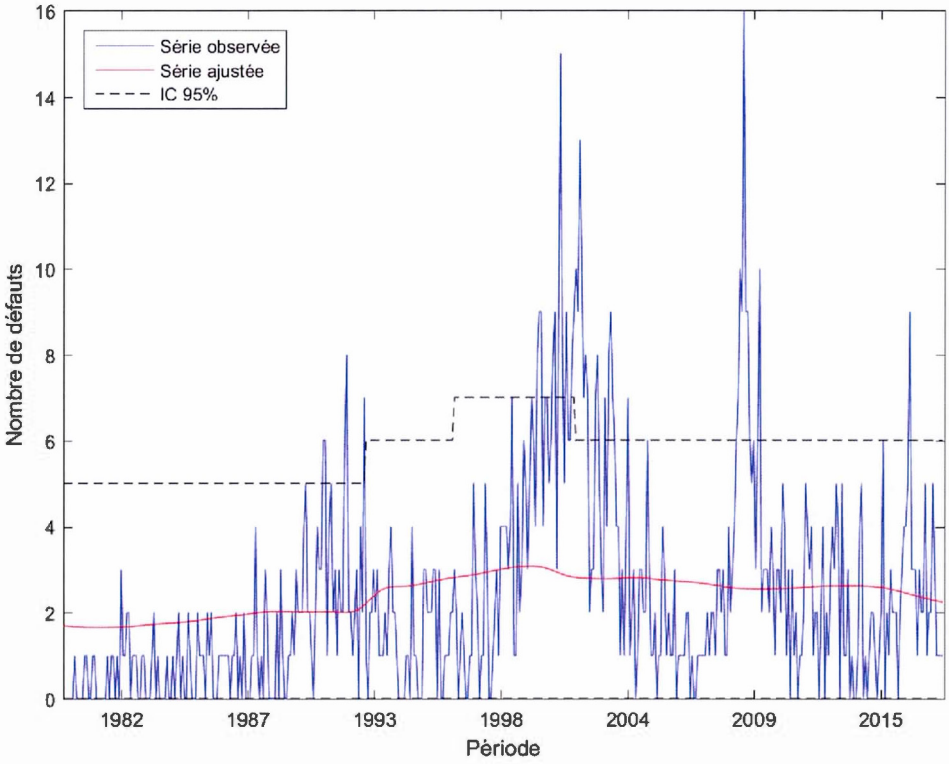


Figure 4.5 Application - Poisson avec exposition

(IC) à 95%, soit la valeur à 2.5% et 97.5% du quantile d'une distribution de Poisson. L'équation (1.1) est utilisée pour calculer la fonction cumulative, soit :

$$\mathbb{P}(Y_t \leq y) = \sum_{k=0}^{y} k \frac{e^{-\lambda_t} \lambda_t^k}{k!}. \tag{4.3}$$

L'intervalle de confiance au temps t sera affiché dans toutes les figures suivantes d'application de modèle. Il est à noter que la borne inférieur de l'IC sera en générale de zéro, donc celle-ci ne sera pas visible.

Dans les deux graphiques ci-dessus, il est possible d'observer que la distribution de Poisson (avec et sans exposition) n'est pas la meilleure modélisation possible. En effet, la ligne représentant la moyenne espérée à chaque période t se trouve parfois très loin de la valeur observée. La période de 2007 à 2010 représente la

dernière crise financière. Si la moyenne selon une loi de Poisson est comparée avec les valeurs empiriques de cette période, il est possible de constater que la Poisson sous-estime cette période. Même aux valeurs limites d'un IC à 95%, ce modèle se trouve parfois très loin de la valeur empirique. La plus grande faiblesse de ce modèle vient du fait que la loi est stationnaire.

La valeur de l'EMV de ces modèles est de $\lambda = 2.4026$ et $\beta_0 = 0.5371$.

4.3 Sélection du modèle

Puisqu'il est parfois avantageux d'ajouter ou d'enlever des régresseurs, 31 modèles possibles ont été créés, soit $2^5 - 1$ possibilités. Le -1 de $2^5 - 1$ représente la possibilité où il n'y a aucun régresseur, soit une simple distribution de Poisson (voir section 4.2). L'utilisation conjointe de quelconque régresseur est possible, car ceux-ci ne sont pas fortement corrélés entre eux (voir sous-section 4.1.1). Le nombre d'états du modèle HMM a, de plus, été testé avec des valeurs entre 2 et 4. Pour ce mémoire, le modèle 13 a été utilisé (voir l'annexe E) avec 2 états. Ce modèle s'avérait celui qui, mathématiquement, s'ajustait mieux à la série de données. Le modèle utilise donc les régresseurs *Spread*, *Fed fund* et S&P 500.

La figure 4.6 illustre les séries temporelles représentant le taux *Spread*, le taux *Fed fund* et les log-rendements du S&P 500.

Dans les modèles qui suivent, si des régresseurs sont présentés, le régresseur β_i sera donc β_1=*Spread*, β_2=*Fed fund* et β_3=S&P 500. L'explication économique de ces régresseurs se trouve à la sous-section 4.1.1.

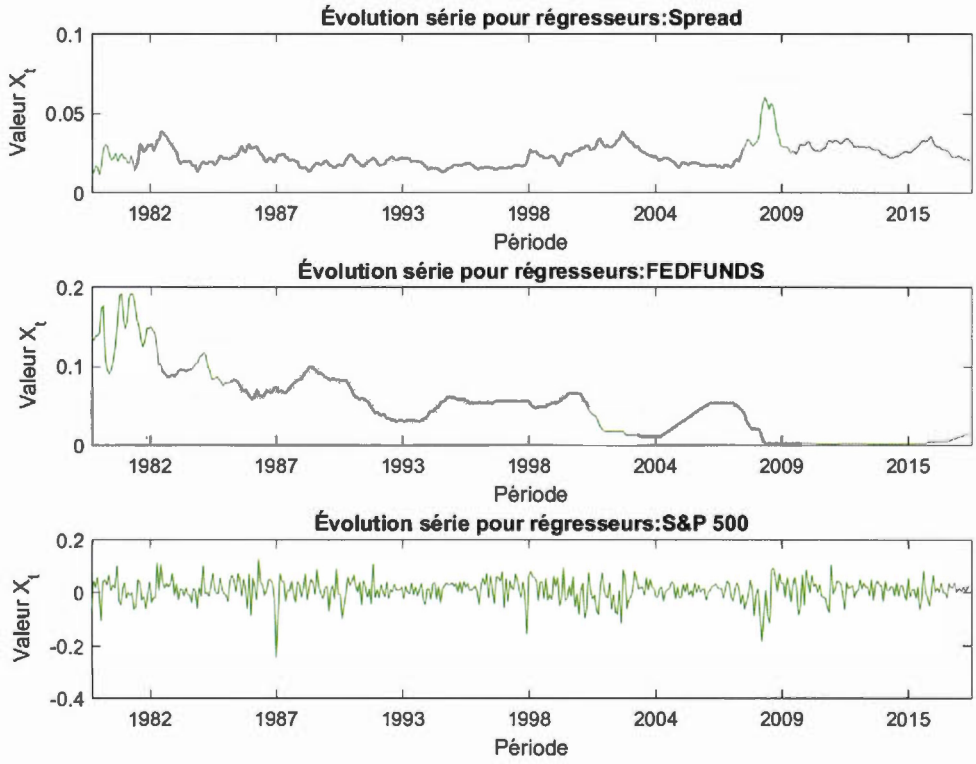


Figure 4.6 Série temporelle des régresseurs du modèle 13

4.4 Application - Poisson avec régresseurs

La figure 4.7 représente l'application d'une distribution de Poisson avec régresseurs à la série de défauts. La technique décrite à la sous-section 1.1.4 a été utilisée pour trouver la valeur des EMV. Les régresseurs utilisés sont ceux du modèle #13.

La distribution de Poisson avec régresseurs semble mieux s'ajuster aux données empiriques en comparaison à la distribution de Poisson sans régresseur. La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (4.2). L'IC est calculé, à chaque période de temps, selon l'équation (4.3). La valeur de λ_t dans l'IC se calcule à partir de l'équation (4.2). Les valeurs maximales

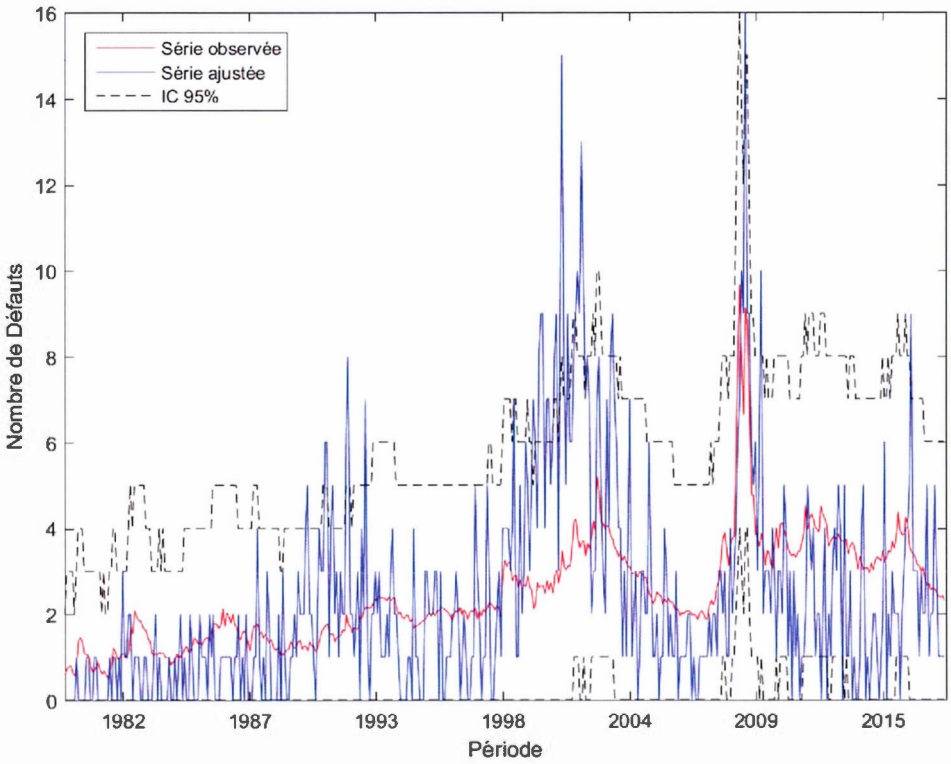


Figure 4.7 Application - Poisson avec régresseurs

de l'IC semblent capter une plus grande portion des observations. Cependant, durant les périodes de grande volatilité, ce modèle ne semble pas très bien capter toute la variabilité des observations. En effet, pour la période de 1998 à 2004 la distribution de Poisson avec régresseurs sous-estime la valeur des y_t.

La valeur des EMV de ce modèle est fournie au tableau 4.1.

Tableau 4.1

EMV β_i - Poisson régresseur

β_i	Valeur β_i
β_1	31.0618
β_2	-5.5884
β_3	0.9904

4.5 Application - HMM Poisson

La figure 4.8 représente l'application d'un modèle de chaîne de Markov cachée simple à la série de défauts. Les distributions de Y conditionnelles à C $(Y|C)$ utilisées sont toutes des distributions de Poisson (sans régresseur). L'algorithme EM a été utilisé pour trouver la valeur des EMV (voir section 2.5).

La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (2.31). L'IC est calculé, à chaque période de temps, selon l'équation (2.30). La valeur de $\lambda_{t,i}$ se calcule à partir de l'équation suivante :

$$\lambda_{t,i} = \exp(\log(E_t))F_i = \lambda_t F_i. \tag{4.4}$$

En comparaison avec le modèle de Poisson avec et sans régresseur, les périodes plus volatiles sont mieux modélisées et l'IC capte presque tout l'éventail des observations. La valeur à 97.5% (IC) capte plusieurs valeurs extrêmes.

La figure 4.9 illustre la probabilité de se situer dans chacun des états à chaque période de temps t, soit le paramètre $\hat{\zeta}_j(t)$ (voir équation (2.22)). L'état 1 semble couvrir des périodes de peu de défauts et l'état 2, les périodes avec le plus de défauts.

La valeur des EMV de ce modèle est fournie au tableau 4.2.

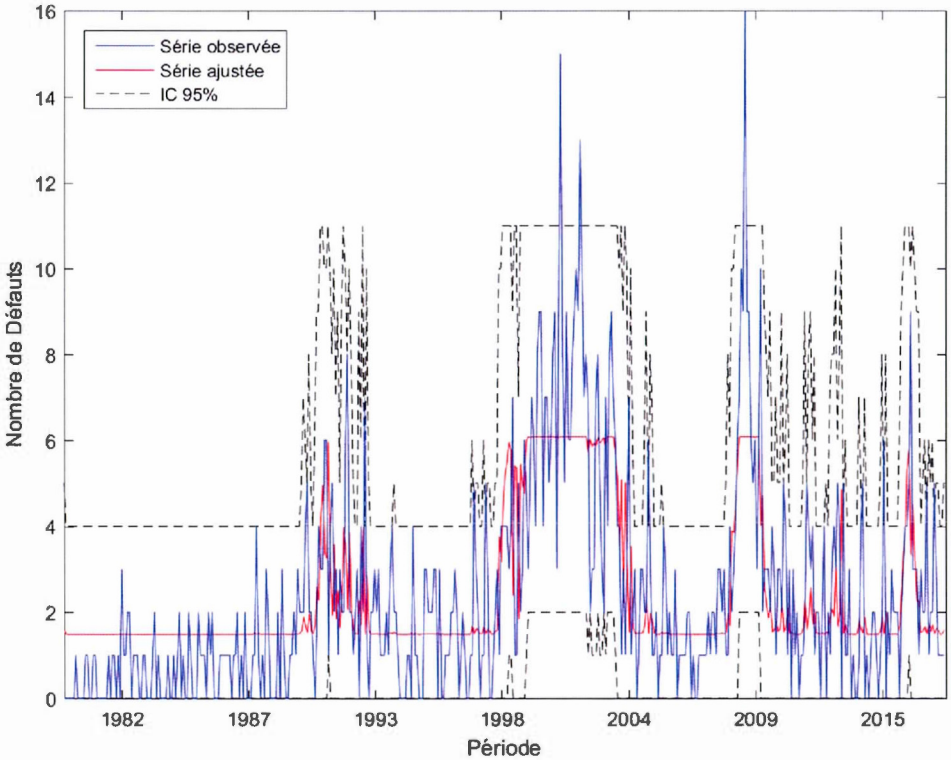


Figure 4.8 Application - HMM Poisson

Tableau 4.2

EMV $\mathbf{\Gamma}$, δ_i et F_i - HMM Poisson

$$\mathbf{\Gamma} = \begin{bmatrix} 0.9760 & 0.0240 \\ 0.0986 & 0.9014 \end{bmatrix}$$

δ_i	Valeur δ_i
δ_1	0.8046
δ_2	0.1954

F_i	Valeur F_i
F_1	0.3848
F_2	1.8083

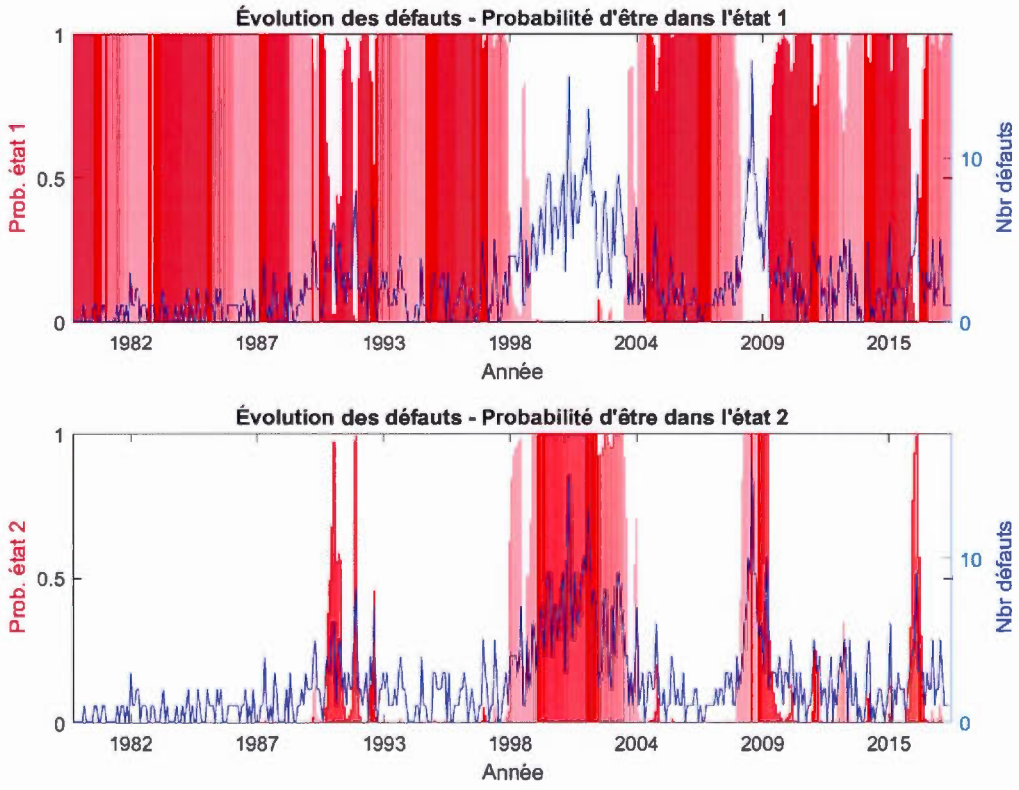


Figure 4.9 Probabilité $\hat{\zeta}_j(t)$ - HMM Poisson

4.6 Application - HMM Poisson avec régresseurs

La figure 4.10 représente l'application d'un modèle de chaîne de Markov cachée à la série de défauts. Les distributions de Y conditionnelles à C $(Y|C)$ utilisées sont des distributions de Poisson avec régresseurs. L'algorithme EM a été utilisé pour trouver la valeur des EMV (voir la section 2.5).

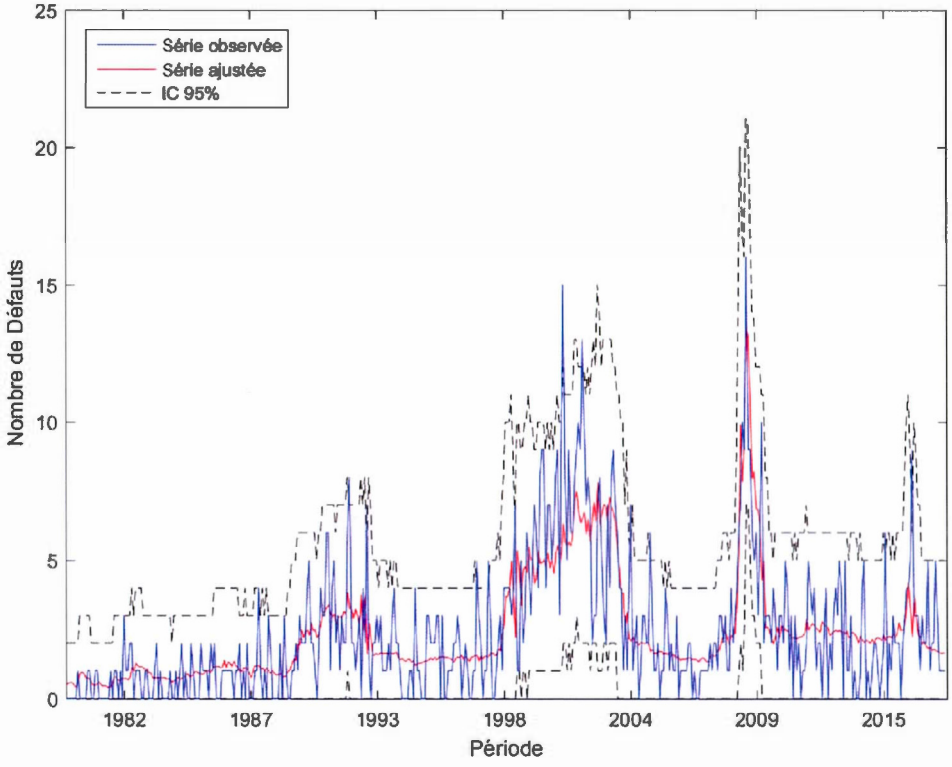


Figure 4.10 Application - HMM Poisson avec régresseurs

La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (2.31). L'IC est calculé, à chaque période de temps, selon l'équation (2.30). La valeur de $\lambda_{t,i}$ se calcule à partir de l'équation suivante :

$$\lambda_{t,i} = \exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \log(E_t))F_i = \lambda_t F_i. \tag{4.5}$$

En comparaison avec le modèle de Poisson avec régresseurs, les périodes plus volatiles sont mieux modélisées et la variance vient capter presque tout l'éventail des observations. Le modèle HMM Poisson avec régresseurs semble mieux ajuster les périodes de variabilité en comparaison au modèle sans régresseurs. De plus, l'IC est beaucoup plus concentré autour des observations lorsqu'il y a présence de régresseurs. En effet, pour la période de 1998 à 2002, le modèle HMM avec régresseur semble suivre la tendance à la hausse sans pour autant faire comme le modèle sans régresseur qui admet simplement un très grand IC.

La figure 4.11 illustre la probabilité de se situer dans chacun des états à chaque période de temps t, soit le paramètre $\hat{\zeta}_j(t)$ (voir equation (2.22)). Le premier état semble représenter les périodes peu à risque. L'état 2 semble représenter les périodes avec les plus grandes valeurs et variabilités.

La valeur des EMV de ce modèle est fournie au tableau 4.3.

Tableau 4.3

EMV $\mathbf{\Gamma}$, δ_i, F_i et β_i - HMM Poisson avec régresseurs

$$\mathbf{\Gamma} = \begin{bmatrix} 0.9882 & 0.0118 \\ 0.0425 & 0.9575 \end{bmatrix}$$

δ_i	Valeur δ_i
δ_1	0.7824
δ_2	0.2176

F_i	Valeur F_i
F_1	-0.2277
F_2	0.8081

β_i	Valeur β_i
β_1	23.2473
β_2	-5.4745
β_3	1.1960

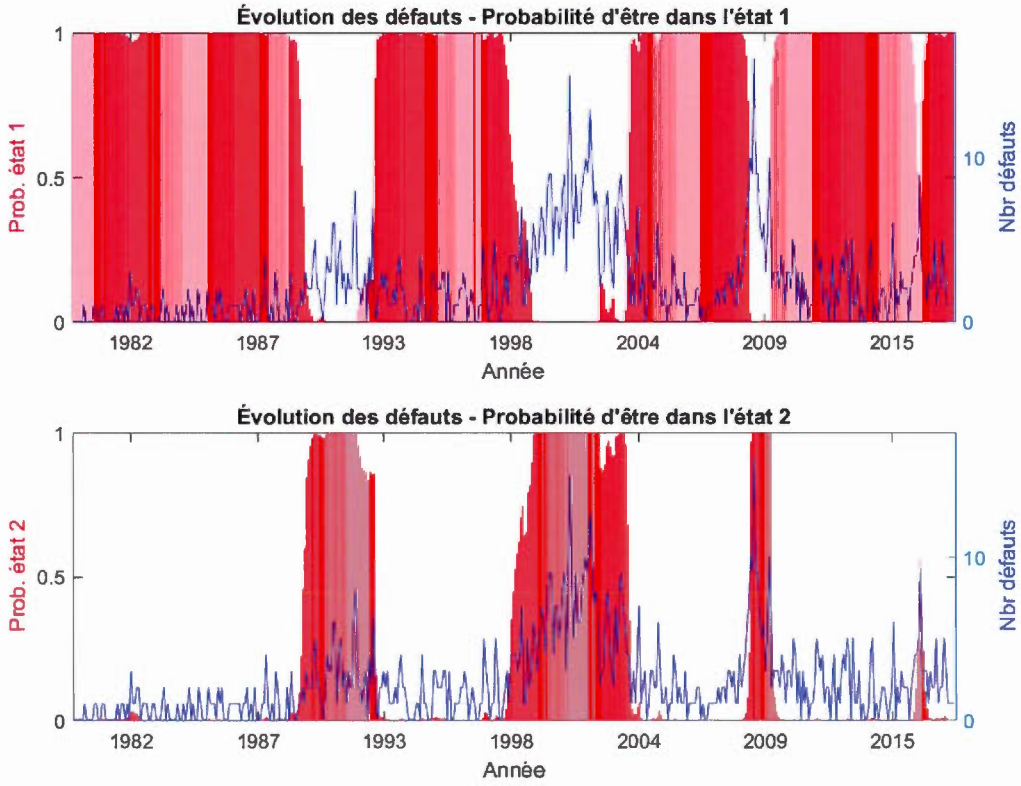


Figure 4.11 Probabilité $\hat{\zeta}_j(t)$ - HMM Poisson avec régresseurs

4.7 Application - HMM second ordre Poisson avec régresseurs $(Y|C_t)$

La figure 4.12 représente l'application d'un modèle de chaîne de Markov cachée du second ordre à la série de défauts. Les distributions de Y conditionnelles à C $(Y|C)$ utilisées sont des distributions de Poisson avec régresseurs. L'algorithme EM a été utilisé pour trouver la valeur des EMV (voir la section 2.5) avec une légère modification sur la matrice $\mathbf{\Gamma}$ (voir la section 2.6).

La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (2.31). L'IC est calculé, à chaque période de temps, selon l'équation (2.30). La valeur de $\lambda_{t,i}$ se calcule à partir de l'équation (4.5).

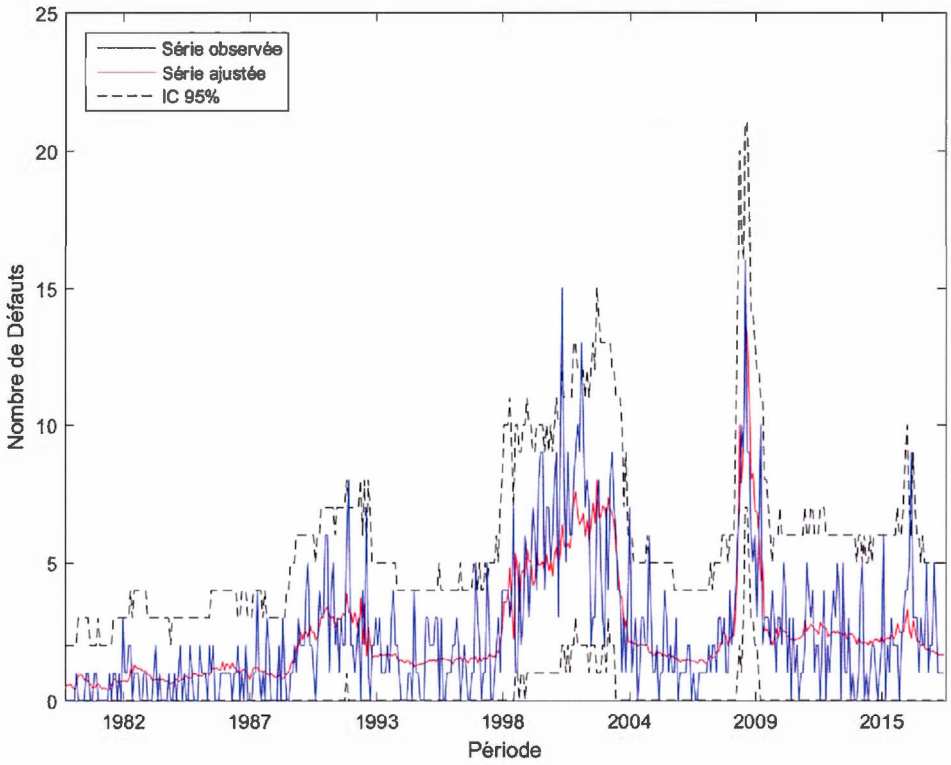


Figure 4.12 Application - HMM second ordre Poisson avec régresseurs $(Y_t|C_t)$

La figure 4.13 illustre la probabilité de se situer dans chacun des états à chaque période de temps t, soit le paramètre $\hat{\zeta}_j(t)$ (voir l'équation (2.22)). Cette figure comporte 4 états pour représenter les 4 mélanges d'états $C_{t-1} = i$ et $C_t = j$ (pour $i, j = 1, 2$) possibles. Le premier état semble représenter les périodes peu à risque. L'état 4 semble représenter les périodes avec les plus grandes valeurs et variabilités. Les états 2 et 3 ne semblent pas être utilisés.

La valeur des EMV de ce modèle est fournie au tableau 4.4.

Bien que la figure 4.12 semble montrer que ce modèle ajuste bien les observations,

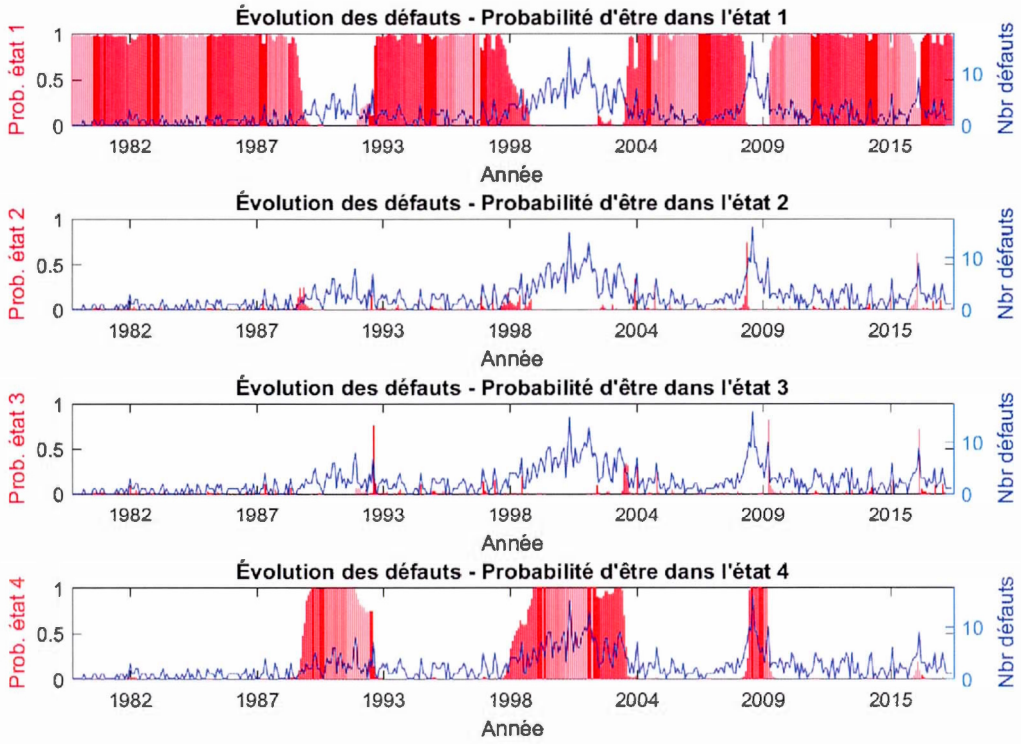


Figure 4.13 Probabilité $\hat{\zeta}_j(t)$ - HMM second ordre Poisson avec régresseurs $(Y_t|C_t)$

l'utilisation de 4 états n'est peut-être pas optimale puisque certaines probabilités γ_{ij} sont proches ou égales à 100%.

Tableau 4.4

EMV $\mathbf{\Gamma}$, δ_i, F_i et β_i - HMM second ordre Poisson avec régresseurs $(Y_t|C_t)$

$$\mathbf{\Gamma} = \begin{bmatrix} 0.9757 & 0.0243 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.5817 & 0.4183 \\ 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0381 & 0.9619 \end{bmatrix}$$

δ_i	Valeur δ_i
δ_1	0.7601
δ_2	0.0185
δ_3	0.0185
δ_4	0.2030

F_i	Valeur F_i
F_1	-0.2456
F_2	0.8073

β_i	Valeur β_i
β_1	23.4084
β_2	-5.4938
β_3	1.2431

4.8 Application - HMM second ordre Poisson avec régresseurs $(Y|C_t, C_{t-1})$

La figure 4.14 représente l'application d'un modèle de chaîne de Markov cachée du second ordre à la série de défauts. Les distributions de Y_t conditionnelles à C_t et C_{t-1} $(Y_t|C_t, C_{t-1})$ utilisées sont des distributions de Poisson avec régresseurs. L'algorithme EM a été utilisé pour trouver la valeur des EMV (voir la section 2.5) avec une légère modification de la matrice $\mathbf{\Gamma}$ (voir la section 2.6) ainsi qu'une modification sur les probabilités $\boldsymbol{P}(y_t)$ (voir la matrice 2.32).

La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (2.31). L'IC est calculé à chaque période de temps selon l'équation (2.30). La valeur de $\lambda_{t,i}$ se calcule à partir de l'équation (4.5).

La figure 4.15 illustre la probabilité de se situer dans chacun des états à chaque période de temps t, soit le paramètre $\hat{\zeta}_j(t)$ (voir l'équation (2.22)). Cette figure comporte 4 états pour représenter les 4 mélanges d'états $C_{t-1} = i$ et $C_t = j$ (pour $i, j = 1, 2$) possibles. Ce modèle ne semble pas bien ajuster les réalisations, puisque

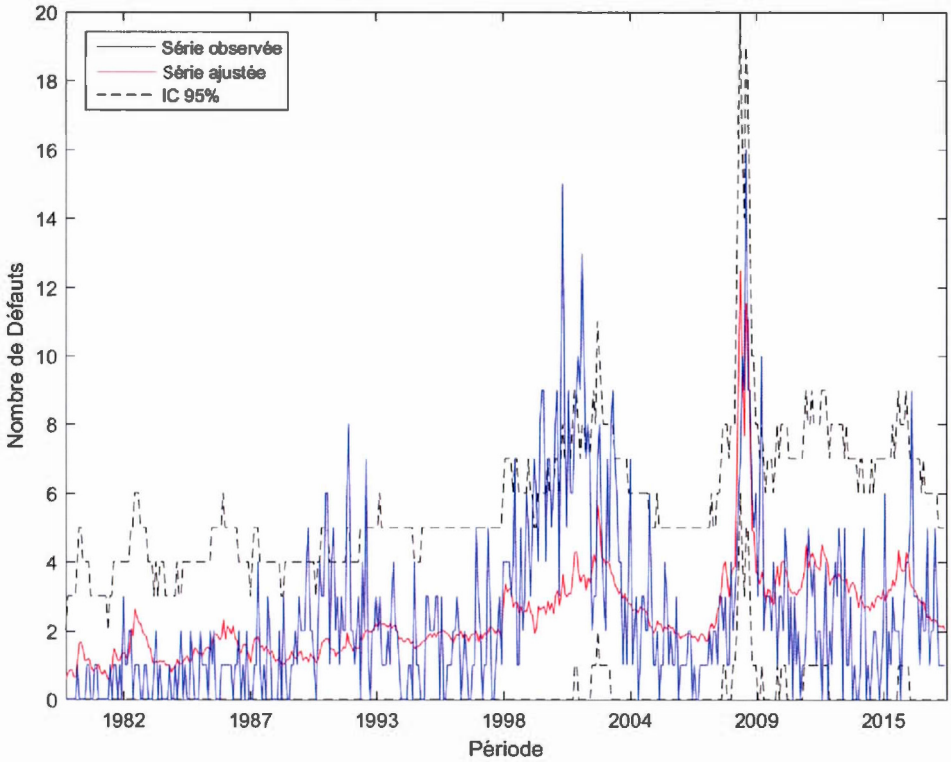


Figure 4.14 Application - HMM second ordre Poisson avec régresseurs $(Y_t|C_{t-1}, C_t)$

seulement l'état 1 semble significatif.

La valeur des EMV de ce modèle est fournie au tableau 4.5.

Encore une fois ici, l'utilisation de 4 états n'est peut-être pas optimale puisque la probabilité δ_1 est de 100%. De plus, le graphique semble moins bien ajuster les réalisations.

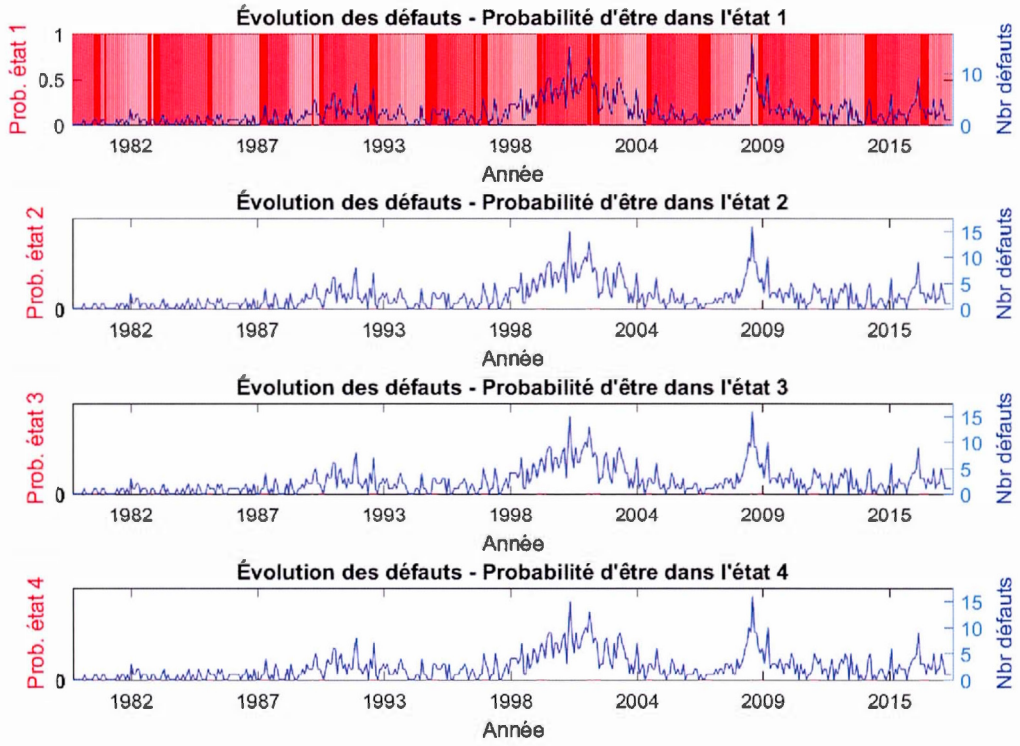


Figure 4.15 Probabilité $\hat{\zeta}_j(t)$ - HMM second ordre Poisson avec régresseurs $(Y_t|C_{t-1}, C_t)$

Tableau 4.5

EMV $\mathbf{\Gamma}$, δ_i, F_i et β_i - HMM second ordre Poisson avec régresseurs $(Y_t|C_{t-1}, C_t)$

$$\mathbf{\Gamma} = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.1495 & 0.8505 \\ 0.9398 & 0.0602 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.7758 & 0.2242 \end{bmatrix}$$

δ_i	Valeur δ_i
δ_1	1.0000
δ_2	0.0000
δ_3	0.0000
δ_4	0.0000

F_i	Valeur F_i
F_1	-0.3489
F_2	1.5824
F_3	3.2509
F_4	4.9469

β_i	Valeur β_i
β_1	41.0366
β_2	-3.7584
β_3	1.2447

4.9 Application - PMC-IN Poisson

La figure 4.16 représente l'application d'un modèle de chaîne de Markov couple à la série de défauts. Les distributions de Y_{t-1} et Y_t conditionnelles à C_{t-1} et C_t $(Y_{t-1}, Y_t | C_{t-1}, C_t)$ utilisées sont des distributions de Poisson indépendantes (voir la sous-section 3.4.2). L'algorithme EM a été utilisé pour trouver la valeur des EMV (voir 3.1).

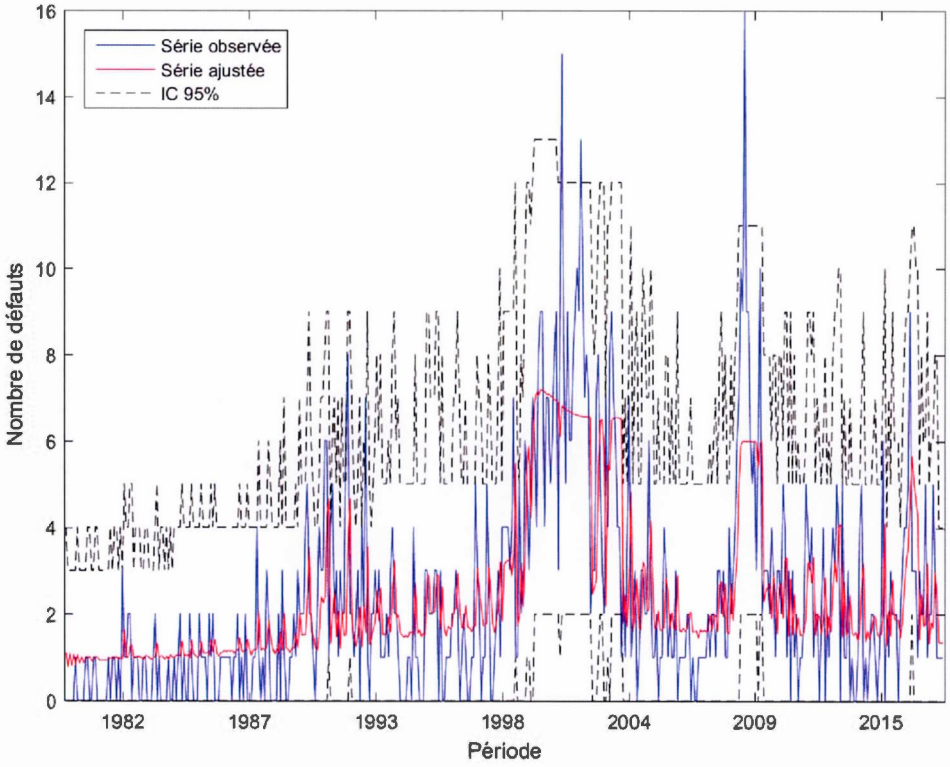


Figure 4.16 Application - PMC-IN Poisson

La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (3.19). L'IC est calculé à chaque période de temps selon l'équation (3.18). La valeur de $\lambda_{t,i}$ se calcule à partir de l'équation (4.4).

La figure 4.17 illustre la probabilité de se situer dans chacun des états à chaque période de temps t, soit $\hat{\zeta}_j(t)$. Il est possible de constater que les 2 états ne sont pas très bien identifiés en comparaison au modèle HMM.

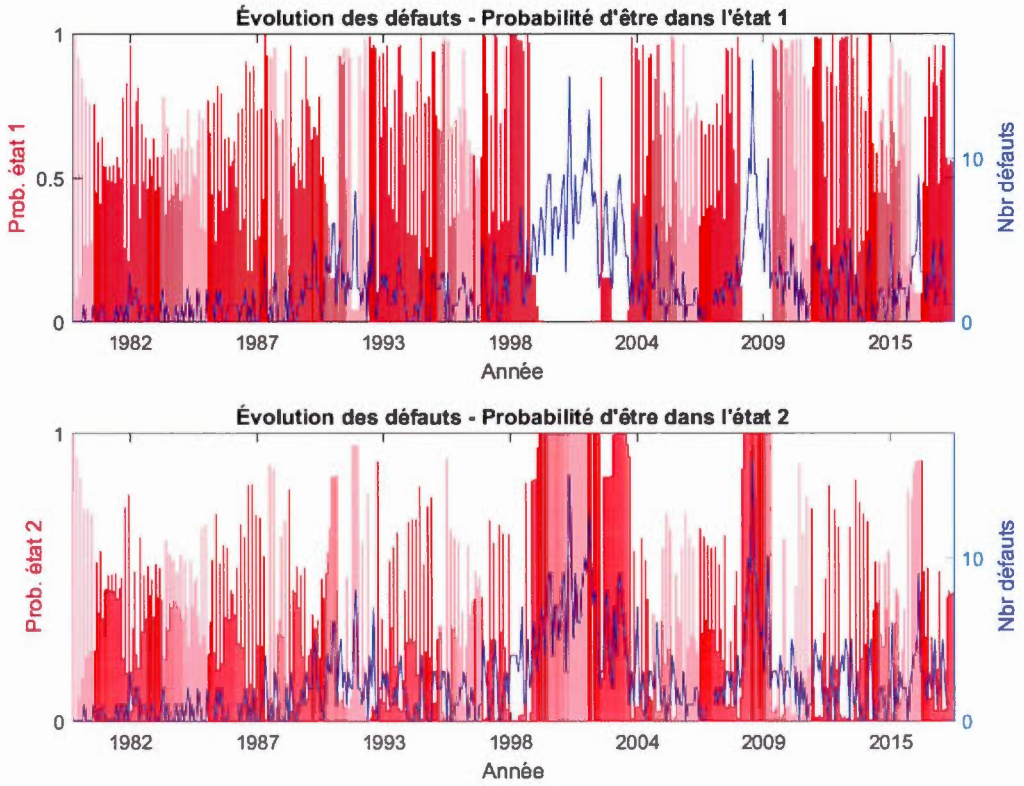


Figure 4.17 Probabilité $\hat{\zeta}_j(t)$ - PMC-IN Poisson

La valeur des EMV de ce modèle est fournie au tableau 4.6.

En comparaison au modèle HMM Poisson sans régresseur, le modèle PMC-IN sans régresseur semble mieux ajuster les observations. Cependant, celui-ci comporte 11 paramètres en comparaison à seulement 4.

Tableau 4.6

EMV c_{ij}, λ_{ij}^1 et λ_{ij}^2 - PMC-IN Poisson

$$c_{ij} = \begin{bmatrix} 0.0008 & 0.0043 \\ 0.9946 & 0.0003 \end{bmatrix}$$

Valeur i, j	Valeur λ_{ij}^1	Valeur λ_{ij}^2
$i = 1, j = 1$	0.0522	1.7896
$i = 1, j = 2$	0.0226	0.5966
$i = 2, j = 1$	0.0161	1.1094
$i = 2, j = 2$	0.5834	4.0096

4.10 Application - PMC-IN Poisson avec régresseurs

La figure 4.16 représente l'application d'un modèle de chaîne de Markov couple à la série de défauts. Les distributions de Y_{t-1} et Y_t conditionnelles à C_{t-1} et C_t $(Y_{t-1}, Y_t | C_{t-1}, C_t)$ utilisées sont des distributions de Poisson indépendantes avec régresseurs. L'algorithme EM a été utilisé pour trouver la valeur des EMV (voir 3.1).

La ligne rouge (série ajustée) est obtenue en appliquant, à chaque période de temps, l'équation (3.19). L'IC est calculé à chaque période de temps selon l'équation (3.18). La valeur de $\lambda_{t,i}$ se calcule à partir de l'équation (4.5).

La figure 4.19 illustre la probabilité de se situer dans chacun des états à chaque période de temps t, soit $\hat{\zeta}_j(t)$. Il y a une amélioration de l'identification des états entre le modèle PMC avec et sans régresseur.

La valeur des EMV de ce modèle est fournie au tableau 4.7.

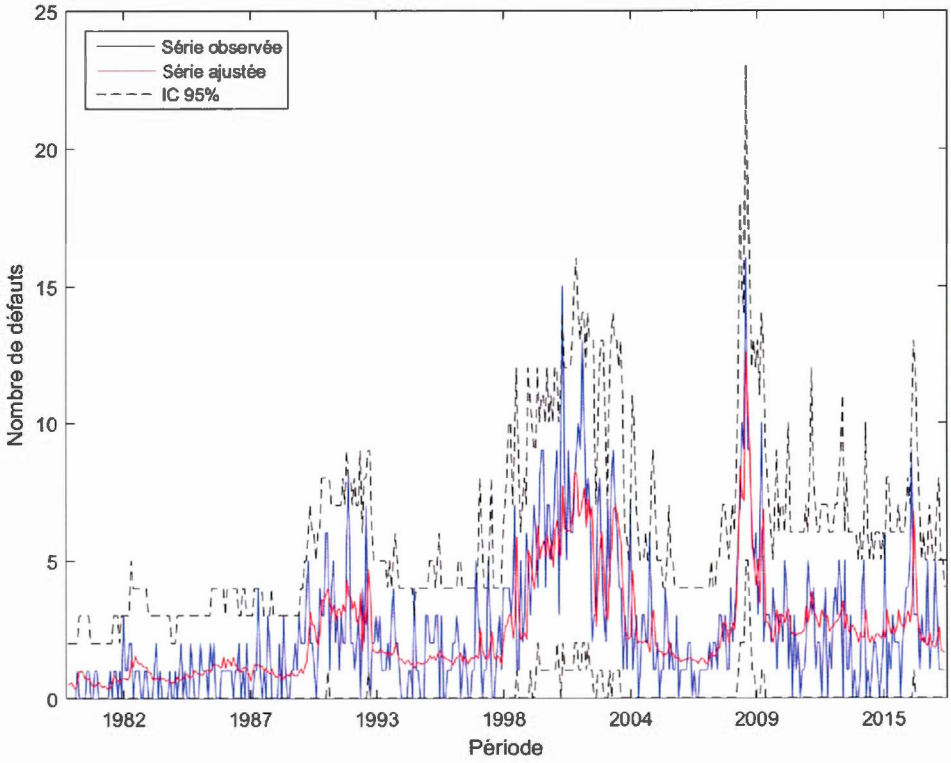


Figure 4.18 Application - PMC-IN Poisson avec régresseurs

Le modèle PMC-IN semble ajuster aussi bien les données que le modèle HMM Poisson avec régresseurs. Cependant, celui-ci comporte 14 paramètres en comparaison à seulement 7.

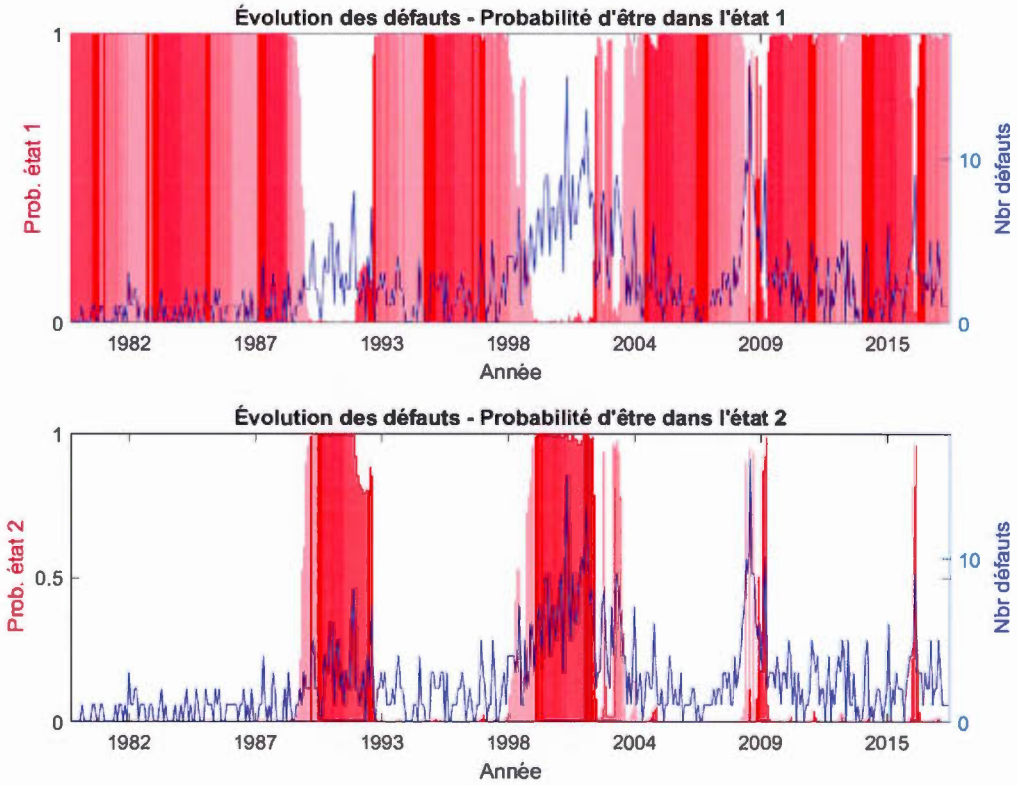


Figure 4.19 Probabilité $\hat{\zeta}_j(t)$ - PMC-IN Poisson avec régresseurs

Tableau 4.7

EMV c_{ij}, F_{ij}^1, F_{ij}^2 et β_i - PMC-IN Poisson avec régresseurs

$$c_{ij} = \begin{bmatrix} 0.9429 & 0.0007 \\ 0.0003 & 0.0561 \end{bmatrix}$$

Valeur i,j	Valeur F_{ij}^1	Valeur F_{ij}^2
$i=1, j=1$	0.0367	0.6748
$i=1, j=2$	0.1132	1.3113
$i=2, j=1$	78.4986	1.0706
$i=2, j=2$	79.6185	2.2012

β_i	Valeur β_i
β_1	31.1987
β_2	-5.7910
β_3	1.8041

4.11 Tableau sommaire des différents modèles

Afin de déterminer le meilleur modèle entre tous ceux appliqués dans le présent chapitre, le tableau 4.8 illustre pour chacun des modèles les valeurs de AIC et BIC (information sur AIC/BIC voir 1.4). De plus, la valeur de la log-vraisemblance et le nombre de paramètres de chacun des modèles y sont présentés.

Dans ce tableau, c'est la log-vraisemblance de chaque modèle qui y est comparé, et non le CDLL (voir 2.2 et 3.2 pour information sur CDLL). En effet, pour les modèles HMM et PMC :

$$\sum_i \alpha_T(i) = \sum_i \mathbb{P}(\boldsymbol{Y}^T, C_T = i) = \mathbb{P}(\boldsymbol{Y}^T) = L_T.$$

Donc, la valeur de la log-vraisemblance s'obtient directement grâce aux probabilités *Forward*.

Tableau 4.8

Tableau sommaire des différents modèles

Modèle #	Nombre de paramètres	$-1\times$Log-Vraisemblance	AIC	BIC
1	1	1065.0932	2132.1865	2136.3111
2	1	992.4796	1986.9592	1991.0838
3	3	893.2929	1792.5857	1804.9598
4	4	851.1057	1694.2114	1726.7102
5	7	784.5388	1583.0776	1611.9504
6	9	784.1560	1586.3120	1623.4341
7	11	888.8276	1799.6552	1845.0267
8	11	806.3172	1634.6344	1680.0059
9	14	771.3811	1570.7622	1628.5078

La liste ci-dessous fait la correspondance des numéros de modèle énoncé dans le tableau 4.8.

1. Poisson
2. Poisson avec exposition
3. Poisson avec régresseurs
4. HMM Poisson
5. HMM Poisson avec régresseurs
6. HMM second ordre Poisson avec régresseurs $(Y_t|C_t)$
7. HMM second ordre Poisson avec régresseurs $(Y_t|C_{t-1}, C_t)$
8. PMC-IN Poisson
9. PMC-IN Poisson avec régresseurs

Le tableau 4.8 présente le résultat de tous les modèles appliqués dans ce mémoire. Les modèles ayant mieux ajusté les données de défaut sont les modèles HMM Poisson avec régresseurs (selon le BIC) et PMC-IN Poisson avec régresseurs (selon le AIC). Le modèle PMC-IN Poisson avec régresseurs a réussi à surpasser les modèles HMM pour la valeur du AIC, mais pas pour la valeur du BIC. Cela s'explique notamment par le fait que les modèles PMC sont formés de beaucoup de paramètres, donc, selon la valeur du BIC, le modèle HMM Poisson avec régresseurs semble être un meilleur modèle que le modèle PMC-IN Poisson avec régresseurs pour des données financières. Il est aussi important de noter que les modèles PMC ne sont pas vraiment adaptés pour les séries temporelles. En effet, par construction, Y_t dépend de la valeur présente et future de l'état caché, soit C_t et C_{t+1}. Or, dans une série temporelle, le futur ne sera jamais connu. En contexte de reconnaissance d'image, le modèle PMC est plausible, mais avec une série temporelle, c'est inadéquat.

CONCLUSION

Le but de ce mémoire était de modéliser une série temporelle du nombre de défauts en risque de crédit à l'aide de modèles de chaînes de Markov cachées ainsi que de modèles de chaînes de Markov couples. Pour ce faire, la base de données du Département de droit de l'Université de Californie à Los Angeles (UCLA), qui contient plus de 1000 enregistrements d'évènements de défaut de grandes compagnies américaines de 1980 à aujourd'hui, a été utilisée.

Pour arriver à effectuer cette tâche, les modèles HMM et PMC ont été définis afin de mieux comprendre leur fonctionnement. Plusieurs formules et résultats indispensables à la bonne compréhension de ces modèles et à l'obtention de résultats ont été démontrés. Finalement, l'application des modèles a été effectuée sur la base de données BRD.

Les résultats obtenus démontrent que les modèles HMM ajustent mieux les observations de défaut que les modèles PMC, selon la valeur du BIC. La performance des modèles HMM et PMC est cependant très semblable. En effet, la valeur du AIC du modèle PMC-IN Poisson avec régresseurs était légèrement supérieure à celle de tous autres modèles utilisés dans ce mémoire.

Les modèles PMC n'étaient pas les meilleurs puisqu'ils comportent beaucoup de paramètres. Le nombre de paramètres de ceux-ci était élevé par rapport à la dimension de la base de données. Les modèles PMC ont été créés à des fins de reconnaissance d'image. Les plus petites images utilisées dans la littérature sont souvent de 126x126 pixels, soit de plus de 15000 données. Il est évident que les modèles PMC n'auront pas la même stabilité avec seulement 457 observations

comme ce qui a été utilisé dans ce mémoire avec la base de données BRD. De plus, en contexte financier, le modèle PMC est inadéquat. En effet, ce modèle admet que l'état caché au temps $t+1$ (C_{t+1}) est connu au temps t. Or, dans une série temporelle de défauts, le futur ne peut pas être connu.

Il serait possible d'appliquer les modèles de chaînes de Markov triplet (Triplet Markov Chain - TMC) (Pieczynski, 2013) créés par M.Pieczynski après la conception du modèle PMC. Encore une fois, l'ajustement à des données financières ne sera peut-être pas idéal, car ces modèles ont aussi été créés à des fins de segmentation d'images.

Il serait aussi possible d'utiliser les modèles HMM multivariés, soit des modèles HMM ayant plus d'une série temporelle à l'étude (voir Zhao, 2010). Il serait intéressant d'utiliser ce type de modèle HMM en utilisant une série pour les compagnies moins bien cotées et une autre série pour les compagnies mieux cotées. En effet, comme la figure 0.1 l'illustre, les compagnies moins bien cotées ont une plus grande probabilité de défaut en période de crise financière.

Enfin, les modèles comme le HMM démontrent énormément de potentiel pour modéliser des séries temporelles financières. En ce sens, une forte croissance de l'application de ces modèles dans le domaine professionnel est très souhaitable.

ANNEXE A

DÉMONSTRATIONS HMM - CHAPITRE 2

Proposition 1.

$$\boxed{\begin{array}{l} \textit{Pour } 0 \leq t \leq T-1 \\ \qquad \mathbb{P}\left(\boldsymbol{Y}_1^T | C_t\right) = \mathbb{P}\left(\boldsymbol{Y}_1^t | C_t\right) \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T | C_t\right) \end{array}} \tag{A.1a}$$

Démonstration. Cette égalité montre l'indépendance conditionnelle entre la distribution de $\boldsymbol{Y}_1^t$ et $\boldsymbol{Y}_{t+1}^T$, sachant C_t.

Premièrement,

$$\begin{aligned} \mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)}) &= \mathbb{P}(\boldsymbol{Y}_1^T, \boldsymbol{C}_0^T) \\ &= \mathbb{P}(C_0) \prod_{k=1}^{T} \mathbb{P}(C_k|C_{k-1}) \mathbb{P}(Y_k|C_k) \\ &= \mathbb{P}(C_0) \prod_{k=1}^{t} \mathbb{P}(C_k|C_{k-1}) \mathbb{P}(Y_k|C_k) \prod_{k=t+1}^{T} \mathbb{P}(C_k|C_{k-1}) \mathbb{P}(Y_t|C_k). \end{aligned} \tag{A.2}$$

Si :

$$\mathbb{P}(\boldsymbol{Y}_{t+1}^T, \boldsymbol{C}_t^T) = \mathbb{P}(C_t) \prod_{k=t+1}^{T} \mathbb{P}(C_k|C_{k-1}) \mathbb{P}(Y_k|C_k),$$

alors

$$\begin{aligned}
&\mathbb{P}(C_0)\prod_{k=1}^{t}\mathbb{P}(C_k|C_{k-1})\mathbb{P}(Y_k|C_k)\prod_{k=t+1}^{T}\mathbb{P}(C_k|C_{k-1})\mathbb{P}(Y_k|C_k)\\
&=\mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_0^t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(C_t)\prod_{k=t+1}^{T}\mathbb{P}(C_k|C_{k-1})\mathbb{P}(Y_k|C_k)\\
&=\mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_0^t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{C}_t^T). \qquad \text{(A.3)}
\end{aligned}$$

Si la somme des $\boldsymbol{c}_0^{t-1}$ et $\boldsymbol{c}_{t+1}^T$ sur $\mathbb{P}(\boldsymbol{Y}^{(T)},\boldsymbol{C}^{(T)})$ est effectuée, alors :

$$\begin{aligned}
&\sum_{c_0,\ldots,c_{t-1}=1}^{m}\sum_{c_{t+1},\ldots,c_T=1}^{m}\mathbb{P}(\boldsymbol{Y}_1^T,\boldsymbol{C}_0^T)\\
&=\sum_{c_0,\ldots,c_{t-1}=1}^{m}\sum_{c_{t+1},\ldots,c_T=1}^{m}\mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_0^t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{C}_t^T)\\
&=\mathbb{P}(\boldsymbol{Y}_1^t,C_t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,C_t).
\end{aligned}$$

Finalement,

$$\begin{aligned}
\sum_{c_0,\ldots,c_{t-1}=1}^{m}\sum_{c_{t+1},\ldots,c_T=1}^{m}\mathbb{P}(\boldsymbol{Y}_1^T,\boldsymbol{C}_0^T)&=\mathbb{P}(\boldsymbol{Y}_1^t,C_t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,C_t)\\
\mathbb{P}(\boldsymbol{Y}_1^T,C_t)&=\mathbb{P}(\boldsymbol{Y}_1^t|C_t)\frac{\mathbb{P}(C_t)}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t)\mathbb{P}(C_t)\\
\mathbb{P}(\boldsymbol{Y}_1^T|C_t)\mathbb{P}(C_t)&=\mathbb{P}(\boldsymbol{Y}_1^t|C_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t)\mathbb{P}(C_t)\\
\mathbb{P}(\boldsymbol{Y}_1^T|C_t)&=\mathbb{P}(\boldsymbol{Y}_1^t|C_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t).
\end{aligned}$$

□

La proposition 1 sera utilisée dans les propositions 5 et 6.

Proposition 2.

Pour $T \geq t+1$

$$\mathbb{P}(\boldsymbol{Y}_1^T,\boldsymbol{C}_t^{t+1})=\mathbb{P}(\boldsymbol{Y}_1^t,C_t)\mathbb{P}(C_{t+1}|C_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_{t+1}) \qquad \text{(A.4a)}$$

Démonstration. Si la somme des $\boldsymbol{c}_0^{t-1}$ et $\boldsymbol{c}_{t+2}^T$ sur l'équation (A.3) est effectuée, alors :

$$\sum_{c_0,\ldots,c_{t-1}=1}^{m}\sum_{c_{t+2},\ldots,c_T=1}^{m}\mathbb{P}(\boldsymbol{Y}^{(T)},\boldsymbol{C}^{(T)}) = \sum_{c_0,\ldots,c_{t-1}=1}^{m}\sum_{c_{t+2},\ldots,c_T=1}^{m}\mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_0^t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{C}_t^T)$$

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}_1^T,\boldsymbol{C}_t^{t+1}) &= \mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{C}_t^{t+1}) \\
&= \mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_t)\frac{1}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{C}_t^{t+1})\mathbb{P}(C_t,C_{t+1}) \\
&= \mathbb{P}(\boldsymbol{Y}_1^t,\boldsymbol{C}_t)\frac{\mathbb{P}(C_t,C_{t+1})}{\mathbb{P}(C_t)}\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_{t+1}) \qquad \text{(A.5)} \\
&= \mathbb{P}(\boldsymbol{Y}_1^t,C_t)\mathbb{P}(C_{t+1}|C_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_{t+1}).
\end{aligned}$$

Dans l'équation (A.5), $\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{C}_t^{t+1}) = \mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_{t+1})$, car $\boldsymbol{Y}_{t+1}^T$ est indépendant de C_t.

La proposition 2 sera utilisée dans les propositions 4 et 8. □

Proposition 3.

$$\boxed{\begin{array}{l} \textit{Pour } 1 \le t \le T-1 \\ \qquad \mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_{t+1}) = \mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t,C_{t+1}) \end{array}} \qquad \text{(A.6a)}$$

Démonstration. Le côté droit de l'équation ci-dessus peut aussi s'écrire comme :

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t,C_{t+1}) &= \frac{1}{\mathbb{P}(C_t,C_{t+1})}\sum_{\boldsymbol{c}_{t+2}^T}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{C}_t^T) \\
&= \frac{1}{\mathbb{P}(C_t,C_{t+1})}\sum_{\boldsymbol{c}_{t+2}^T}\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{C}_t^T)\mathbb{P}(\boldsymbol{C}_t^T) \\
&= \frac{1}{\mathbb{P}(C_t,C_{t+1})}\sum_{\boldsymbol{c}_{t+2}^T}\prod_{k=t+2}^{T}\mathbb{P}(C_k|C_{k-1})\mathbb{P}(C_t,C_{t+1})\prod_{k=t+1}^{T}\mathbb{P}(Y_k|C_k) \\
&= \sum_{\boldsymbol{c}_{t+2}^T}\prod_{k=t+2}^{T}\mathbb{P}(C_k|C_{k-1})\prod_{k=t+1}^{T}\mathbb{P}(Y_k|C_k).
\end{aligned}$$

Le côté gauche de l'équation ci-dessus peut aussi s'écrire comme :

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_{t+1}) &= \frac{1}{\mathbb{P}(C_{t+1})}\sum_{\boldsymbol{c}_{t+2}^T}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{C}_{t+1}^T) \\
&= \frac{1}{\mathbb{P}(C_{t+1})}\sum_{\boldsymbol{c}_{t+2}^T}\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{C}_{t+1}^T)\mathbb{P}(\boldsymbol{C}_{t+1}^T) \\
&= \frac{1}{\mathbb{P}(C_{t+1})}\sum_{\boldsymbol{c}_{t+2}^T}\prod_{k=t+2}^{T}\mathbb{P}(C_k|C_{k-1})\mathbb{P}(C_{t+1})\prod_{k=t+1}^{T}\mathbb{P}(Y_k|C_k) \\
&= \sum_{\boldsymbol{c}_{t+2}^T}\prod_{k=t+2}^{T}\mathbb{P}(C_k|C_{k-1})\prod_{k=t+1}^{T}\mathbb{P}(Y_k|C_k).
\end{aligned}$$

Il y a donc une égalité entre le terme de droite et de gauche. □

La proposition 3 sera utilisée dans la proposition 5.

Proposition 4.

Pour $1 \leq t \leq T$

$$\alpha_t(j) = \mathbb{P}(C_t = j, \boldsymbol{Y}^{(t)}), \tag{A.7a}$$

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \tag{A.7b}$$

et

$$\boldsymbol{\alpha}_t = \boldsymbol{\delta}\boldsymbol{\Gamma}\boldsymbol{P}(y_1)\boldsymbol{\Gamma}\boldsymbol{P}(y_2)...\boldsymbol{\Gamma}\boldsymbol{P}(y_t) \tag{A.7c}$$

Démonstration. Sachant que $\boldsymbol{\alpha}_0 = \boldsymbol{\delta}$, alors :

$$\begin{aligned}
\alpha_1(j) &= \delta_j p_j(y_1) \\
&= \mathbb{P}(C_1 = j)\mathbb{P}(Y_1 = y_1|C_1 = j) \\
\alpha_1(j) &= \mathbb{P}(C_1 = j, Y_1).
\end{aligned}$$

Il s'en suit que

$$\boldsymbol{\alpha}_1 = \boldsymbol{\delta}\boldsymbol{P}(y_1).$$

Pour la période de temps $t+1$ la démonstration s'effectue à partir de la formule récursive (2.15), soit :

$$\begin{aligned}
\boldsymbol{\alpha}_{t+1} &= \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \\
\Rightarrow \alpha_{t+1}(j) &= \sum_{i=1}^{m} \alpha_t(i) \gamma_{ij} p_j(y_{t+1}) \\
&= \sum_i \mathbb{P}(\boldsymbol{Y}^{(t)}, C_t = i) \mathbb{P}(C_{t+1} = j | C_t = i) \mathbb{P}(Y_{t+1} | C_{t+1} = j).
\end{aligned}$$

La proposition 2 est utilisée pour obtenir :

$$\begin{aligned}
&\sum_i \mathbb{P}(\boldsymbol{Y}^{(t)}, C_t = i) \mathbb{P}(C_{t+1} = j | C_t = i) \mathbb{P}(Y_{t+1} | C_{t+1} = j) \\
&= \sum_i \mathbb{P}(\boldsymbol{Y}^{(t+1)}, C_t = i, C_{t+1} = j) \\
&= \mathbb{P}(\boldsymbol{Y}^{(t+1)}, C_{t+1} = j).
\end{aligned}$$

La formule récursive étant démontrée pour tout $t+1$, alors il est admis par induction que celle-ci est vrai pour tout t $(0 \leq t \leq T)$. Finalement, puisque $\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1})$, alors

$$\begin{aligned}
\alpha_{t+1} &= \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \\
&= \boldsymbol{\alpha}_{t-1} \boldsymbol{\Gamma} \boldsymbol{P}(y_t) \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \\
&\ldots \\
&= \boldsymbol{\alpha}_0 \boldsymbol{\Gamma} \boldsymbol{P}(y_1) \ldots \boldsymbol{\Gamma} \boldsymbol{P}(y_t) \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \\
&= \boldsymbol{\delta} \boldsymbol{\Gamma} \boldsymbol{P}(y_1) \ldots \boldsymbol{\Gamma} \boldsymbol{P}(y_t) \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}).
\end{aligned}$$

□

La proposition 4 est définie pour la première fois aux équations (2.14), (2.15) et(2.16).

Proposition 5.

Pour $1 \leq t \leq T-1$

$$\beta_t(i) = \mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t = i) \tag{A.8a}$$

$$\boldsymbol{\beta}_t' = \boldsymbol{\Gamma}\boldsymbol{P}(y_{t+1})\boldsymbol{\beta}_{t+1}' \tag{A.8b}$$

et

$$\boldsymbol{\beta}_t' = \boldsymbol{\Gamma}\boldsymbol{P}(y_{t+1})...\boldsymbol{\Gamma}\boldsymbol{P}(y_T)1' \tag{A.8c}$$

Démonstration. Premièrement, le relation suivante est démontrée, soit

$$\boldsymbol{\beta}_{T-1}' = \boldsymbol{\Gamma}\boldsymbol{P}(y_1)\mathbf{1}'.$$

La vecteur $\mathbf{1}'$ vient de l'hypothèse que $\boldsymbol{\beta}_T = \mathbf{1}'$. Pour un élément (une probabilité) du vecteur $\boldsymbol{\beta}_{T-1}$, soit $\beta_{T-1}(i) = \mathbb{P}(\boldsymbol{Y}_T | C_{T-1} = i)$, l'équation matricielle ci-dessus peut ce réécrire ainsi,

$$\beta_{T-1}(i) = \sum_j \mathbb{P}(C_T = j | C_{T-1} = i)\mathbb{P}(Y_T = y_T | C_T = j).$$

Avec la proposition 3, il s'en suit que :

$$\begin{aligned}
&\sum_j \mathbb{P}(C_T = j | C_{T-1} = i)\mathbb{P}(Y_T = y_T | C_T = j) \\
&= \sum_j \mathbb{P}(C_T = j | C_{T-1} = i)\mathbb{P}(Y_T = y_T | C_{T-1} = i, C_T = j) \\
&= \sum_j \mathbb{P}(C_T = j | C_{T-1} = i)\frac{\mathbb{P}(Y_T = y_T, C_{T-1} = i, C_T = j)}{\mathbb{P}(C_T = j, C_{T-1} = i)} \\
&= \sum_j \frac{\mathbb{P}(Y_T = y_T, C_{T-1} = i, C_T = j)}{\mathbb{P}(C_{T-1} = i)} \\
&= \frac{1}{\mathbb{P}(C_{T-1} = i)} \sum_j \mathbb{P}(Y_T = y_T, C_{T-1} = i, C_T = j) \\
&= \frac{\mathbb{P}(Y_T = y_T, C_{T-1} = i)}{\mathbb{P}(C_{T-1} = i)} \\
&= \mathbb{P}(Y_T = y_T | C_{T-1} = i).
\end{aligned}$$

De ceci, $\beta_{T-1}(i) = \mathbb{P}(Y_T = y_T | C_{T-1} = i)$. Pour la période de temps t la démonstration s'effectue à partir de la formule récursive (2.19), soit :

$$\begin{aligned} \boldsymbol{\beta}'_t &= \boldsymbol{\Gamma} \boldsymbol{P}(y_{t+1}) \boldsymbol{\beta}'_{t+1} \\ \Rightarrow \beta_t(i) &= \sum_j \gamma_{ij} \mathbb{P}(Y_{t+1} = y_{t+1} | C_{t+1} = j) \mathbb{P}(\boldsymbol{Y}_{t+2}^T = \boldsymbol{y}_{t+2}^T | C_{t+1} = j). \end{aligned}$$

Avec les proposition 1 et 3, il s'en suit que :

$$\mathbb{P}(Y_{t+1} = y_{t+1} | C_{t+1} = j) \mathbb{P}(\boldsymbol{Y}_{t+2}^T = \boldsymbol{y}_{t+2}^T | C_{t+1} = j) = \mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t, C_{t+1}). \quad \text{(A.9)}$$

Si $\mathbb{P}(Y_{t+1} = y_{t+1} | C_{t+1} = j) \mathbb{P}(\boldsymbol{Y}_{t+2}^T = \boldsymbol{y}_{t+2}^T | C_{T+1} = j)$ est substitué par (A.9) dans $\beta_t(i)$, alors

$$\begin{aligned} &\sum_j \gamma_{ij} \mathbb{P}(Y_{t+1} = y_{t+1} | C_{t+1} = j) \mathbb{P}(\boldsymbol{Y}_{t+2}^T = \boldsymbol{y}_{t+2}^T | C_{t+1} = j) \\ &= \sum_j \mathbb{P}(C_{t+1} = j | C_t = i) \mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t = i, C_{t+1} = j) \\ &= \sum_j \mathbb{P}(C_{t+1} = j | C_t = i) \frac{\mathbb{P}(\boldsymbol{Y}_{t+1}^T, C_t = i, C_{t+1} = j)}{\mathbb{P}(C_t = i, C_{t+1} = j)} \\ &= \frac{1}{\mathbb{P}(C_t = i)} \sum_j \mathbb{P}(\boldsymbol{Y}_{t+1}^T, C_t = i, C_{t+1} = j) \\ &= \frac{\mathbb{P}(\boldsymbol{Y}_{t+1}^T, C_t = i)}{\mathbb{P}(C_t = i)} \\ &= \mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t = i). \end{aligned}$$

La formule récursive étant démontrée pour tout t, alors il est admis par induction que celle-ci est vrai pour tout temps t $(0 \leq t \leq T)$. Finalement, puisque $\boldsymbol{\beta}'_t =$

$\boldsymbol{\Gamma P}(y_{t+1})\boldsymbol{\beta}'_{t+1}$, alors :

$$\begin{aligned}\boldsymbol{\beta}'_t &= \boldsymbol{\Gamma P}(y_{t+1})\boldsymbol{\beta}'_{t+1} \\ &= \boldsymbol{\Gamma P}(y_{t+1})\boldsymbol{\Gamma P}(y_{t+2})\boldsymbol{\beta}'_{t+2} \\ &\dots \\ &= \boldsymbol{\Gamma P}(y_{t+1})...\boldsymbol{\Gamma P}(y_T)\boldsymbol{\beta}'_T \\ &= \boldsymbol{\Gamma P}(y_{t+1})...\boldsymbol{\Gamma P}(y_T)1'.\end{aligned}$$

□

La proposition 5 est définie pour la première fois aux équations (2.17), (2.19) et (2.20).

Proposition 6.

$$\boxed{\begin{array}{l}\textit{Pour } 0 \leq t \leq T \\ \qquad \alpha_t(i)\beta_t(i) = \mathbb{P}\left(\boldsymbol{Y}^{(T)}, C_t = i\right)\end{array}} \tag{A.10a}$$

Démonstration. Premièrement :

$$\begin{aligned}\alpha_t(i)\beta_t(i) &= \mathbb{P}(C_t = i, \boldsymbol{Y}^{(t)})\mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t = i) \\ &= \mathbb{P}(C_t = i)\mathbb{P}(\boldsymbol{Y}^{(t)} | C_t = i)\mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t = i).\end{aligned}$$

Avec la proposition 1, il s'en suit que :

$$\begin{aligned}&\mathbb{P}(C_t = i)\mathbb{P}(\boldsymbol{Y}^{(t)} | C_t = i)\mathbb{P}(\boldsymbol{Y}_{t+1}^T | C_t = i) \\ &= \mathbb{P}(C_t = i)\mathbb{P}\left(\boldsymbol{Y}_1^T | C_t\right) \\ &= \mathbb{P}\left(\boldsymbol{Y}_1^T, C_t\right).\end{aligned}$$

□

La proposition 6 sera utilisée dans les propositions 7 et 9. De plus, la proposition 6 est définie pour la première fois à l'équation (2.21).

Proposition 7.

$$\boxed{\begin{array}{l} \textit{Pour } 0 \leq t \leq T \\ \qquad \dfrac{\alpha_t(i)\beta_t(i)}{L_T} = \mathbb{P}(C_t = i|\boldsymbol{Y}^{(T)}) \end{array}} \tag{A.11a}$$

Démonstration. À l'aide de la proposition 6 :

$$\begin{aligned} \frac{\alpha_t(i)\beta_t(i)}{L_T} &= \frac{\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, C_t = i)}{\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})} \\ &= \frac{\mathbb{P}(C_t = i|\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})}{\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})} \\ &= \mathbb{P}(C_t = i|\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}). \end{aligned}$$

□

La proposition 7 est définie pour la première fois à l'équation (2.22).

Proposition 8.

$$\boxed{\begin{array}{l} \textit{Pour } 0 \leq t \leq T \\ \dfrac{\alpha_{t-1}(i)\gamma_{ij}p_j(y_t)\beta_t(j)}{L_T} = \mathbb{P}(C_{t-1} = i, C_t = j|\boldsymbol{Y}^{(T)}) \end{array}} \tag{A.12a}$$

Démonstration. Premièrement, puisque $p_j(y_t) = \mathbb{P}(Y_t|C_t = j)$ et que

$$\mathbb{P}(Y_t|C_t = j)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t = j) = \mathbb{P}(\boldsymbol{Y}_t^T|C_t = j),$$

alors,

$$\begin{aligned} \frac{\alpha_{t-1}(i)\gamma_{ij}p_j(y_t)\beta_t(j)}{L_T} &= \frac{\mathbb{P}(C_{t-1} = i, \boldsymbol{Y}^{(t-1)})\gamma_{ij}p_j(y_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|C_t = j)}{\mathbb{P}(\boldsymbol{Y}^{(T)})} \\ &= \frac{\mathbb{P}(C_{t-1} = i, \boldsymbol{Y}^{(t-1)})\gamma_{ij}\mathbb{P}(\boldsymbol{Y}_t^T|C_t = j)}{\mathbb{P}(\boldsymbol{Y}^{(T)})}. \end{aligned} \tag{A.13}$$

La proposition 2 est utilisée sur le numérateur de (A.13), soit :

$$\begin{aligned}
&\frac{\mathbb{P}(C_{t-1}=i,\boldsymbol{Y}^{(t-1)})\gamma_{ij}\mathbb{P}(\boldsymbol{Y}_t^T|C_t=j)}{\mathbb{P}(\boldsymbol{Y}^{(T)})}\\
&=\frac{\mathbb{P}(C_{t-1}=i,C_t=j,\boldsymbol{Y}^{(T)})}{\mathbb{P}(\boldsymbol{Y}^{(T)})}\\
&=\frac{\mathbb{P}(C_{t-1}=i,C_t=j|\boldsymbol{Y}^{(T)})\mathbb{P}(\boldsymbol{Y}^{(T)})}{\mathbb{P}(\boldsymbol{Y}^{(T)})}\\
&=\mathbb{P}(C_{t-1}=i,C_t=j|\boldsymbol{Y}^{(T)}).
\end{aligned}$$

□

La proposition 8 est définie pour la première fois à l'équation (2.23).

Proposition 9.

$$\boxed{\begin{array}{l}\textit{Pour } 0\leq t\leq T\\ \qquad\qquad \boldsymbol{\alpha}_t\boldsymbol{\beta}_t'=L_T\end{array}} \tag{A.14a}$$

Démonstration. À l'aide de la proposition 6 :

$$\begin{aligned}
\boldsymbol{\alpha}_t\boldsymbol{\beta}_t' &= \sum_i \alpha_t(i)\beta_t(i)\\
&= \sum_i \mathbb{P}(\boldsymbol{Y}^{(T)}=\boldsymbol{y}^{(T)},C_t=i)\\
&= \sum_i \mathbb{P}(\boldsymbol{Y}^{(T)}=\boldsymbol{y}^{(T)}|C_t=i)\mathbb{P}(C_t=i)\\
&= \mathbb{P}(\boldsymbol{Y}^{(T)}=\boldsymbol{y}^{(T)})\\
&= L_T.
\end{aligned}$$

□

La proposition 9 est définie pour la première fois à l'équation (2.24).

ANNEXE B

DÉMONSTRATIONS PMC - CHAPITRE 3

Proposition 10.

$$\boxed{\begin{aligned} c_{ij} &= \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{T-1} \qquad \text{(B.1a)} \\ f_{i,j}(l,m) &= \frac{\sum_{t=2:(y_{t-1},y_t=(l,m))} \hat{\psi}_t^{(n)}(i,j)}{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)} \qquad \text{(B.1b)} \end{aligned}}$$

Démonstration. L'équation L suivante (définit à l'équation (3.14)) :

$$\begin{aligned} L(c_{ij}, f_{ij}, \lambda_1, \lambda_{j,m}, \lambda_3) = &\sum_{t=2}^{T} \sum_{i,j \in \Omega^2} \hat{\psi}_t^{(n)}(i,j) \log c_{ij} - \sum_{t=2}^{T-1} \sum_{i \in \Omega} \hat{\zeta}_t^{(n)}(i) \log b_j(m) \\ &+ \sum_{t=2}^{T} \sum_{i,j \in \Omega^2} \hat{\psi}_t^{(n)}(i,j) \log f_{i,j}(l,m) - \lambda_1 \left(\sum_{(i,j) \in \Omega^2} c_{ij} - 1 \right) \\ &- \sum_{(j,m) \in \Omega \times \mathbb{R}} \lambda_{j,m} \left(\sum_{(i,l) \in \Omega \times \mathbb{R}} c_{ij} f_{i,j}(l,m) - b_j(m) \right) \\ &- \lambda_3 \left(\int_{(l,m) \in \mathbb{R}^2} f_{i,j}(l,m) - 1 \right), \end{aligned}$$

sera dérivée par rapport à c_{ij}, f_{ij}, λ_1, λ_3 et les $\lambda_{j,m}$. Ainsi, pour les différents paramètres, il s'en suit que : λ_1 :

$$\begin{aligned} \frac{\partial L}{\partial \lambda_1} = 0 &= \sum_{(i,j) \in \Omega^2} c_{ij} - 1 \\ \Rightarrow \sum_{(i,j) \in \Omega^2} c_{ij} &= 1, \qquad \text{(B.2)} \end{aligned}$$

$\lambda_{j,m}$:

$$\frac{\partial L}{\partial \lambda_{j,m}} = 0 = \sum_{(i,l)\in\Omega\times\mathbb{R}} c_{ij} f_{i,j}(l,m) - b_j(m)$$

$$\Rightarrow \sum_{(i,l)\in\Omega\times\mathbb{R}} c_{ij} f_{i,j}(l,m) = b_j(m),$$

λ_3 :

$$\frac{\partial L}{\partial \lambda_3} = 0 = \int_{(l,m)\in\mathbb{R}^2} f_{i,j}(l,m) - 1$$

$$\Rightarrow \int_{(l,m)\in\mathbb{R}^2} f_{i,j}(l,m) = 1 \qquad \text{(Version continue)},$$

$$\Rightarrow \sum_{(l,m)\in\mathbb{N}^2} f_{i,j}(l,m) = 1 \qquad \text{(Version discrète)}, \tag{B.3}$$

c_{ij} :

$$\frac{\partial L}{\partial c_{ij}} = 0 = \sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)\frac{1}{c_{ij}} - \lambda_1 - \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m)$$

$$\Rightarrow c_{ij} = \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_1 + \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m)}. \tag{B.4}$$

Si le résultat en (B.2) est utilisé, alors :

$$\sum_{(i,j)\in\Omega^2} c_{ij} = 1 = \frac{\sum_{(i,j)\in\Omega^2} \sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_1 + \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m)}$$

$$= \frac{T-1}{\lambda_1 + \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m)}$$

$$\Rightarrow \lambda_1 = (T-1) - \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m).$$

En remettant cette dernière équation dans (B.4), ceci donne les résultats suivants :

$$c_{ij} = \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_1 + \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m)}$$

$$= \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{(T-1) - \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m) + \sum_{m\in\mathbb{R}} \lambda_{j,m} \sum_{l\in\mathbb{R}} f_{i,j}(l,m)}$$

$$= \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{T-1},$$

$f_{ij}(l,m)$:

$$\frac{\partial L}{\partial f_{ij}(l,m)} = 0 = \sum_{t=2:(y_{t-1},y_t=(l,m))}^{T} \hat{\psi}_t^{(n)}(i,j)\frac{1}{f_{i,j}(l,m)} - \lambda_3$$

$$\Rightarrow f_{i,j}(l,m) = \frac{\sum_{t=2:(y_{t-1},y_t=(l,m))}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_3}. \tag{B.5}$$

Avec le résultat en (B.3), il s'en suit que :

$$\sum_{(l,m)\in\mathbb{N}^2} f_{i,j}(l,m) = 1 = \sum_{(l,m)\in\mathbb{N}^2} \frac{\sum_{t=2:(y_{t-1},y_t=(l,m))}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_3}$$

$$= \frac{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}{\lambda_3}$$

$$\Rightarrow \lambda_3 = \sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j).$$

Si cette dernière égalité est remise dans l'équation (B.5), alors :

$$f_{i,j}(l,m) = \frac{\sum_{t=2:(y_{t-1},y_t=(l,m))} \hat{\psi}_t^{(n)}(i,j)}{\sum_{t=2}^{T} \hat{\psi}_t^{(n)}(i,j)}.$$

□

La proposition 10 est définie pour la première fois aux équations (3.15) et (3.16).

Proposition 11.

Pour $0 \leq t \leq T-1$

$$\mathbb{P}\left(\boldsymbol{Y}^{(-t)}|\boldsymbol{Z}_t\right) = \mathbb{P}\left(\boldsymbol{Y}_1^{t-1}|\boldsymbol{Z}_t\right)\mathbb{P}\left(\boldsymbol{Y}_{t+1}^{T}|\boldsymbol{Z}_t\right) \tag{B.6a}$$

Démonstration. Premièrement,

$$\mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)}) = \mathbb{P}(\boldsymbol{Z}_1)\prod_{k=1}^{T-1}\mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k)$$

$$= \mathbb{P}(\boldsymbol{Z}_1)\prod_{k=1}^{t-1}\mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k)\prod_{k=t}^{T-1}\mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k). \tag{B.7}$$

Si

$$\mathbb{P}(\boldsymbol{Y}_t^T, \boldsymbol{C}_t^T) = \mathbb{P}(\boldsymbol{Z}_t) \prod_{k=t}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k),$$

alors

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{Z}_1) \prod_{k=1}^{t-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k) \prod_{k=t}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k) \\
&= \mathbb{P}(\boldsymbol{Y}^{(t)}, \boldsymbol{C}^{(t)}) \frac{\mathbb{P}(\boldsymbol{Z}_t)}{\mathbb{P}(\boldsymbol{Z}_t)} \prod_{k=t}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k) \\
&= \mathbb{P}(\boldsymbol{Y}^{(t)}, \boldsymbol{C}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T, \boldsymbol{C}_{t+1}^T, \boldsymbol{Z}_t\right). \qquad \text{(B.8)}
\end{aligned}$$

Si la somme des $\boldsymbol{c}_1^{t-1}$ et $\boldsymbol{c}_{t+1}^T$ sur $\mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)})$ est effectuée, alors :

$$\begin{aligned}
&\sum_{c_1,\ldots,c_{t-1}=1}^{m} \sum_{c_{t+1},\ldots,c_T=1}^{m} \mathbb{P}(\boldsymbol{Y}_1^T, \boldsymbol{C}_1^T) \\
&= \sum_{c_1,\ldots,c_{t-1}=1}^{m} \sum_{c_{t+1},\ldots,c_T=1}^{m} \mathbb{P}(\boldsymbol{Y}^{(t)}, \boldsymbol{C}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T, \boldsymbol{C}_{t+1}^T, \boldsymbol{Z}_t\right) \\
&= \mathbb{P}(\boldsymbol{Y}^{(t-1)}, \boldsymbol{Z}_t) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T, \boldsymbol{Z}_t\right).
\end{aligned}$$

Finalement,

$$\begin{aligned}
\sum_{c_1,\ldots,c_{t-1}=1}^{m} \sum_{c_{t+1},\ldots,c_T=1}^{m} \mathbb{P}(\boldsymbol{Y}_1^T, \boldsymbol{C}_1^T) &= \mathbb{P}(\boldsymbol{Y}^{(t-1)}, \boldsymbol{Z}_t) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T, \boldsymbol{Z}_t\right) \\
\mathbb{P}(\boldsymbol{Y}^{(-t)}, \boldsymbol{Z}_t) &= \mathbb{P}(\boldsymbol{Y}^{(t-1)}|\boldsymbol{Z}_t) \frac{\mathbb{P}(\boldsymbol{Z}_t)}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t\right) \mathbb{P}(\boldsymbol{Z}_t) \\
\mathbb{P}(\boldsymbol{Y}^{(-t)}|\boldsymbol{Z}_t)\mathbb{P}(\boldsymbol{Z}_t) &= \mathbb{P}(\boldsymbol{Y}^{(t-1)}|\boldsymbol{Z}_t) \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t\right) \mathbb{P}(\boldsymbol{Z}_t) \\
\mathbb{P}(\boldsymbol{Y}^{(-t)}|\boldsymbol{Z}_t) &= \mathbb{P}(\boldsymbol{Y}^{(t-1)}|\boldsymbol{Z}_t) \mathbb{P}\left(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t\right).
\end{aligned}$$

□

La proposition 11 sera utilisée dans la proposition 16.

Proposition 12.

$$\boxed{\begin{aligned} &\textit{Pour } T \geq t+1 \\ &\mathbb{P}(\boldsymbol{Y}_1^{t-1}, \boldsymbol{Y}_{t+2}^{T}, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}) = \frac{\mathbb{P}(\boldsymbol{Y}_1^{t-1}, \boldsymbol{Z}_t)}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}(\boldsymbol{Y}_{t+2}^{T}, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}) \qquad \text{(B.9a)} \\ &\qquad\qquad = \mathbb{P}(\boldsymbol{Y}_1^{T}, C_t, C_{t+1}) \qquad \text{(B.9b)} \end{aligned}}$$

Démonstration. Cette démonstration débute à partir de l'équation (B.8), soit :

$$\begin{aligned} &\mathbb{P}(\boldsymbol{Y}^{(t)}, \boldsymbol{C}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+1}^{T}, \boldsymbol{C}_{t+1}^{T}, \boldsymbol{Z}_t\right) \\ &= \mathbb{P}(\boldsymbol{Y}^{(t)}, \boldsymbol{C}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+2}^{T}, \boldsymbol{C}_{t+2}^{T}, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}\right). \end{aligned}$$

Finalement, si la somme des $\boldsymbol{c}_1^{t-1}$ et $\boldsymbol{c}_{t+2}^{T}$ sur $\mathbb{P}(\boldsymbol{Y}^{(T)}, \boldsymbol{C}^{(T)})$ est effectuée, alors :

$$\sum_{c_1,\ldots,c_{t-1}=1}^{m} \sum_{c_{t+2},\ldots,c_T=1}^{m} \mathbb{P}(\boldsymbol{Y}_1^{T}, \boldsymbol{C}_1^{T}) = \sum_{c_1,\ldots,c_{t-1}=1}^{m} \sum_{c_{t+2},\ldots,c_T=1}^{m} \mathbb{P}(\boldsymbol{Y}^{(t)}, \boldsymbol{C}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+2}^{T}, \boldsymbol{C}_{t+2}^{T}, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}\right)$$

$$\mathbb{P}(\boldsymbol{Y}_1^{T}, C_t, C_{t+1}) = \mathbb{P}(\boldsymbol{Y}^{(t-1)}, \boldsymbol{Z}_t) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}\left(\boldsymbol{Y}_{t+2}^{T}, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}\right).$$

□

La proposition 12 sera utilisée dans les propositions 14 et 18.

Proposition 13.

$$\boxed{\begin{aligned} &\textit{Pour } 1 \leq t \leq T-1 \\ &\qquad \mathbb{P}(\boldsymbol{Y}_{t+2}^{T} | \boldsymbol{Z}_{t+1}) = \mathbb{P}(\boldsymbol{Y}_{t+2}^{T} | \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}) \end{aligned}} \qquad \text{(B.10a)}$$

Démonstration. Le côté droit de l'équation ci-dessus peut aussi s'écrire comme :

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}_{t+2}^T|\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}) &= \frac{1}{\mathbb{P}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \mathbb{P}(\boldsymbol{Y}_{t+2}^T, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}\boldsymbol{C}_{t+2}^T) \\
&= \frac{1}{\mathbb{P}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \mathbb{P}(\boldsymbol{Y}_t^T, \boldsymbol{C}_t^T) \\
&= \frac{1}{\mathbb{P}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \mathbb{P}(\boldsymbol{Z}_t)\mathbb{P}(\boldsymbol{Z}_{t+1}|\boldsymbol{Z}_t) \prod_{k=t+1}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k) \\
&= \frac{\mathbb{P}(\boldsymbol{Z}_t)\mathbb{P}(\boldsymbol{Z}_{t+1}|\boldsymbol{Z}_t)}{\mathbb{P}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \prod_{k=t+1}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k) \\
&= \sum_{\boldsymbol{c}_{t+2}^T} \prod_{k=t+1}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k).
\end{aligned}$$

Le côté gauche de l'équation ci-dessus peut aussi s'écrire comme :

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Y}_{t+2}^T|\boldsymbol{Z}_{t+1}) &= \frac{1}{\mathbb{P}(\boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \mathbb{P}(\boldsymbol{Y}_{t+2}^T, \boldsymbol{Z}_{t+1}, \boldsymbol{C}_{t+2}^T) \\
&= \frac{1}{\mathbb{P}(\boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \mathbb{P}(\boldsymbol{Y}_{t+1}^T, \boldsymbol{c}_{t+1}^T) \\
&= \frac{1}{\mathbb{P}(\boldsymbol{Z}_{t+1})} \sum_{\boldsymbol{c}_{t+2}^T} \mathbb{P}(\boldsymbol{Z}_{t+1}) \prod_{k=t+1}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k) \\
&= \sum_{\boldsymbol{c}_{t+2}^T} \prod_{k=t+1}^{T-1} \mathbb{P}(\boldsymbol{Z}_{k+1}|\boldsymbol{Z}_k).
\end{aligned}$$

Il y a donc une égalité entre le terme de droite et de gauche. □

La proposition 13 sera utilisée dans les propositions 15 et 18.

Proposition 14.

$$\boxed{\begin{array}{l} \textit{Pour } 1 \leq t \leq T \\ \qquad \alpha_t(j) = \mathbb{P}(C_t = j, \boldsymbol{Y}^{(t)}) \end{array}} \tag{B.11a}$$

Démonstration. Puisque $\boldsymbol{\alpha}_1 = \mathbb{P}(\boldsymbol{Z}_1) = \mathbb{P}(Y_1, C_1)$, alors :

$$\alpha_1(i) = \sum_{j \in \Omega} \mathbb{P}(i,j) \sum_{y_{t+1}} f_{i,j}(y_t, y_{t+1}).$$

De ceci, $\alpha_1(i) = \mathbb{P}(C_1 = i, Y_1)$. Pour $t+1$, cette démonstration débute avec la récursive (3.8) :

$$\begin{aligned}
\alpha_{t+1}(j) &= \sum_{i=1}^{m} \alpha_t(i) \mathbb{P}(\boldsymbol{Z}_{t+1} | \boldsymbol{Z}_t) \\
&= \sum_{i=1}^{m} \mathbb{P}(C_t = i, \boldsymbol{Y}^{(t)}) \mathbb{P}(\boldsymbol{Z}_{t+1} | \boldsymbol{Z}_t) \\
&= \sum_{i=1}^{m} \mathbb{P}(C_t = i, \boldsymbol{Y}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}).
\end{aligned}$$

Par la suite, la proposition 12 est utilisée pour obtenir :

$$\begin{aligned}
&\sum_{i=1}^{m} \mathbb{P}(C_t = i, \boldsymbol{Y}^{(t)}) \frac{1}{\mathbb{P}(\boldsymbol{Z}_t)} \mathbb{P}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+1}) \\
&= \sum_{i} \mathbb{P}(\boldsymbol{Y}^{(t-1)}, \boldsymbol{Z}_t(i), \boldsymbol{Z}_{t+1}(j)) \\
&= \sum_{i} \mathbb{P}(\boldsymbol{Y}^{(t+1)}, C_t = i, C_{t+1} = j) \\
&= \mathbb{P}(\boldsymbol{Y}^{(t+1)}, C_{t+1} = j),
\end{aligned}$$

où $\boldsymbol{Z}_t(i) = (Y_t = y_t, C_t = i)$. □

La proposition 14 est définie pour la première fois à l'équation (3.8).

Proposition 15.

$$\boxed{\begin{array}{l} \textit{Pour } 1 \leq t \leq T-1 \\ \qquad \beta_t(i) = \mathbb{P}(\boldsymbol{Y}_{t+1}^{T} | \boldsymbol{Z}_t(i)) \end{array}} \tag{B.12a}$$

Démonstration. Cette démonstration débute avec la formule récursive (3.9) :

$$\begin{aligned}\beta_t(i) &= \sum_j \beta_{t+1}(j)\mathbb{P}(\boldsymbol{Z}_{t+1}(j)|\boldsymbol{Z}_t(i)) \\ &= \sum_j \mathbb{P}(\boldsymbol{Y}_{t+2}^T|\boldsymbol{Z}_{t+1}(j))\mathbb{P}(\boldsymbol{Z}_{t+1}(j)|\boldsymbol{Z}_t(i)).\end{aligned}$$

Avec la proposition 13, il s'en suit que

$$\begin{aligned}&\sum_j \mathbb{P}(\boldsymbol{Y}_{t+2}^T|\boldsymbol{Z}_{t+1}(j))\mathbb{P}(\boldsymbol{Z}_{t+1}(j)|\boldsymbol{Z}_t(i)) \\ &= \sum_j \mathbb{P}(\boldsymbol{Z}_{t+1}(j)|\boldsymbol{Z}_t(i))\mathbb{P}(\boldsymbol{Y}_{t+2}^T|\boldsymbol{Z}_t(i), \boldsymbol{Z}_{t+1}(j)) \\ &= \sum_j \frac{\mathbb{P}(\boldsymbol{Z}_t(i), \boldsymbol{Z}_{t+1}(j))}{\mathbb{P}(\boldsymbol{Z}_t(i))}\mathbb{P}(\boldsymbol{Y}_{t+2}^T|\boldsymbol{Z}_t(i), \boldsymbol{Z}_{t+1}(j)) \\ &= \frac{1}{\mathbb{P}(\boldsymbol{Z}_t(i))}\sum_j \mathbb{P}(\boldsymbol{Y}_{t+2}^T, \boldsymbol{Z}_t(i), \boldsymbol{Z}_{t+1}(j)) \\ &= \frac{\mathbb{P}(\boldsymbol{Y}_{t+1}^T, \boldsymbol{Z}_t(i))}{\mathbb{P}(\boldsymbol{Z}_t(i))} \\ &= \mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(i)).\end{aligned}$$

□

La proposition 15 est définie pour la première fois à l'équation (3.9).

Proposition 16.

$$\boxed{\begin{aligned}&\textit{Pour } 0 \leq t \leq T \\ &\qquad \alpha_t(i)\beta_t(i) = \mathbb{P}\left(\boldsymbol{Y}^{(T)}, C_t = i\right)\end{aligned}} \tag{B.13a}$$

Démonstration. Premièrement,

$$\begin{aligned}\alpha_t(i)\beta_t(i) &= \mathbb{P}(C_t = i, \boldsymbol{Y}^{(t)})\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(i)) \\ &= \mathbb{P}(\boldsymbol{Y}^{(t-1)}|\boldsymbol{Z}_t(i))\mathbb{P}(\boldsymbol{Z}_t(i))\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(i)).\end{aligned}$$

Avec la proposition 11, il s'en suit que :

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{Y}^{(t-1)}|\boldsymbol{Z}_t(i))\mathbb{P}(\boldsymbol{Z}_t(i))\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(i)) \\
&= \mathbb{P}(\boldsymbol{Z}_t(i))\mathbb{P}\left(\boldsymbol{Y}^{(-t)}|\boldsymbol{Z}_t(i)\right) \\
&= \mathbb{P}\left(\boldsymbol{Y}^{(-t)}, \boldsymbol{Z}_t(i)\right) \\
&= \mathbb{P}\left(\boldsymbol{Y}_1^T, C_t = i\right).
\end{aligned}$$

□

La proposition 16 sera utilisée dans les propositions 17 et 19. De plus, la proposition 16 est définie pour la première fois à l'équation (3.10).

Proposition 17.

$$\boxed{\begin{array}{c} \textit{Pour } 0 \leq t \leq T \\ \dfrac{\alpha_t(i)\beta_t(i)}{\sum_i \alpha_t(i)\beta_t(i)} = \mathbb{P}(C_t = i|\boldsymbol{Y}^{(T)}) \end{array}} \tag{B.14a}$$

Démonstration. Avec l'aide de la proposition 16 :

$$\begin{aligned}
\frac{\alpha_t(i)\beta_t(i)}{\sum_i \alpha_t(i)\beta_t(i)} &= \frac{\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, C_t = i)}{\sum_i \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, C_t = i)} \\
&= \frac{\mathbb{P}(C_t = i|\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})}{\mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)})} \\
&= \mathbb{P}(C_t = i|\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}).
\end{aligned}$$

□

La proposition 17 est définie pour la première fois à l'équation (3.11).

Proposition 18.

$$\boxed{\begin{array}{c} \textit{Pour } 0 \leq t \leq T \\ \dfrac{\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}{\sum_{i,j\in\Omega^2} \alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)} = \mathbb{P}(C_{t-1} = i, C_t = j|\boldsymbol{Y}^{(T)}) \end{array}} \tag{B.15a}$$

Démonstration. Premièrement,

$$\frac{\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}{\sum_{i,j\in\Omega^2}\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)} = \frac{\mathbb{P}(C_{t-1}=i,\boldsymbol{Y}^{(t-1)})\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(j))}{\sum_{i,j\in\Omega^2}\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}$$
$$= \frac{\mathbb{P}(C_{t-1}=i,\boldsymbol{Y}^{(t-1)})\frac{1}{\mathbb{P}(\boldsymbol{Z}_{t-1})}\mathbb{P}(\boldsymbol{Z}_{t-1},\boldsymbol{Z}_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(j))}{\sum_{i,j\in\Omega^2}\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}.$$

En utilisant les propositions 12 et 13, il s'en suit que

$$\begin{aligned}
&\frac{\mathbb{P}(C_{t-1}=i,\boldsymbol{Y}^{(t-1)})\frac{1}{\mathbb{P}(\boldsymbol{Z}_{t-1})}\mathbb{P}(\boldsymbol{Z}_{t-1},\boldsymbol{Z}_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_t(j))}{\sum_{i,j\in\Omega^2}\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}\\
&= \frac{\mathbb{P}(C_{t-1}=i,\boldsymbol{Y}^{(t-1)})\frac{1}{\mathbb{P}(\boldsymbol{Z}_{t-1})}\mathbb{P}(\boldsymbol{Z}_{t-1},\boldsymbol{Z}_t)\mathbb{P}(\boldsymbol{Y}_{t+1}^T|\boldsymbol{Z}_{t-1}(i),\boldsymbol{Z}_t(j))}{\sum_{i,j\in\Omega^2}\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}\\
&= \frac{\mathbb{P}(C_{t-1}=i,\boldsymbol{Y}^{(t-1)})\frac{1}{\mathbb{P}(\boldsymbol{Z}_{t-1})}\mathbb{P}(\boldsymbol{Y}_{t+1}^T,\boldsymbol{Z}_{t-1}(i),\boldsymbol{Z}_t(j))}{\sum_{i,j\in\Omega^2}\alpha_{t-1}(i)\mathbb{P}(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})\beta_t(j)}\\
&= \frac{\mathbb{P}(\boldsymbol{Y}_1^T,C_{t-1},C_t)}{\sum_{i,j\in\Omega^2}\mathbb{P}(\boldsymbol{Y}_1^T,C_{t-1},C_t)}\\
&= \frac{\mathbb{P}(\boldsymbol{Y}_1^T,C_{t-1},C_t)}{\mathbb{P}(\boldsymbol{Y}_1^T)}\\
&= \frac{\mathbb{P}(C_{t-1},C_t|\boldsymbol{Y}_1^T)\mathbb{P}(\boldsymbol{Y}_1^T)}{\mathbb{P}(\boldsymbol{Y}_1^T)}\\
&= \mathbb{P}(C_{t-1},C_t|\boldsymbol{Y}_1^T).
\end{aligned}$$

□

La proposition 18 est définie pour la première fois à l'équation (3.12).

Proposition 19.

$$\boxed{\begin{array}{l}\textit{Pour } 0\leq t\leq T\\ \qquad\qquad \boldsymbol{\alpha}_t\boldsymbol{\beta}_t' = L_T\end{array}} \tag{B.16a}$$

Démonstration. Avec l'aide de la proposition 16 :

$$\begin{aligned}
\boldsymbol{\alpha}_t\boldsymbol{\beta}_t' &= \sum_i \alpha_t(i)\beta_t(i) \\
&= \sum_i \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}, C_t = i) \\
&= \sum_i \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)} | C_t = i)\mathbb{P}(C_t = i) \\
&= \mathbb{P}(\boldsymbol{Y}^{(T)} = \boldsymbol{y}^{(T)}) \\
&= L_T.
\end{aligned}$$

□

La proposition 19 est définie pour la première fois à l'équation (3.13).

ANNEXE C

BASE DE DONNÉES BRD - PRIX IPC

Tableau C.1

BRD - Actif requis

Année	Index IPC	Actif requis	Année	Index IPC	Actif requis
1980	82.4	$100,000,000	1999	166.6	$202,184,466
1981	90.9	$110,315,534	2000	172.2	$208,980,000
1982	96.5	$117,111,650	2001	177.1	$214,927,184
1983	99.6	$120,873,786	2002	179.9	$218,325,200
1984	103.9	$126,092,233	2003	184.0	$223,300,971
1985	107.6	$130,582,524	2004	188.9	$229,247,573
1986	109.6	$133,009,709	2005	195.3	$237,014,563
1987	113.6	$137,864,078	2006	201.6	$244,660,941
1988	118.3	$143,567,961	2007	207.342	$251,628,640
1989	124.0	$150,485,437	2008	215.303	$261,290,049
1990	130.7	$158,616,505	2009	214.537	$260,360,437
1991	136.2	$165,291,262	2010	218.056	$264,631,068
1992	140.3	$170,266,990	2011	224.939	$272,653,333
1993	144.5	$175,364,078	2012	229.594	$278,633,495
1994	148.2	$179,854,369	2013	232.957	$282,714,806
1995	152.4	$184,951,456	2014	236.736	$287,300,971
1996	156.9	$190,412,621	2015	237.017	$287,641,990
1997	160.5	$194,781,553	2016	240.007	$291,270,631
1998	163.0	$197,815,534	2017	245.120	$297,475,700

ANNEXE D

NOMBRE ANNUEL DE COMPAGNIES ACTIVES

Tableau D.1

Nombre annuel de compagnies actives

Année	Nombre de compagnies	Année	Nombre de compagnies
1979	2311	1998	4030
1980	2242	1999	4137
1981	2231	2000	4114
1982	2249	2001	3861
1983	2320	2002	3768
1984	2373	2003	3761
1985	2440	2004	3791
1986	2545	2005	3735
1987	2622	2006	3673
1988	2712	2007	3599
1989	2729	2008	3484
1990	2710	2009	3444
1991	2709	2010	3445
1992	2802	2011	3472
1993	3404	2012	3518
1994	3519	2013	3531
1995	3642	2014	3521
1996	3784	2015	3435
1997	3888	2016	3248

ANNEXE E

MODÈLES ANALYSÉS

Les 31 combinaisons possibles des 5 régresseurs sont présentées dans cette annexe. La première colonne représente le numéro du modèle et les autres colonnes représentent chacun des régresseurs. S'il y a un 1 dans une colonne, alors le régresseur est utilisé, sinon le régresseur n'est pas utilisé et la valeur sera de zéro.

# modèle	Chômage	*Spread*	*Fed funds*	IPC	S&P 500
1	0	0	0	0	1
2	0	0	0	1	0
3	0	0	0	1	1
4	0	0	1	0	0
5	0	0	1	0	1
6	0	0	1	1	0
7	0	0	1	1	1
8	0	1	0	0	0
9	0	1	0	0	1
10	0	1	0	1	0
11	0	1	0	1	1
12	0	1	1	0	0
13	0	1	1	0	1
14	0	1	1	1	0

# modèle	Chômage	*Spread*	*Fed funds*	IPC	S&P 500
15	0	1	1	1	1
16	1	0	0	0	0
17	1	0	0	0	1
18	1	0	0	1	0
19	1	0	0	1	1
20	1	0	1	0	0
21	1	0	1	0	1
22	1	0	1	1	0
23	1	0	1	1	1
24	1	1	0	0	0
25	1	1	0	0	1
26	1	1	0	1	0
27	1	1	0	1	1
28	1	1	1	0	0
29	1	1	1	0	1
30	1	1	1	1	0
31	1	1	1	1	1

RÉFÉRENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Artyomov, E., Rivenson, Y., Levi, G. et Yadid-Pecht, O. (2005). Morton (z) scan based real-time variable resolution cmos image sensor. *IEEE transactions on circuits and systems for video technology*, *15*(7), 947–952.

Baum, L. E. (1972). An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, *3*, 1–8.

Baum, L. E., Eagon, J. A. *et al.* (1967). An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bull. Amer. Math. Soc*, *73*(3), 360–363.

Baum, L. E. et Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Baum, L. E., Petrie, T., Soules, G. et Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, *41*(1), 164–171.

Bayes, M. et Price, M. (1763). An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs. *Philosophical Transactions (1683-1775)*, 370–418.

Benmiloud, B. et Pieczynski, W. (1995). Estimation des paramètres dans les chaînes de markov cachées et segmentation d'images. *TS. Traitement du signal*, *12*(5), 433–454.

BIS (2016). History of the basel committee. Récupéré de `http://www.bis.org/bcbs/history.htm`

Black, F. et Cox, J. C. (1976). Valuing corporate securities : Some effects of bond indenture provisions. *The Journal of Finance*, *31*(2), 351–367.

Black, F. et Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, *81*(3), 637–654.

Borio, C. (2014). The financial cycle and macroeconomics : What have we learnt ? *Journal of Banking & Finance*, *45*, 182–198.

Boucher, J.-P. et Hainault, D. (2013). Time series of correlated count data using multifractal process.

Boudaren, M. E. Y. (2014). *Modèles graphiques évidentiels*. (Thèse de doctorat). Evry, Institut national des télécommunications.

Bowers, N. L. N. L. *et al.* (1986). *Actuarial mathematics*. Numéro 517/A18.

Buyer, T. B. (2017). Mergers & acquisitions : Moody's to buy credit risk analyzer kmv for $210 million.

Byrne, C. et Eggermont, P. P. (2011). Em algorithms. In *Handbook of Mathematical Methods in Imaging* 271–344. Springer.

Calvet, L. E. et Fisher, A. (2008). *Multifractal volatility : theory, forecasting, and pricing*. Academic Press.

Cameron, A. C. et Trivedi, P. K. (2001). Essentials of count data regression. *A companion to theoretical econometrics*, *331*.

Cameron, A. C. et Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Cameron, A. C., Trivedi, P. K., Milne, F. et Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in australia. *The Review of economic studies*, *55*(1), 85–106.

Chorafas, D. N. (2004). *Economic capital allocation with Basel II : Cost, benefit and implementation procedures*. Butterworth-Heinemann.

Committee, B. *et al.* (2010). Group of governors and heads of supervision announces higher global minimum capital standards.

Compustat (2017). Base de données sur le nombre annuel de compagnies. Récupéré de `https://wrds-web.wharton.upenn.edu/wrds/`

CSFBI (2017). Creditrisk+ - a credit risk management framework. Récupéré de `http://www.csfb.com/institutional/research/assets/creditrisk.pdf`

Denuit, M., Maréchal, X., Pitrebois, S. et Walhin, J.-F. (2007). *Actuarial modelling of claim counts : Risk classification, credibility and bonus-malus systems*. John Wiley & Sons.

Derrode, S. et Pieczynski, W. (2004). Signal and image segmentation using pairwise markov chains. *IEEE Transactions on Signal Processing*, *52*(9), 2477–2489.

Dowd, K. et Rowe, D. (2004). The professional risk manager's international association, the prm handbook–iii.

Durrett, R. (2010). *Probability : theory and examples*. Cambridge university press.

Fisher, R. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, *41*, 155–160.

Fisher, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, *85*(4), 597–612.

FRED (2017a). Fred economic data. Récupéré de `https://fred.stlouisfed.org/`

FRED (2017b). Publications sur *Economic indicators*. Récupéré de `https://www.stlouisfed.org/Publications`

Frees, E. W. (2004). *Longitudinal and panel data : analysis and applications in the social sciences*. Cambridge University Press.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss*. sumtibus Frid. Perthes et IH Besser.

Giampieri, G., Davis, M. et Crowder, M. (2005). Analysis of default data using hidden markov models. *Quantitative Finance*, *5*(1), 27–34.

Griffiths, J. (1985). Table-driven algorithms for generating space-filling curves. *Computer-Aided Design*, *17*(1), 37–41.

Griliches, Z. (1990). *Patent statistics as economic indicators : a survey*. Rapport technique, National Bureau of Economic Research.

Guldberg, A. (1934). On discontinuous frequency functions of two variables. *Scandinavian Actuarial Journal*, *1934*(1), 89–117.

Hamdy, A. et Hussein, W. B. (2016). Credit risk assessment model based using principal component analysis and artificial neural network. Dans *MATEC Web of Conferences*, volume 76, p. 02039. EDP Sciences.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica : Journal of the Econometric Society*, 357–384.

Hilbert, D. (1891). Ueber die stetige abbildung einer line auf ein flächenstück. *Mathematische Annalen*, *38*(3), 459–460.

Hodrick, R. J. et Prescott, E. C. (1997). Postwar us business cycles : an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, *30*(1), 175–193.

Judd, K. L. (1998). *Numerical methods in economics*. MIT press.

Jung, R. C. et Liesenfeld, R. (2001). Estimating time series models for count data using efficient importance sampling. *AStA Advances in Statistical Analysis*, *4*(85), 387–407.

Jung, R. C. et Tremayne, A. (2011). Useful models for time series of counts or simply wrong ones ? *AStA Advances in Statistical Analysis*, *95*(1), 59–91.

Katti, S. et Rao, A. V. (1968). Handbook of the poisson distribution. *Technometrics*, *10*(2), 412–412.

Lagrange, L. (1788). Mécanique analytique, paris. *MJ Bertrand. Mallet-Bachelier.)[aPJHS]*.

Lahiri, K. et Moore, G. H. (1992). *Leading economic indicators : new approaches and forecasting records*. Cambridge University Press.

Lanchantin, P. (2006). *Chaınes de Markov triplets et segmentation non supervisée de signaux*. (Thèse de doctorat). PhD thesis, Institut National des Télécommunications, Evry, France.

Merton, R. C. (1974). On the pricing of corporate debt : The risk structure of interest rates. *The Journal of finance*, *29*(2), 449–470.

Moody's (2008). Moody's places lehman's a2 rating on review with direction uncertain. Récupéré de `https://www.moodys.com/research/Moodys-places-Lehmans-A2-rating-on-review-with-direction-uncertain--PR_162621`

Moody's (2017). Riskfrontier. Récupéré de `http://www.moodysanalytics.com/Products-and-Solutions/Enterprise-Risk-Solutions/Portfolio-Management/RiskFrontier`

Okun, A. M. (1963). *Potential GNP : its measurement and significance.* Yale University, Cowles Foundation for Research in Economics.

Paroli, R. et Spezia, L. (2000). *Something else about gaussian hidden Markov models and air pollution data.* Universita cattolica del Sacro Cuore, Istituto di statistica.

Peano, G. (1890). Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen, 36*(1), 157–160.

Pearson, K. (1900). X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50*(302), 157–175.

Pfanzagl, J. (1994). *Parametric statistical theory.* Walter de Gruyter.

Pieczynski, W. (1992). Statistical image segmentation. *Machine graphics and vision, 1*(1/2), 261–268.

Pieczynski, W. (2003). Pairwise markov chains. *IEEE Transactions on pattern analysis and machine intelligence, 25*(5), 634–639.

Pieczynski, W. (2013). Triplet markov chains and image segmentation. *Inverse problems in Vision and 3D Tomography*, 123–153.

Poisson, S. D. (1837). Probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés. *Paris, France : Bachelier, 1*, 1837.

Quandl (2017). The united states economic indicators. Récupéré de `https://www.quandl.com/collections/usa/usa-economy-data`

Rafi, S., Castella, M. et Pieczynski, W. (2011). Pairwise markov model applied to unsupervised image separation. Dans *SPPRA'11 : The Eighth IASTED International Conference on Signal Processing, Pattern Recognition, and Applications.* Acta Press.

Reuters, J. . (2017). Riskmetrics(tm) — technical document. Récupéré de `https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a`

Rose, N. L. (1990). Profitability and product quality : Economic determinants of airline safety performance. *Journal of Political Economy*, 944–964.

Schumpeter (2017). Life in the public eye. *The Economist.*

Schwarz, G. *et al.* (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

SEC (2017a). Form 10-k. Récupéré de `https://www.sec.gov/answers/form10k.htm`

SEC (2017b). What every investor should know ... corporate bankruptcy. Récupéré de `https://www.sec.gov/investor/pubs/bankrupt.htm`

Sullivan, A. (2003). Economics : Principles in action.

Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

UCLA (2017). A window on the world of big-case bankruptcy. Récupéré de `http://lopucki.law.ucla.edu/index.htm`

Welch, L. R. (2003). Hidden markov models and the baum-welch algorithm. *IEEE Information Theory Society Newsletter*, *53*(4), 10–13.

Wikipedia (2014). Bond credit rating. Récupéré de `http://en.wikipedia.org/wiki/Bond_credit_rating#cite_note-MRSD-3`

Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. *Journal of Business & Economic Statistics*, *13*(4), 467–474.

Woodcock, P. S. (2014). Notes de cours en econ 837 - 11 : Asymptotic properties of maximum likelihood estimators. Récupéré de `http://www.sfu.ca/~swoodcoc/teaching/sp2014/econ837/11.mle.pdf`

Ypma, T. J. (1995). Historical development of the newton-raphson method. *SIAM review*, *37*(4), 531–551.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, *75*(4), 621–629.

Zhao, J. Y. (2010). Hidden markov models with multiple observation processes. *arXiv preprint arXiv :1010.1042.*

Zucchini, W. et MacDonald, I. L. (2009). *Hidden Markov models for time series : an introduction using R*, volume 150. CRC press.