

Sur l'utilisation des modèles multi-états  
pour la mesure et la gestion des risques  
d'un contrat d'assurance

Annexes

Quentin GUIBERT

ISFA (Laboratoire SAF EA2429), Université de Lyon, Université Claude  
Bernard Lyon 1  
Autorité de Contrôle Prudentiel et de Résolution

7 Décembre 2015  
Soutenance de thèse - Lyon

Sous la direction de F. PLANCHET (ISFA)

## Lien entre probabilités et intensités de transition

- ▶ **Cas markovien** : les probabilités de transition sont solutions des équations *forward* et *backward* de Kolmogorov (matricielles)

$$\frac{\partial}{\partial t} \mathbf{p}(s, t) = \mathbf{p}(s, t) \boldsymbol{\mu}(t) \quad \text{et} \quad \frac{\partial}{\partial s} \mathbf{p}(s, t) = \mathbf{p}(s, t) \boldsymbol{\mu}(s) \quad \text{avec} \quad \mathbf{p}(s, s) = \mathbf{Id}.$$

- ▶ L'équation *forward* admet une unique solution s'exprimant sous la forme d'un produit intégral

$$\begin{aligned} \mathbf{p}(s, t) &= \lim_{\max_k |t_k - t_{k-1}| \rightarrow 0} \prod_{k=1}^n (\mathbf{Id} + \mathbf{A}(t_k) - \mathbf{A}(t_{k-1})) \\ &= \mathcal{P}_{\tau \in ]s, t]} (\mathbf{Id} + d\mathbf{A}(\tau)). \end{aligned}$$

avec  $0 < t_1 < \dots < t_K$  une partition de  $]s, t]$  et  $\mathbf{A}(t) = \left( \int_0^t \mu_{ij}(\tau) d\tau \right)_{i,j \in S}$  la matrice des fonctions **d'intensité cumulée**.

## Lien entre probabilités et intensités de transition

- ▶ Si les fonctions d'intensité cumulée sont des fonctions en escalier, le produit intégral devient un produit classique

$$\mathbf{p}(s, t) = \prod_{k=1}^K (\mathbf{Id} + \Delta \mathbf{A}(t_k)) \text{ avec } \Delta \mathbf{A}(t) = \mathbf{A}(t_k) - \mathbf{A}(t_{k-1}).$$

- ▶ **Cas homogène** : avec la condition initiale  $\mathbf{p}(0) = \mathbf{Id}$

$$\mathbf{p}(s, t) = \mathbf{p}(0, t - s) = \exp \{ \mathbf{A} \times (t - s) \}.$$

- ▶ **Cas homogène par morceaux** : intensité de transition  $\mathbf{A}_k$  pour le segment  $]t_{k-1}, t_k]$

$$\mathbf{p}(s, t) = \exp \{ \mathbf{A}_{[s]} \times (t_{[s]} - s) \} \times \prod_{\{k: s < t_k < t\}} \exp \{ \mathbf{A}_k \times (t_{k+1} \wedge t - t_k) \},$$

avec  $[s] = \min_k \{k : s < t_k\}$ .

## Cas Markov : inférence non-paramétrique pour données censurées

- ▶ **Cas Markov** : méthode d'inférence générale pour tout type de modèle multi-états.
- ▶  $C \equiv$  censure à droite indépendante et non informative .
- ▶ Fondée sur la théorie des processus de comptage (Andersen *et al.*, 1993)

$$N_{hj}(t) = \# \{0 \leq \tau \leq t : X_\tau = j, X_{\tau-} = h, 0 \leq \tau \leq C\} ,$$

$$L_h(t) = \mathbb{1}_{\{X_{\tau-} = h, 0 \leq t \leq C\}} \text{ and } N(t) = \sum_{h,j} N_{hj}(t) .$$

- ▶ Avec  $(\mathcal{F}_t)$  la filtration canonique engendrée par  $N$  et  $X_0$ , on a pour tout  $h \rightarrow j$

$$N_{hj}(t) - \int_0^t L_h(\tau) dA_{hj}(\tau) \stackrel{\text{si abs. continu}}{=} N_{hj} - \int_0^t L_h(\tau) \mu_{hj}(\tau) d\tau ,$$

sont des martingales.

## Cas Markov : inférence non-paramétrique pour données censurées

- ▶ Intensités de transition estimées à partir de l'estimateur de [Nelson-Aalen](#)

$$\widehat{A}_{hj}(t) = \int_0^t \frac{dN_{hj}(\tau)}{L_h(\tau)} = \sum_{\{k:t_k \leq t\}} \frac{d_{hj}(t_k)}{L_h(t_k)}.$$

- ▶ L'hétérogénéité de la population.  
Ex : modèle de Cox à hasard proportionnel

$$\mu_{hj}(t | \mathbf{Z}_{hj,i}, \boldsymbol{\theta}) = \mu_{0hj}(t) \exp\left(\boldsymbol{\theta}^\top \mathbf{Z}_{hj,i}\right).$$

- ▶ Les matrices de probabilités de transition  $\mathbf{p}$  sont obtenues au moyen de l'estimateur d'[Aalen-Johansen](#)

$$\widehat{\mathbf{p}}(s, t) = \mathcal{P}_{\tau \in ]s, t]} \left( \mathbf{Id} + d\widehat{\mathbf{A}}(\tau) \right),$$

avec  $\mathcal{P}$  l'opérateur produit intégral.

- ▶ Généralisation de l'estimateur de Kaplan-Meier (KM) pour des données de survie (Kaplan and Meier, 1958).

## Modèle semi-Markov

- ▶ Soit  $S_1 < S_2 < \dots < S_k < \dots$  les temps de saut du processus  $X$ .
- ▶ Soit  $J_k$  le processus discret donnant l'état occupé par  $X$  entre les dates  $S_k$  et  $S_{k+1}$ .
- ▶  $U_t = t - S_{N(t)} \equiv$  temps de séjour dans l'état courant.

### Définition

Si le processus discret  $(S_k, J_k)$  est markovien, il est appelé **processus de renouvellement markovien** et est caractérisé par une distribution initiale et un **noyau semi-markovien**

$$Q_{hj}(s, t) = \mathbb{P}(\Delta S_{k+1} \leq t, J_{k+1} = j \mid S_k = s, J_k = h).$$

$\Rightarrow (X_t, U_t)$  est Markov et le processus  $(X_t)$  est appelé processus semi-Markov selon sa filtration canonique  $(\mathcal{F}_t)$ .

- ▶  $p_{hj}(s, t, u, v) = \mathbb{P}(X_t = j, U_t \leq v \mid X_s = h, U_s = u) \equiv$  probabilité de transition.
- ▶  $\mu_{hj}(t, u) = \lim_{\Delta t \rightarrow 0} \frac{p_{hj}(t, t + \Delta t, u, \infty)}{\Delta t} \equiv$  intensité de transition.

## Inférence d'un processus semi-Markov homogène

- ▶ Processus homogène :  $Q_{hj}(s, t) = Q_{hj}(t)$ .
- ▶ Inférence non-paramétrique (paramétrique ne pose pas de problème).
- ▶ **Modèle sans boucle** : peut être estimé comme un modèle Markov → se produit souvent en science actuarielle.
- ▶ **Modèle avec boucles** : le noyau semi-markovien est estimé non-paramétriquement (Gill, 1980) par

$$\widehat{Q}_{hj}(t) = \int_0^t \left(1 - \widehat{H}_h(\tau)\right) \frac{dN_{hj}(\tau)}{L_h(\tau)}.$$

où  $H_h(u) = \mathbb{P}(\Delta S_{k+1} \leq u \mid J_k = h)$ , estimée par Kaplan-Meier.

- ▶  $N_{hj}(u)$  et  $L_h(u)$  dépendent du temps  $u$  passé dans l'état  $h$ .
- ▶ **Mais** les probabilités de transition  $\Psi_{hj}(t) = \mathbb{P}(X_t = j \mid X_0 = h)$  sont complexes à calculer (Spitoni *et al.*, 2012).

## Inférence pour modèles semi-Markov non-homogènes

### ► Cas semi-Markov non-homogène **sans boucle** :

- Généralement, une des deux variables temporelles est assimilée à une covariable ;
- Approche *splitting of state* (Haberman and Pitacco, 1998) ;
- Approche du praticien pour les modèles d'invalidité : estimation des probabilités de survie dans l'état d'invalidité, par exemple par Kaplan-Meier, en stratifiant par âge (entier) d'entrée ;
- Modèle semi-Markov de Cox (Andersen and Perme, 2008)

$$\mu_{hj}(t \mid \mathbf{Z}_{hj,i}, U_t, \boldsymbol{\theta}) = \mu_{0hj}(t) \exp \left( \theta_0 f(U_t) + \boldsymbol{\theta}^\top \mathbf{Z}_{hj,i} \right).$$

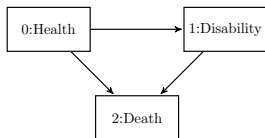
### ► Cas semi-Markov non-homogène **avec boucles** :

- Pas de cadre général pour l'estimation non-paramétrique ;
- Spécification de type Cox applicable (Dabrowska, 1995) ;
- Approches paramétriques où les intensités ou bien les noyaux sont écrits comme le produit de deux fonctions unidimensionnelles (Monteiro *et al.*, 2006; Mathieu *et al.*, 2007) .



## Approches fondées sur un calcul direct des probabilités

- ▶ Meira-Machado *et al.* (2006) → probabilité de transition pour un modèle *illness-death* acyclique sans l'hypothèse de Markov.
- ▶ Soit  $S$  la durée de vie dans l'état sain et  $T$  la durée de vie totale.
- ▶ Probabilités de transition vues comme des fonctionnelles sous la distribution jointe de  $(S, T)$  et estimées en utilisant des intégrales Kaplan-Meier.



$$p_{00}(s, t) = \frac{\mathbb{P}(S > t)}{\mathbb{P}(S > s)}$$

$$p_{01}(s, t) = \frac{\mathbb{P}(s < S \leq t < T)}{\mathbb{P}(S > s)} = \frac{\mathbb{E} \left[ \varphi_{st}^{(1)}(S, T) \right]}{\mathbb{P}(S > s)}$$

$$p_{11}(s, t) = \frac{\mathbb{P}(S \leq s, t < T)}{\mathbb{P}(S \leq s < T)} = \frac{\mathbb{E} \left[ \varphi_{st}^{(2)}(S, T) \right]}{\mathbb{E} \left[ \varphi_{ss}^{(2)}(S, T) \right]}$$

## Assurance dépendance

- ▶ Hausse de dépendance de santé pour les personnes âgées ces dernières décennies.
- ▶ Grande incertitude sur la tendance mais devrait continuer à croître ...
- ▶ En France, produits d'assurance en complément de la couverture de la sécurité sociale.
- ▶ Le niveau des paiements dépend du degré de dépendance (fixé au contrat).
- ▶ Plusieurs types de grilles définissent la sévérité de la dépendance. Elles s'appuient généralement sur plusieurs AVQ (se laver, manger, s'habiller, se déplacer, etc.).
- ▶ Large gamme de produits (garanties annuelles ou pluriannuelles, individuelles ou collectives).
- ▶ Assortis de nombreuses clauses.

## Données d'assurance dépendance

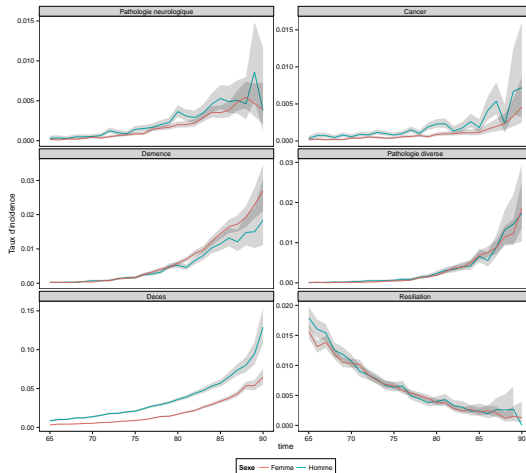
- ▶ Base de données d'un assureur français avec **observations individuelles**
- ▶ Dépendance lourde
- ▶  $\simeq$  210 000 individus observés sur la période 1998-2010 (exposition majoritairement féminine), dont environ 70 % sort par **censure** (concerne surtout les âges au-delà de 90 ans).
- ▶ Entrée en dépendance décomposée par **cause** (4 types de pathologies avec des périodes de franchise différentes).

### Limites des données

- ▶ Règles de codage des pathologies (évolutions, poly-pathologiques).
- ▶ Certains décès parfois considérés comme des résiliations.
- ▶ Forte censure aux grands âges.
- ▶ Peu de covariables.

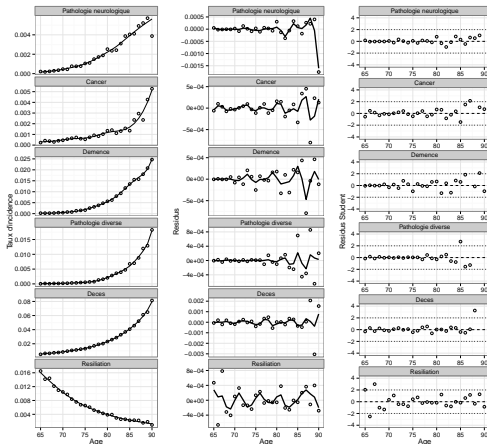
## Taux d'incidence bruts

- Taux d'incidence bruts par cause en fonction du sexe.



## Taux d'incidence lissés

- Taux d'incidence lissés, résidus et résidus de Student par cause.



- Lissage des taux bruts par la technique de **Whittaker-Henderson**.
- Pondération en fonction de l'exposition  $E_t$

$$\frac{E_t}{\max(E_t)}$$

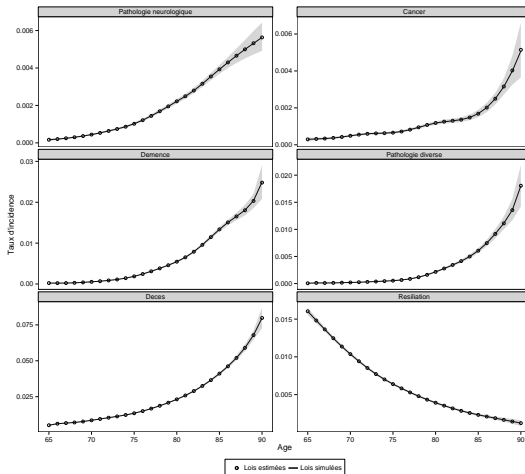
- Paramètres de lissage sélectionnés avec l'analyse des résidus et *via* les critères :
  - CV et GCV ;
  - AIC et AICC.

## Mesure du risque d'estimation

- ▶ La construction des lois induit deux sources de risques :
  - ▶ fluctuation d'échantillonnage (estimation des taux bruts) ;
  - ▶ erreur produite par le lissage.
- ▶ Le risque d'estimation est mesuré en utilisant une méthode de *bootstrap* avec 1 000 rééchantillonnages :
  - 1 Échantillonner avec remplacement des observations  $(Y_i, \gamma_i, \gamma_i V_i)$  ;
  - 2 Calculer les taux d'incidence bruts avec la population rééchantillonnée ;
  - 3 Lisser des taux bruts *via* la technique de Whittaker-Henderson ;
  - 4 Répéter les trois étapes précédentes.

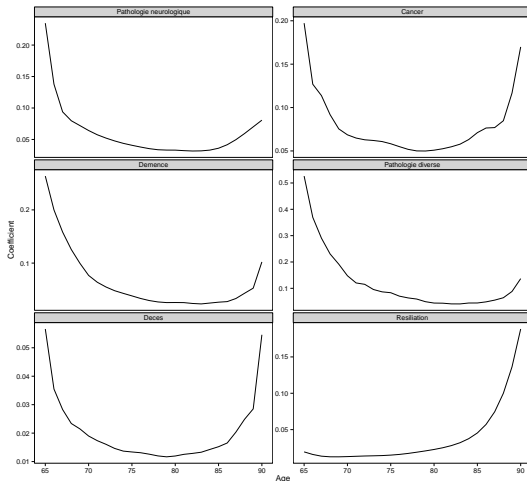
## Mesure du risque d'estimation

- Taux d'incidence et intervalles de confiance à 95 % simulés.



## Mesure du risque d'estimation

- Coefficient de dispersion par âge et par cause



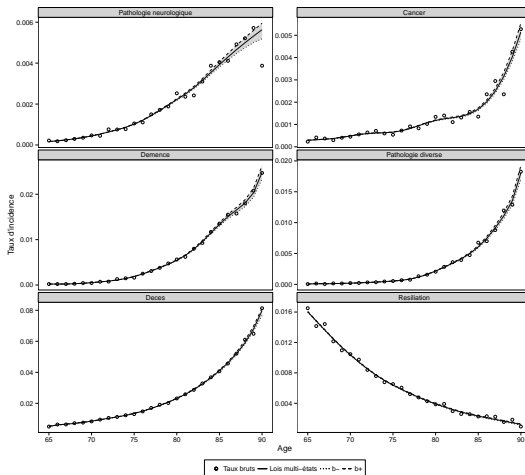


## Lissage par une approche *Generalized Additive Model*

- ▶ Approche **Poisson-GAM**
  - ▶ Hypothèse de **constance par morceaux** des intensités de transition et régression selon les âges ;
  - ▶ Choix d'une méthode de lissage (ex : splines cubiques) ;
  - ▶ Inférence par algorithme de *backfitting* ;
  - ▶ Mesure de la qualité du modèle et résidus ;
  - ▶ Possibilité de contrôle des intensités lissées par rapport aux intensités obtenues avec l'estimateur de Nelson-Aalen ;
  - ▶ Calcul des probabilités de transition (taux d'incidence).
- ▶ Possibilité de travailler directement avec les taux d'incidence  
→ Modèle de régression multinomiale *logit* GAM.

## Comparaison avec l'approche marginale

- Comparaison entre les taux lissés selon les deux approches. Représentation des deux bornes.



## Inférence des probabilité de transition ?

- ▶ La littérature s'appuie majoritairement sur des données nationales agrégées.  
⇒ Inférence *via* des modèles de type **GLM Poisson** (CMIR12, 1991).  
Exemple :

$$\eta \left( \mathbb{E} \left[ \frac{d_{hj}(x)}{e_h(x)} \right] \right) = a_k x^k + a_{k-1} x^{k-1} + \dots + a_0.$$

- ▶ Praticiens recourent à des techniques marginales, dérivées de l'analyse de survie.
- ▶ Recours à l'hypothèse de **Markov**.  
⇒ Mais clairement non respectée dans les états d'invalidité.

### Estimation non-paramétrique

- ▶ Données individuelles (longitudinales) avec censure indépendante à droite (et troncature à gauche).
- ▶ Pas d'estimateur des probabilités de transition général si l'hypothèse de Markov n'est pas satisfaite.

## Évaluation des engagements

- ▶ Évaluer des provisions techniques (ou des primes) nécessite de projeter les flux futurs d'un contrat.
- ▶ Exemple : contrat d'assurance de personnes (simplifié → pas de prestations lors d'une transition)
  - ▶  $B_h(t, u) \equiv$  primes et annuités (continues, prévisibles) versées dans l'état  $h$  en date  $t$  et durée de séjour  $u$  (flux net).
  - ▶  $v(s, t) \equiv$  coefficient d'actualisation pour la période  $[s, t]$ .

### Provisions techniques dans un état $h$

#### Formule générique cas semi-Markov

$$V_h(t, u) = \sum_j \int_t^\infty \int_t^{u+\tau-t} v(t, \tau) p_{hj}(t, \tau, u, dv) dB_j(\tau, v).$$

## Évaluation des engagements

- ▶ Évaluer des provisions techniques (ou des primes) nécessite de projeter les flux futurs d'un contrat.
- ▶ Exemple : contrat d'assurance de personnes (simplifié → pas de prestations lors d'une transition)
  - ▶  $B_h(t, u) \equiv$  primes et annuités (continues, prévisibles) versées dans l'état  $h$  en date  $t$  et durée de séjour  $u$  (flux net).
  - ▶  $v(s, t) \equiv$  coefficient d'actualisation pour la période  $[s, t]$ .

qui se réécrit ...

### Provisions techniques dans un état $h$

$$\begin{aligned}
 V_h(t, u) &= \int_t^{+\infty} v(t, \tau) (1 - \bar{p}_{hh}(t, \tau, u)) dB_h(\tau, u + \tau - v) \\
 &+ \sum_{j \neq h} \int_t^{+\infty} v(t, \tau) (1 - \bar{p}_{hh}(t, \tau, u)) \mu_{hj}(\tau, u + \tau - t) V_j(\tau, 0) d\tau.
 \end{aligned}$$

## Évaluation des engagements

- ▶ Évaluer des provisions techniques (ou des primes) nécessite de projeter les flux futurs d'un contrat.
- ▶ Exemple : contrat d'assurance de personnes (simplifié → pas de prestations lors d'une transition)
  - ▶  $B_h(t, u) \equiv$  primes et annuités (continues, prévisibles) versées dans l'état  $h$  en date  $t$  et durée de séjour  $u$  (flux net).
  - ▶  $v(s, t) \equiv$  coefficient d'actualisation pour la période  $[s, t]$ .

- ▶ Discrétisation de ces équations dans la pratique.
- ▶ Clauses plus contraignantes (franchise, carence, temporaire, etc.)  
⇒ complexifient ces expressions.  
Exemple : période de franchise  $f$  après l'entrée dans l'état  $j$

$$\mathbb{P}(X_t = j, f < U_t \leq v \mid (X_s, U_s) = (h, u)).$$

- ▶ Comment estimer ces quantités d'intérêt ?

## Modèles à risques concurrents

- ▶ L'estimateur d'Aalen-Johansen pour la **fonction d'incidence cumulée** (CIF) relative au modèle à risques concurrents  $(T, V)$  est

$$F^{(v)}(t) = \mathbb{P}(T \leq t, V = v).$$

- ▶  $C \equiv$  censure à droite indépendante

$$Y = \min(T, C) \text{ et } \gamma = \mathbb{1}_{\{T \leq C\}}.$$

### Estimateur non-paramétrique de la CIF

- ▶  $(Y_i, \gamma_i, \gamma_i V_i)_{1 \leq i \leq n}$  observations i.i.d.
- ▶ En ordonnant les valeurs de  $Y$ , l'estimateur s'exprime

$$\widehat{F}_n^{(v)}(y) = \sum_{i=1}^n \widetilde{W}_{in} J_{[i:n]}^{(v)} \mathbb{1}_{\{Y_{i:n} \leq y\}}, \quad \widetilde{W}_{in} = \frac{\delta_{[i:n]}}{n - i + 1} \prod_{j=1}^{i-1} \left( \frac{n - j}{n - j + 1} \right)^{\delta_{[j:n]}}.$$

- ▶  $\widetilde{W}_{in} \equiv$  poids de Kaplan-Meier et  $J_i^{(v)} = \mathbb{1}_{\{V_i = v\}}$ .
- ▶  $\widehat{F}_n^{(v)}(\cdot)$  converge p.s. vers  $F^{(v)}(\cdot)$  et est asymptotiquement normal.

## Techniques de *bootstrap* pour l'estimateur de la CIF

- ▶ Méthodologie de Efron (Efron, 1979, 1981)  
→ rééchantillonnage pour obtenir la distribution de  $(Y_i, \gamma_i, \gamma_i V_i)$ .
- ▶ *Wild bootstrap* (Beyersmann *et al.*, 2013; Dobler and Pauly, 2014)

$$W_n(t) = \sqrt{n} \left( \widehat{F}^{(v)}(t) - F^{(v)}(t) \right).$$

Admet une représentation martingale. Exemple avec deux causes

$$\widehat{W}_n(t) = \sqrt{n} \sum_{i=1}^n \left\{ \int_0^t \frac{(1 - \widehat{F}^{(2)}(\tau)) G_{01,i} dN_{01,i}(\tau)}{L(\tau)} + \frac{\widehat{F}^{(1)}(\tau) G_{02,i} dN_{02,i}(\tau)}{L(\tau)} - \widehat{F}^{(1)}(t) \int_0^t \frac{G_{01,i} dN_{01,i}(\tau) + G_{02,i} dN_{02,i}(\tau)}{L(\tau)} \right\};$$

avec  $G_{01,i}$  et  $G_{02,i}$  des variables de moyenne 0 et de variance 1 (Ex : Poisson, normale).



## Analyse des hypothèses

### Hypothèse 3

$$\int \frac{\varphi(S, T)^2 \delta}{(1 - G(T))^2} d\mathbb{P} < \infty.$$

- ▶ S'interprète comme l'existence d'un moment d'ordre 2.
- ▶ Hypothèse minimale mais pas forcément vérifiable.
- ▶ La censure à droite ne doit pas occulter trop d'information.

### Hypothèse 4

$$\int |\varphi(S, T)| \sqrt{C_0(T)} \mathbb{1}_{\{T < \tau_Z\}} d\mathbb{P} < \infty.$$

- ▶ Hypothèse purement technique reprise de Stute (1995) (argument de tension).
- ▶  $C_0(t) \leq \frac{1}{1 - M(x)}$ .
- ▶ Akritas (2000) s'en affranchit pour les intégrales KM.

## Analyse des hypothèses

$$S_{\infty}^{(v)}(\varphi) = \int \mathbb{1}_{\{t < \tau_Z\}} \varphi(s, t, \mathbf{u}) F_0^{(v)}(ds, dt, d\mathbf{u}).$$

- ▶ Même configuration que Meira-Machado *et al.* (2006).
- ▶ Sélectionne la population telle que  $\mathbb{1}_{\{T < \tau_Z\}}$ .
- ▶ de Uña-Álvarez and Meira-Machado (2015) optimisent leurs résultats précédents en réarrangeant les probabilités de transition. Par exemple,

$$p_{01}(s, t) = 1 - p_{00}(s, t) - p_{02}(s, t) \quad \text{à la Pepe (1991).}$$

- ▶ Estimateur consistant si  $\mathbb{1}_{\{t < \tau_Y\}}$ .
- ▶ Ajustement des poids en remplaçant  $n$  par  $n_s$  la taille de population en bonne santé en  $s$ .

## Troncature à gauche

- ▶  $L$  troncature à gauche et  $C$  censure à droite.

### Intégrale selon l'estimateur d'Aalen-Johansen

$$\widehat{S}_n^{(v)}(\varphi) = \sum_{i=1}^n \widetilde{W}_{in}^{(v)} \varphi(Y_{[i:n]}, Z_{i:n}), \quad 0 \leq s \leq t \leq \tau_Z,$$

avec

$$\widetilde{W}_{in}^{(v)} = \frac{\delta_{[i:n]} J_{[i:n]}^{(v)}}{nC_n(Z_{i:n})} \prod_{j=1}^{i-1} \left( 1 - \frac{1}{nC_n(Z_{i:n})} \right)^{\delta_{[j:n]}},$$

où  $C_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}_{L_i \leq x \leq Z_i}$ .

- ▶ Hypothèses :  $(C, L)$  indépendant de  $(S, T, V)$  et  $C$  indépendant de  $L$ .
- ▶ Résultats de convergence se démontrent en suivant l'approche de Sánchez-Sellero *et al.* (2005).

## Estimateurs des probabilités de transition

$$\hat{p}_{0e}(s, t, \eta) = \frac{\widehat{S}_n^{(e, C_e)}(\varphi_{s, t, \eta}^{(1)})}{1 - \widehat{H}_n(s)}, \text{ avec } \varphi_{s, t, \eta}^{(1)}(x, y) = \mathbb{1}_{\{s < x \leq t - \eta, y > t\}}.$$

$$\hat{p}_{ee}(s, t) = \frac{\widehat{S}_n^{(e, C_e)}(\varphi_{s, t}^{(2)})}{\widehat{S}_n^{(e, C_e)}(\varphi_{s, s}^{(2)})}, \text{ avec } \varphi_{s, t}^{(2)}(x, y) = \mathbb{1}_{\{x \leq s, y > t\}}.$$

$$\hat{p}_{ed}(s, \eta, \zeta) = \frac{\widehat{S}_n^{(e, d)}(\varphi_{s, \zeta}^{(3)})}{\widehat{S}_n^{(e, C_e)}(\varphi_{s, \eta}^{(4)})}, \text{ avec } \varphi_{s, \zeta}^{(3)}(x, y) = \mathbb{1}_{\{s < x \leq t, \eta < y - x \leq \zeta\}}.$$

$$\varphi_{s, \eta}^{(4)}(x, y) = \mathbb{1}_{\{s < x \leq t, \eta < y - x\}} \text{ et } \widehat{H}_n \text{ l'estimateur KM de la f.d.r de } S.$$

⇒ Ces estimateurs généralisent ceux de Meira-Machado *et al.* (2006).

## Variance asymptotique

- ▶ Rappel du résultats de type TCL :  $\sqrt{n} \left\{ \widehat{S}_n(\varphi) - S(\varphi) \right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\varphi))$ .

- ▶ **Variance asymptotique**

$$\sigma^{(v)}(\varphi) = \mathbb{V}\text{ar} \left( \frac{\varphi(Y, Z) \delta J^{(v)}}{1 - G(Z)} + \lambda_1^{(v)}(Z)(1 - \delta) - \lambda_2^{(v)}(Z) \right) = \mathbb{V}\text{ar}(\psi(\varphi)).$$

- ▶ Un estimateur naturel est obtenu en remplaçant  $G$ ,  $\lambda_1^{(v)}$  et  $\lambda_2^{(v)}$  par leur estimateur non-paramétrique, mais pas de garantie de consistance si  $\varphi$  n'est pas à support compact.
- ▶ Les probabilités de transition s'expriment sous la forme

$$p(\varphi_1, \varphi_2) = \frac{S(\varphi_1)}{S(\varphi_2)}.$$

- ▶ Leur variance asymptotique est (méthode Delta)

$$\begin{aligned} \sigma(p(\varphi_1, \varphi_2)) &= \frac{\mathbb{V}\text{ar}(\psi(\varphi_1))}{S(\varphi_2)^2} - 2\text{Cov}(\psi(\varphi_1), \psi(\varphi_2)) \frac{S(\varphi_1)}{S(\varphi_2)^3} \\ &\quad + \frac{\mathbb{V}\text{ar}(\psi(\varphi_2)) S(\varphi_1)^2}{S(\varphi_2)^4} \end{aligned}$$

## Mesures d'association

- ▶ On introduit une **mesure d'association locale** s'appuyant sur le *cross odds ratio*

$$\pi_0^{(e,d)}(s, t) = \frac{\text{odds}(T \leq t, V_2 = d \mid S \leq s, V_1 = e)}{\text{odds}(T \leq t, V_2 = d \mid V_1 = e)},$$

$$\text{où } \text{odds}(A) = \frac{\mathbb{P}(A)}{1 - \mathbb{P}(A)}.$$

- ▶ Mesure la dépendance entre la durée de vie en bonne santé et la durée de vie totale par cause.
- ▶ **Estimateur non-paramétrique**

$$\widehat{\pi}_{0n}^{(e,d)}(s, t) = \frac{\widehat{F}_{0n}^{(e,d)}(s, t)}{\frac{\widehat{H}_{0n}^{(e)}(s) - \widehat{F}_{0n}^{(e,d)}(s, t)}{\widehat{F}_n^{(e,d)}(t)}},$$

$$\frac{\widehat{H}_{0n}^{(e)}(\infty) - \widehat{F}_n^{(e,d)}(t)}$$

où  $\widehat{H}_{0n}^{(e)}$  est l'estimateur de la CIF associée à  $S$  pour la cause  $V_1 = e$  et  $\widehat{F}_n^{(e,d)}$  est celui de  $T$  pour la cause  $V = (e, d)$ .

## Développements sous R

- ▶ Package **mstate** permet de calculer les intensités et probabilités de transition d'un modèle multi-états markovien (intervalles de confiance et modèle de Cox).
- ▶ Existence d'autres bibliothèques comme **etm**, **SemiMarkov** (modèles paramétriques), **p3state.msm** (méthode non-paramétrique pour les modèles à 3 états).
- ▶ Estimateurs **non-paramétriques** des probabilités de transition :
  - 1 Transformer un jeu de données pour le faire correspondre avec la structure comprenant 2 modèles à risques concurrents ;
  - 2 Calcul des poids d'Aalen-Johansen pour ces modèles à risques concurrents ;
  - 3 Évaluation individuelle de chaque probabilités de transition ;  
→ Très rapide.
  - 4 Estimation des intervalles de confiance par *bootstrap*.  
→ Très consommateur de ressources sur les étapes 1 et 2.  
→ Plusieurs jours de calcul avec 500 rééchantillonnages.

## Développements sous R

- ▶ Package **mstate** permet de calculer les intensités et probabilités de transition d'un modèle multi-états markovien (intervalles de confiance et modèle de Cox).
- ▶ Existence d'autres bibliothèques comme *etm*, *SemiMarkov* (modèles paramétriques), *p3state.msm* (méthode non-paramétrique pour les modèles à 3 états).
- ▶ Estimateurs **non-paramétriques** des probabilités de transition :
  - 1 Transformer un jeu de données pour le faire correspondre avec la structure comprenant 2 modèles à risques concurrents ;
  - 2 Calcul des poids d'Aalen-Johansen pour ces modèles à risques concurrents ;
  - 3 Évaluation individuelle de chaque probabilités de transition ;  
→ Très rapide.
  - 4 Estimation des intervalles de confiance par *bootstrap*.  
→ Très consommateur de ressources sur les étapes 1 et 2.  
→ Plusieurs jours de calcul avec 500 rééchantillonnages.



## Développements sous R

- ▶ Package **mstate** permet de calculer les intensités et probabilités de transition d'un modèle multi-états markovien (intervalles de confiance et modèle de Cox).
- ▶ Existence d'autres bibliothèques comme **etm**, **SemiMarkov** (modèles paramétriques), **p3state.msm** (méthode non-paramétrique pour les modèles à 3 états).
- ▶ Estimateurs **non-paramétriques** des probabilités de transition :
  - 1 Transformer un jeu de données pour le faire correspondre avec la structure comprenant 2 modèles à risques concurrents ;
  - 2 Calcul des poids d'Aalen-Johansen pour ces modèles à risques concurrents ;
  - 3 Évaluation individuelle de chaque probabilités de transition ;  
→ Très rapide.
  - 4 Estimation des intervalles de confiance par *bootstrap*.  
→ Très consommateur de ressources sur les étapes 1 et 2.  
→ Plusieurs jours de calcul avec 500 rééchantillonnages.

## Développements sous R

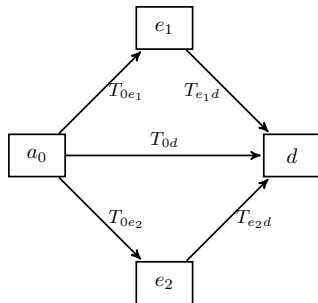
- ▶ Package `mstate` permet de calculer les intensités et probabilités de transition d'un modèle multi-états markovien (intervalles de confiance et modèle de Cox).
- ▶ Existence d'autres bibliothèques comme `etm`, `SemiMarkov` (modèles paramétriques), `p3state.msm` (méthode non-paramétrique pour les modèles à 3 états).
- ▶ Estimateurs **non-paramétriques** des probabilités de transition :
  - 1 Transformer un jeu de données pour le faire correspondre avec la structure comprenant 2 modèles à risques concurrents ;
  - 2 Calcul des poids d'Aalen-Johansen pour ces modèles à risques concurrents ;
  - 3 Évaluation individuelle de chaque probabilités de transition ;  
→ Très rapide.
  - 4 Estimation des intervalles de confiance par *bootstrap*.  
→ Très consommateur de ressources sur les étapes 1 et 2.  
→ Plusieurs jours de calcul avec 500 rééchantillonnages.

## Développements sous R

- ▶ Package **mstate** permet de calculer les intensités et probabilités de transition d'un modèle multi-états markovien (intervalles de confiance et modèle de Cox).
- ▶ Existence d'autres bibliothèques comme *etm*, *SemiMarkov* (modèles paramétriques), *p3state.msm* (méthode non-paramétrique pour les modèles à 3 états).
- ▶ Estimateurs **non-paramétriques** des probabilités de transition :
  - 1 Transformer un jeu de données pour le faire correspondre avec la structure comprenant 2 modèles à risques concurrents ;
  - 2 Calcul des poids d'Aalen-Johansen pour ces modèles à risques concurrents ;
  - 3 Évaluation individuelle de chaque probabilités de transition ;  
→ Très rapide.
  - 4 Estimation des intervalles de confiance par *bootstrap*.  
→ Très consommateur de ressources sur les étapes 1 et 2.  
→ Plusieurs jours de calcul avec 500 rééchantillonnages.

## Mesure de la performance : approche par simulation

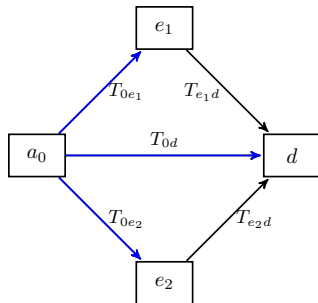
S'appuie sur l'algorithme proposé par Rotolo *et al.* (2013).



- 1 Simuler les **durées de vie latentes**  $T_{0e}$ ,  $T_{ed}$  et  $T_{0d}$  pour  $e = e_1, e_2$ , supposées suivre des lois Wei  $(\lambda, \rho)$ .
- 2 Lier à partir d'une **copule de Clayton**  $\mathcal{C}_{\theta_0}$  les durées issues de  $a_0$ .
- 3 Pour chaque  $e$ , définir une copule de Clayton  $\mathcal{C}_{\theta_e}$  pour lier  $T_{0e}$  et ses "enfants"  $T_{ed}$ .
- 4 Simuler trois variables de **censure** (pour chaque nœud) selon une distribution exponentielle Wei  $(\lambda, 1)$ .

## Mesure de la performance : approche par simulation

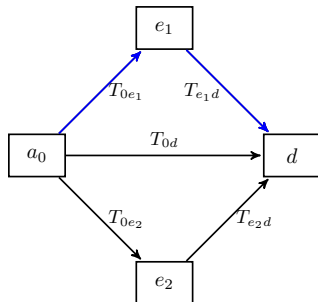
S'appuie sur l'algorithme proposé par Rotolo *et al.* (2013).



- 1 Simuler les **durées de vie latentes**  $T_{0e}$ ,  $T_{ed}$  et  $T_{0d}$  pour  $e = e_1, e_2$ , supposées suivre des lois Wei  $(\lambda, \rho)$ .
- 2 Lier à partir d'une **copule de Clayton**  $\mathfrak{C}_{\theta_0}$  les durées issues de  $a_0$ .
- 3 Pour chaque  $e$ , définir une copule de Clayton  $\mathfrak{C}_{\theta_e}$  pour lier  $T_{0e}$  et ses "enfants"  $T_{ed}$ .
- 4 Simuler trois variables de **censure** (pour chaque nœud) selon une distribution exponentielle Wei  $(\lambda, 1)$ .

## Mesure de la performance : approche par simulation

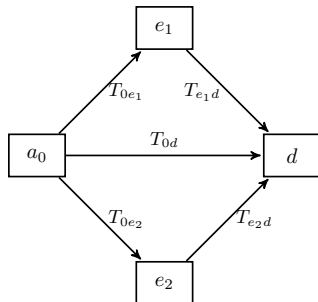
S'appuie sur l'algorithme proposé par Rotolo *et al.* (2013).



- 1 Simuler les **durées de vie latentes**  $T_{0e}$ ,  $T_{ed}$  et  $T_{0d}$  pour  $e = e_1, e_2$ , supposées suivre des lois Wei  $(\lambda, \rho)$ .
- 2 Lier à partir d'une **copule de Clayton**  $\mathfrak{C}_{\theta_0}$  les durées issues de  $a_0$ .
- 3 Pour chaque  $e$ , définir une copule de Clayton  $\mathfrak{C}_{\theta_e}$  pour lier  $T_{0e}$  et ses "enfants"  $T_{ed}$ .
- 4 Simuler trois variables de **censure** (pour chaque nœud) selon une distribution exponentielle Wei  $(\lambda, 1)$ .

## Mesure de la performance : approche par simulation

S'appuie sur l'algorithme proposé par Rotolo *et al.* (2013).



- 1 Simuler les **durées de vie latentes**  $T_{0e}$ ,  $T_{ed}$  et  $T_{0d}$  pour  $e = e_1, e_2$ , supposées suivre des lois Wei  $(\lambda, \rho)$ .
- 2 Lier à partir d'une **copule de Clayton**  $\mathfrak{C}_{\theta_0}$  les durées issues de  $a_0$ .
- 3 Pour chaque  $e$ , définir une copule de Clayton  $\mathfrak{C}_{\theta_e}$  pour lier  $T_{0e}$  et ses "enfants"  $T_{ed}$ .
- 4 Simuler trois variables de **censure** (pour chaque nœud) selon une distribution exponentielle Wei  $(\lambda, 1)$ .

## Mesure de la performance

- ▶ Mesure de la performance sur  $\widehat{p}_{0e}(s, t, 0)$  and  $\widehat{p}_{ee}(s, t)$  *versus* estimateur d'Aalen-Johansen classique.
- ▶ Calcul du biais intégré, de la variance intégrée et de l'erreur quadratique intégrée  $[s, s + 50]$ .

Paramètres	$T_{0e_1}$	$T_{0e_2}$	$T_{0d}$	$C_0$	$T_{e_1d}$	$C_1$	$T_{e_2d}$	$C_2$
<i>Scénario de censure modérée</i>								
Échelle $\lambda$	25	15	40	35	2.5	10	5	15
Forme $\rho$	0.9	0.8	1.1	1	0.5	1	0.8	1
<i>Scénario de censure moyenne</i>								
Échelle $\lambda$	25	15	40	27	2.5	3	5	7
Forme $\rho$	0.9	0.8	1.1	1	0.5	1	0.8	1

- ▶ Scénario de censure modérée :  $S$  censurée à 21 % et  $T$  à 34 %.
- ▶ Scénario de censure moyenne :  $S$  censurée à 24 % et  $T$  à 47 %.

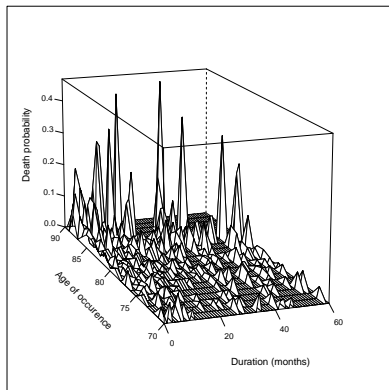


## Mesure de la performance

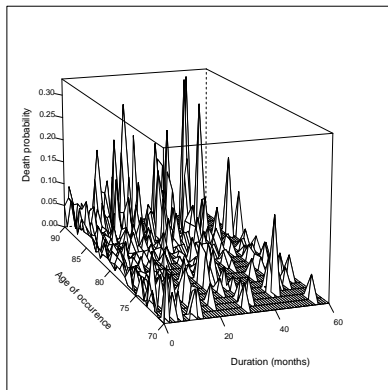
- ▶ Trois échantillons de taille  $n = 200, 400, 800$  et répétés  $K = 1000$ .
- ▶ Évaluation aux dates  $s = 5, s = 10$  et  $s = 15$ .
- ▶  $p_{ee}(s, t)$  : Sensiblement moins biaisé que l'estimateur d'AJ (systématique) mais plus incertain.
- ▶  $p_{0e}(s, t)$  :
  - ▶ Performances plus complexes à analyser ;
  - ▶ Varie selon l'intensité de la censure et la force de la dépendance entre la durée de vie en bonne santé et la durée de maintien ;
  - ▶ Localement meilleures performances lorsque la dépendance est importante.

## Application à la dépendance

- Estimation de la **surface de taux de décès mensuels** pour chaque état de dépendance → qualité faible pour certaines zones par manque d'observations.



(a)  $e_1$ -Pathologie neurologique



(b)  $e_2$ -Pathologie diverse

## Résultats : mesure d'association

- ▶ **Mesure d'association** entre la durée en bonne santé et la durée de vie totale à partir des *cross-odds ratios*.
- ▶ Intervalle de confiance à 95 % (entre parenthèses), calculé à partir de 500 rééchantillonnages.
- ▶ **Exemple : Démence**

$s$	1 mois	6 mois	12 mois	18 mois	24 mois
70	4.45 (4.226, 4.673)	4.469 (4.241, 4.698)	4.403 (4.165, 4.641)	4.282 (4.027, 4.537)	4.174 (3.907, 4.442)
75	3.396 (3.27, 3.522)	3.333 (3.199, 3.467)	3.258 (3.117, 3.4)	3.115 (2.969, 3.26)	2.994 (2.845, 3.142)
80	2.333 (2.233, 2.433)	2.272 (2.168, 2.376)	2.193 (2.082, 2.304)	2.102 (1.989, 2.214)	2.007 (1.889, 2.125)
85	1.487 (1.38, 1.594)	1.434 (1.323, 1.545)	1.379 (1.264, 1.493)	1.313 (1.192, 1.434)	1.234 (1.107, 1.36)
90	0.885 (0.737, 1.033)	0.876 (0.718, 1.034)	0.826 (0.651, 1.002)	0.775 (0.583, 0.968)	0.751 (0.53, 0.973)

## Assurance crédit

- ▶ Difficile d'envisager une dimension prospective à partir de ces données dépendance et solutions limitées pour la gestion dynamique des risques.
  - ▶ Conditions de crédit décrites par des [matrices de migration](#).
  - ▶ Dépendent de facteurs financiers et macroéconomiques.
- ▶ Comment faire évoluer dynamiquement les matrices de migration ?
  - ▶ Comment l'assureur peut-il ajuster son exposition en cas d'aggravation du risque de certaines polices suite à certains chocs des facteurs macros (*stress tests*) ?

## Spécifications économétriques

- ▶ Les *rating* sont naturellement ordonnés

$$\mathbb{P}(X_k(t) \leq j \mid X_k(t-1) = h, \mathbf{Z}(t)) = g\left(\mu_{hj,k} + \boldsymbol{\theta}_{hj,k}^\top \tilde{\mathbf{Z}}(t)\right).$$

- ▶  $g : \mathbb{R} \rightarrow [0, 1]$  une fonction lien et les  $\mu_{hj}$  représentent les coefficients de seuil d'un *rating* à l'autre.
- ▶ Les probabilités de transitions s'obtiennent par

$$\begin{cases} p_{h1,k}(t) = g\left(\mu_{h1,k} + \boldsymbol{\theta}_{h1,k}^\top \mathbf{Z}(t)\right), \\ \vdots \\ p_{hj,k}(t) = g\left(\mu_{hj,k} + \boldsymbol{\theta}_{hj,k}^\top \mathbf{Z}(t)\right) - g\left(\mu_{hj-1,k} + \boldsymbol{\theta}_{hj-1,k}^\top \mathbf{Z}(t)\right), \\ \vdots \\ p_{hR,k}(t) = 1 - g\left(\mu_{hR-1,k} + \boldsymbol{\theta}_{hR-1,k}^\top \mathbf{Z}(t)\right). \end{cases}$$

## Dynamique des facteurs macros

- ▶ Modèle VAR( $m$ ) avec  $m \geq 1$

$$\mathbf{Z}(t) = \sum_{i=1}^m \mathbf{A}_i \mathbf{Z}(t-i) + \mathbf{b}_0 + \mathbf{b}_1 t + \boldsymbol{\zeta}(t) ,$$

avec  $\boldsymbol{\zeta}(t) \equiv$  vecteur gaussien de variance-covariance  $\boldsymbol{\Sigma}$ .

- ▶ Transformation en modèle VECM pour gérer les problèmes de stationnarité.
- ▶  $R(t) \equiv$  indicatrice de censure (indépendante et non informative).
- ▶ Vraisemblance du modèle

$$\mathcal{L}(X, R, \mathbf{Z}; (\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\kappa})) = \mathcal{L}(\mathbf{Z}; \boldsymbol{\kappa}) \times \mathcal{L}(X, R; (\boldsymbol{\mu}, \boldsymbol{\theta})) .$$

- ▶ Permet l'inférence du modèle en deux temps par maximum de vraisemblance.

## Données d'assurance crédit

- ▶ Système de *rating* interne constitué de 10 classes, regroupés en 5 classes et les défauts  $P$  et  $I$

$$\mathcal{R} = (1, \dots, 5, P, I).$$

- ▶ Estimation sur base *trimestrielle* : regroupement des transitions et ajustement pour prendre en compte des transitions intermédiaires.
- ▶ l'état  $P$  est supposé vérifier l'hypothèse de Markov d'ordre 2.
- ▶ Reclassification de certains *buyers*, appartenant aux classes 1 et 2 vers la classe 5.

## Méthode de simulation

- Projection de la population entre  $t$  et  $t + 1$ , connaissant la situation en  $t - 1$ . Pour chaque simulation  $s$  :

1 Tirer  $\zeta(t + 1)^s$  selon une loi  $\mathcal{N}(0, \Sigma)$ .

2 Générer  $Z(t + 1)^s$  à partir de

$$Z(t + 1)^s = \sum_{i=1}^m \widehat{A}_i Z(t + 1 - i) + \widehat{b}_0 + \widehat{b}_1 t + \zeta(t + 1)^s.$$

3 Transformer  $Z(t + 1)^s$  en  $\widetilde{Z}(t + 1)^s$ .

4 Calculer  $p_k(\widetilde{Z}(t + 1)^s)$  via le modèle de migration.

5 Simuler l'évolution de la population entre  $t$  et  $t + 1$  au moyen de loi multinomiale.



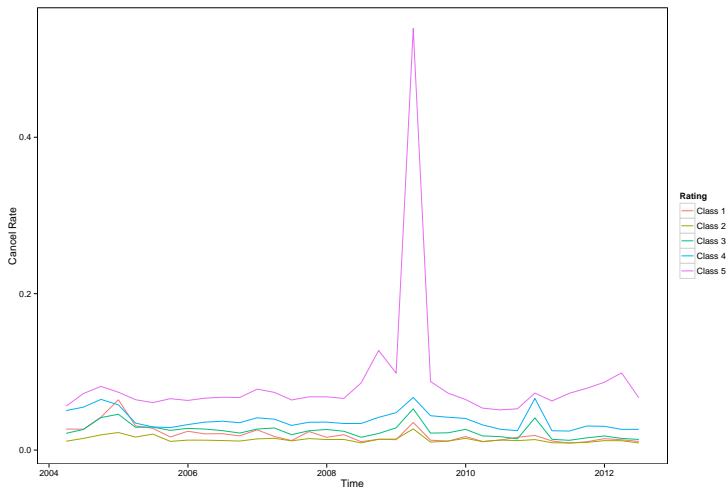
## Données d'assurance crédit

- ▶ Trajectoire de *ratings* individuels d'un assureur crédit (1 604 533 *buyers*) avec fréquence mensuelle.
- ▶ Période d'observation : Mars 2004 - Septembre 2012.
- ▶ Expositions nationales et internationales → on se limite aux données françaises.
- ▶ Hétérogénéité en termes de taille (PME et grosses entreprises) et de secteur.
- ▶ Regroupement en 5 secteurs.

Secteur	Nom	Nombre de <i>buyers</i>
1	Agriculture	93 697
2	Finance/Immobilier	148 582
3	Produits finis	187 857
4	Matières premières	471 270
5	Services/Commerce	703 127

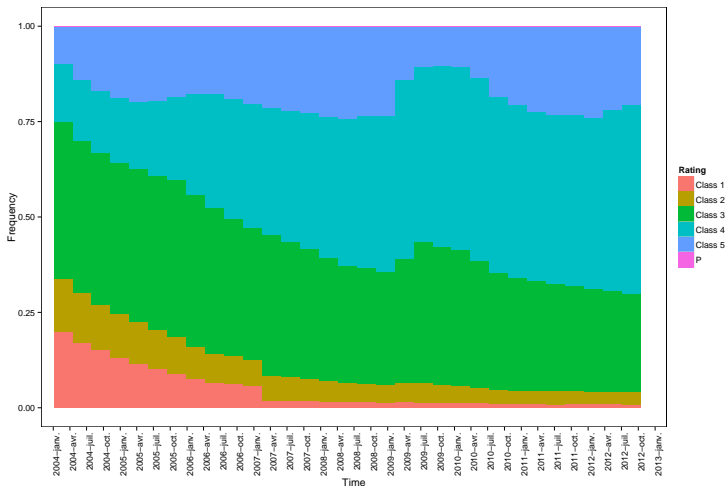
## Données

► Taux de clôture.



## Données

### ► Répartition par *rating*.



## Covariables

- Choix des facteurs macros selon leur prédictivité et leur utilité pour la gestion.

Variable	Définition	Moyenne	Écart-type
<i>Conditions macroéconomiques générales</i>			
$\tilde{Z}_1(t) = UR(t) - UR(t-4)$	$\Delta$ annuelle du log taux de chômage $UR(t)$	1.14%	8.96%
$\tilde{Z}_2(t) = CPI(t) - CPI(t-4)$	$\Delta$ annuelle du log $CPI(t)$	1.71%	0.82%
$\tilde{Z}_3(t) = EB(t) - EB(t-4)$	$\Delta$ annuelle du log du nb de créations d'entreprises $EB(t)$	0.52%	11.34%
<i>Direction de l'économie</i>			
$\tilde{Z}_4(t) = GDP(t) - GDP(t-4)$	$\Delta$ annuelle du log du $GDP(t)$	1.18%	1.85%
$\tilde{Z}_5(t) = IP(t) - IP(t-4)$	$\Delta$ annuelle du log de la production industrielle $IP(t)$	-1.04%	6.12%
<i>Conditions sur les marchés financiers</i>			
$\tilde{Z}_6(t) = 0.25 \ln(1 + RF(t))$	Transformation trimestrielle des taux souverains français à 10 ans $RF(t)$	0.90%	0.14%
$\tilde{Z}_7(t) = CAC(t) - CAC(t-4)$	$\Delta$ annuelle du log du CAC40 $CAC(t)$	-0.01%	21.62%
$\tilde{Z}_8(t) = \sigma(t)$	Vol trimestrielle du CAC40	21.94%	10.60%

## Covariables

- Choix des facteurs macros selon leur prédictivité et leur utilité pour la gestion.

Variable	Définition	Moyenne	Écart-type
<i>Conditions macroéconomiques générales</i>			
$\tilde{Z}_1(t) = UR(t) - UR(t-4)$	$\Delta$ annuelle du log taux de chômage $UR(t)$	1.14%	8.96%
$\tilde{Z}_2(t) = CPI(t) - CPI(t-4)$	$\Delta$ annuelle du log $CPI(t)$	1.71%	0.82%
$\tilde{Z}_3(t) = EB(t) - EB(t-4)$	$\Delta$ annuelle du log du nb de créations d'entreprises $EB(t)$	0.52%	11.34%
<i>Direction de l'économie</i>			
$\tilde{Z}_4(t) = GDP(t) - GDP(t-4)$	$\Delta$ annuelle du log du $GDP(t)$	1.18%	1.85%
$\tilde{Z}_5(t) = IP(t) - IP(t-4)$	$\Delta$ annuelle du log de la production industrielle $IP(t)$	-1.04%	6.12%
<i>Conditions sur les marchés financiers</i>			
$\tilde{Z}_6(t) = 0.25 \ln(1 + RF(t))$	Transformation trimestrielle des taux souverains français à 10 ans $RF(t)$	0.90%	0.14%
$\tilde{Z}_7(t) = CAC(t) - CAC(t-4)$	$\Delta$ annuelle du log du CAC40 $CAC(t)$	-0.01%	21.62%
$\tilde{Z}_8(t) = \sigma(t)$	Vol trimestrielle du CAC40	21.94%	10.60%

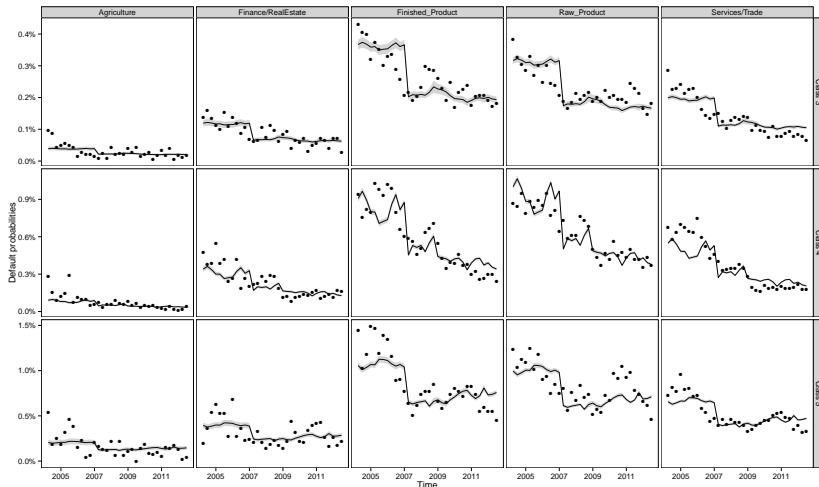
## Comparaison des modèles

### ► Modèles avec transitions non-parallèles.

Modèle	1	2	3	4	5
<b>M1*</b> : $(\bar{Z}_1(t), \bar{Z}_2(t), \bar{Z}_3(t))$					
AIC	15417.54	12875.74	60663.97	126999.99	157815.43
BIC	22977.54	20435.74	71247.97	137583.99	166887.43
Log Vraisemblance	-7663.77	-6392.87	-30268.99	-63437.00	-78853.71
McFadden R <sup>2</sup> ajusté	0.53	0.36	0.54	0.51	0.50
<b>M2*</b> : $(\bar{Z}_4(t), \bar{Z}_5(t))$					
AIC	16081.48	13569.05	64006.56	107196.71	98392.70
BIC	22801.48	<b>20289.05</b>	73414.56	116604.71	106456.70
Log Vraisemblance	-8000.74	-6744.53	-31947.28	-53542.35	-49148.35
McFadden R <sup>2</sup> ajusté	0.51	0.46	0.51	0.59	0.69
<b>M3*</b> : $(\bar{Z}_6(t), \bar{Z}_7(t), \bar{Z}_8(t))$					
AIC	14414.26	13514.79	62888.78	126951.08	125985.13
BIC	<b>21974.26</b>	21074.79	73472.78	137535.08	135057.13
Log Vraisemblance	-7162.13	-6712.39	-31381.39	-63412.54	-62938.57
McFadden R <sup>2</sup> ajusté	0.57	0.43	0.53	0.51	0.60
<b>M4*</b> : $(\bar{Z}_1(t), \dots, \bar{Z}_8(t))$					
AIC	<b>12592.34</b>	<b>11190.12</b>	<b>50472.16</b>	<b>85440.00</b>	<b>70664.23</b>
BIC	24352.34	22950.12	<b>66936.16</b>	<b>101904.00</b>	<b>84776.23</b>
Log Vraisemblance	<b>-6226.17</b>	<b>-5525.06</b>	<b>-25138.08</b>	<b>-42622.00</b>	<b>-35248.12</b>
McFadden R <sup>2</sup> ajusté	<b>0.64</b>	<b>0.56</b>	<b>0.63</b>	<b>0.67</b>	<b>0.78</b>
<b>M5*</b> : $(\bar{Z}_5(t), \bar{Z}_6(t), \bar{Z}_8(t))$					
AIC	14764.97	12996.75	59513.61	102565.07	94791.34
BIC	22324.97	20556.75	70097.61	113149.07	103863.34
Log Vraisemblance	-7337.49	-6453.38	-29693.81	-51219.54	-47341.67
McFadden R <sup>2</sup> ajusté	0.55	0.45	0.55	0.61	0.70

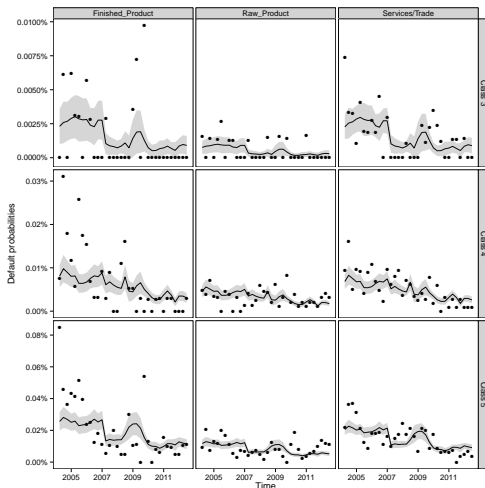
## Probabilités de transition avec M5\*

- Estimation des probabilités de défaut  $P$  trimestrielles.



## Probabilités de transition avec M5\*

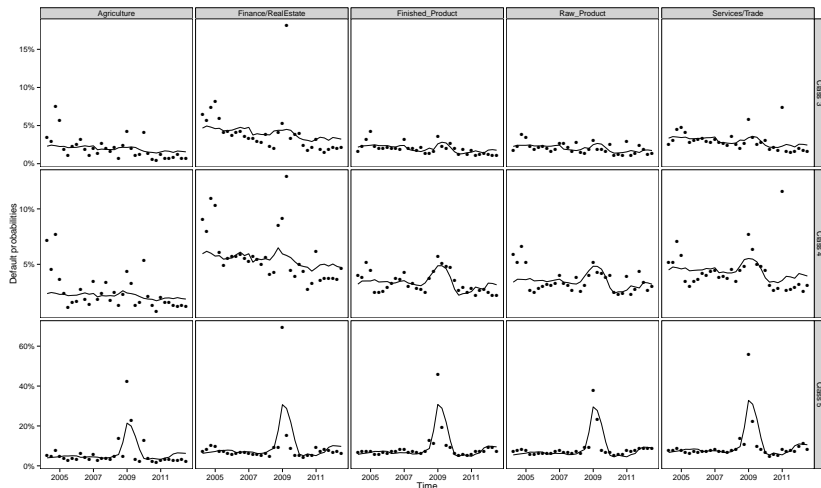
- Estimation des probabilités de défaut  $I$  trimestrielles.





## Probabilités de transition avec M5\*

- Estimation des probabilités de **clôture C** trimestrielles.

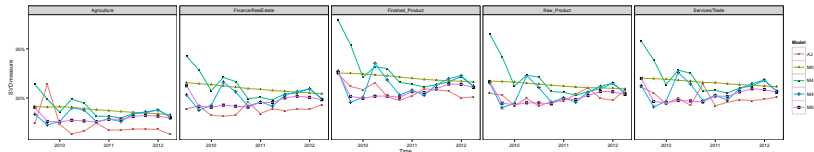


## Performances des modèles de migration

- ▶ Performances *out-of-sample* sur les matrices de transition trimestrielles.

$$SVD_k(t) = \frac{1}{R} \sum_{i=1}^R \sqrt{\lambda_i \left( (\tilde{\mathbf{p}}_k(t))^T \tilde{\mathbf{p}}_k(t) \right)}$$

avec  $\tilde{\mathbf{p}}_k(t) = \mathbf{p}_k(t) - Id$ , où  $\lambda_i(\mathbf{p})$  retourne la  $i$ -ème valeur propre de la matrice  $\mathbf{p}$ .

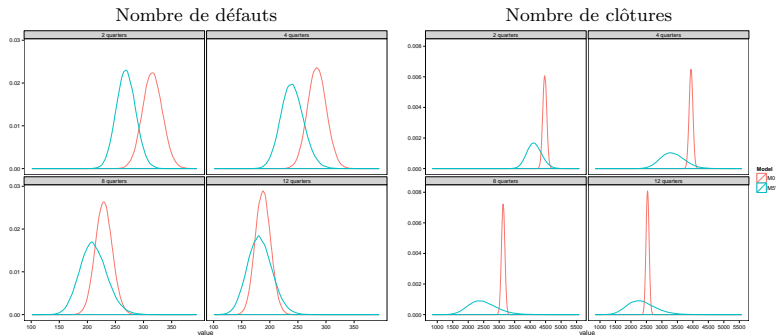


## Estimation du modèle pour les facteurs macros

- ▶ Test de Dickey-Fuller augmenté sur  $(UR(t), CPI(t), GDP(t), IP(t), r(t), CAC(t), \sigma(t))$ .
- ▶  $(UR(t), \Delta CPI(t), GDP(t), IP(t), r(t), CAC(t), \sigma(t))$  sont au moins  $I(1)$ .
- ▶ Modèle VAR(2) avec constante et une tendance déterministe.
- ▶ Quelques problèmes d'autocorrélation et de normalité des résidus.
- ▶ Analyse de la cointégration  $\Rightarrow r = 4$ .
- ▶ Approche simplifiée dans laquelle la normalisation du modèle n'est pas faite en intégrant des contraintes déduites de la théorie économique.

## Application à la gestion des risques

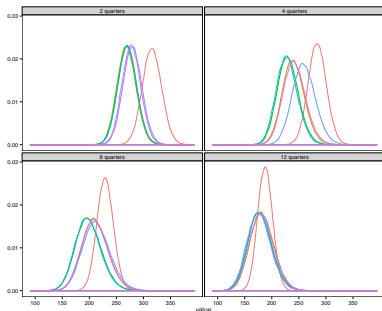
- Comparaison du modèle M0 (sans facteur) et de M5\*.



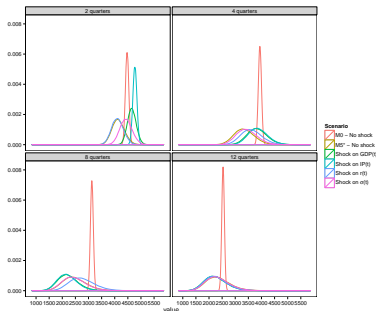
## Application à la gestion des risques

- ▶ Chocs survenant durant le 1<sup>er</sup> trimestre :
  - ▶ choc de  $-2,58\sigma$  sur  $GDP(t)$  ;
  - ▶ choc de  $-2,58\sigma$  sur  $IP(t)$  ;
  - ▶ choc de  $+2,58\sigma$  sur  $r(t)$  ;
  - ▶ choc de  $+2,58\sigma$  sur  $\sigma(t)$ .

Nombre de défauts



Nombre de clôtures



Scenario  
 MS - No shock  
 MS\* - No shock  
 Shock on GDP(t)  
 Shock on IP(t)  
 Shock on r(t)  
 Shock on  $\sigma(t)$

## Quelques Références I

Akritis, M. G. (2000). The central limit theorem under censoring. *Bernoulli*, **6**(6), 1109–1120.

Andersen, P. K. and Perme, M. P. (2008). Inference for outcome probabilities in multi-state models. *Lifetime data analysis*, **14**(4), 405–431.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag New York Inc.

Beyersmann, J., Termini, S. D., and Pauly, M. (2013). Weak Convergence of the Wild Bootstrap for the Aalen–Johansen Estimator of the Cumulative Incidence Function of a Competing Risk. *Scandinavian Journal of Statistics*, **40**(3), 387–402.

CMIR12 (1991). The Analysis of Permanent Health Insurance Data. Technical report, Continuous Mortality Investigation Bureau, The Institute of Actuaries and the Faculty of Actuaries.

Dabrowska, D. (1995). Estimation of transition probabilities and bootstrap in a semiparametric markov renewal model. *Journal of Nonparametric Statistics*, **5**(3), 237–259.

de Uña-Álvarez, J. and Meira-Machado, L. (2015). Nonparametric estimation of transition probabilities in the non-Markov illness-death model : A comparative study. *Biometrics*.

## Quelques Références II

- Dobler, D. and Pauly, M. (2014). How to Bootstrap Aalen-Johansen Processes for Competing Risks ? Handicaps, Solutions and Limitations. *arXiv :1401.7801 [math, stat]*.
- Efron, B. (1979). Bootstrap Methods : Another Look at the Jackknife. *The Annals of Statistics*, **7**(1), 1–26.
- Efron, B. (1981). Censored Data and the Bootstrap. *Journal of the American Statistical Association*, **76**(374), 312–319.
- Gill, R. D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Probability Theory and Related Fields*, **53**(1), 97–116.
- Haberman, S. and Pitacco, E. (1998). *Actuarial Models for Disability Insurance*. Chapman and Hall/CRC, 1 edition.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**(282), 457–481.
- Mathieu, E., Foucher, Y., Dellamonica, P., and Daures, J. P. (2007). Parametric and Non Homogeneous Semi-Markov Process for HIV Control. *Methodology and Computing in Applied Probability*, **9**(3), 389–397.

## Quelques Références III

- Meira-Machado, L., de Uña-Álvarez, J., and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness–death model. *Lifetime Data Analysis*, **12**(3), 325–344.
- Monteiro, A., Smirnov, G. V., and Lucas, A. (2006). Non-parametric Estimation for Nonhomogeneous Semi-Markov Processes : An Application to Credit Risk. Discussion Paper TI 2006-024/2, Tinbergen Institute, Amsterdam.
- Pepe, M. S. (1991). Inference for Events With Dependent Risks in Multiple Endpoint Studies. *Journal of the American Statistical Association*, **86**(415), 770–778.
- Rotolo, F., Legrand, C., and Van Keilegom, I. (2013). A simulation procedure based on copulas to generate clustered multi-state survival data. *Computer Methods and Programs in Biomedicine*, **109**(3), 305–312.
- Sánchez-Sellero, C., Manteiga, W. G., and Van Keilegom, I. (2005). Uniform Representation of Product-Limit Integrals with Applications. *Scandinavian Journal of Statistics*, **32**(4), 563–581.
- Spitoni, C., Verduijn, M., and Putter, H. (2012). Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi-State Models. *The International Journal of Biostatistics*, **8**(1).
- Stute, W. (1995). The Central Limit Theorem Under Random Censorship. *The Annals of Statistics*, **23**(2), 422–439.