

**100% ACTUAIRES &  
100% DATA SCIENCE**

INSTITUT DES  
**ACTUAIRES**



**16 Novembre 2018**  
Hôtel Marriott Rive Gauche  
Paris 14ème

# Séries temporelles en grande dimension appliquées à la projection de la mortalité

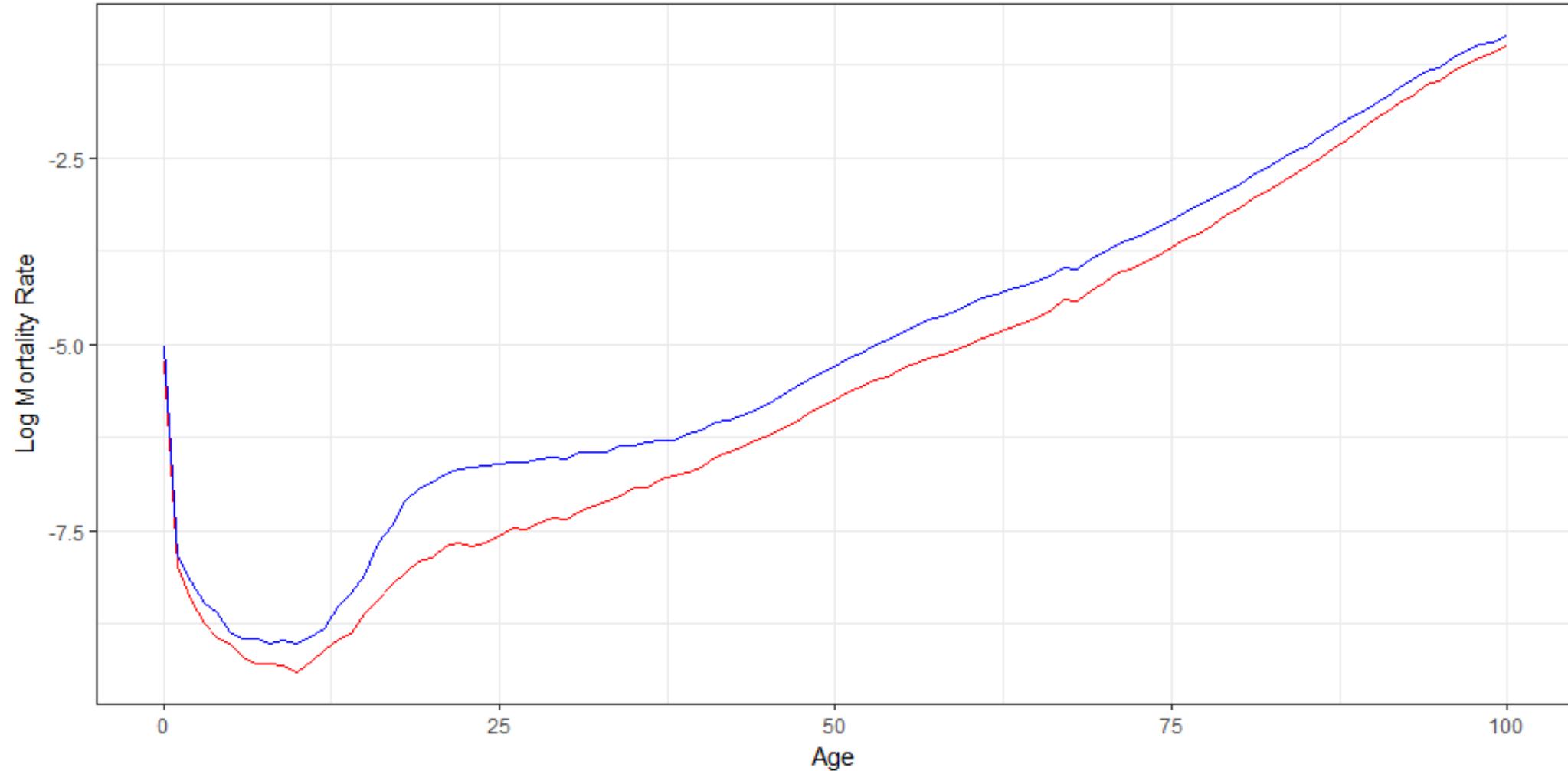
Quentin Guibert (Prim'Act / Univ. Paris Dauphine)  
Pierrick Piette (Sinalys / Univ. Lyon 1)

Travaux en collaboration avec  
Olivier Lopez (UPMC)

- **Introduction**
- **Modèle VAR en grande dimension**
- **Résultats**
- **Conclusion**

- **Introduction**
- Modèle VAR en grande dimension
- Résultats
- Conclusion

Mortality rates in the US (2014) for males (blue) and females (red)



- Lee-Carter (1992) et **Renchow-Haberman (2006)**

$$\ln(m_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t + \beta_x^{(2)} \gamma_{t-x}$$

- CBD (2006) et **M7 (2009)**

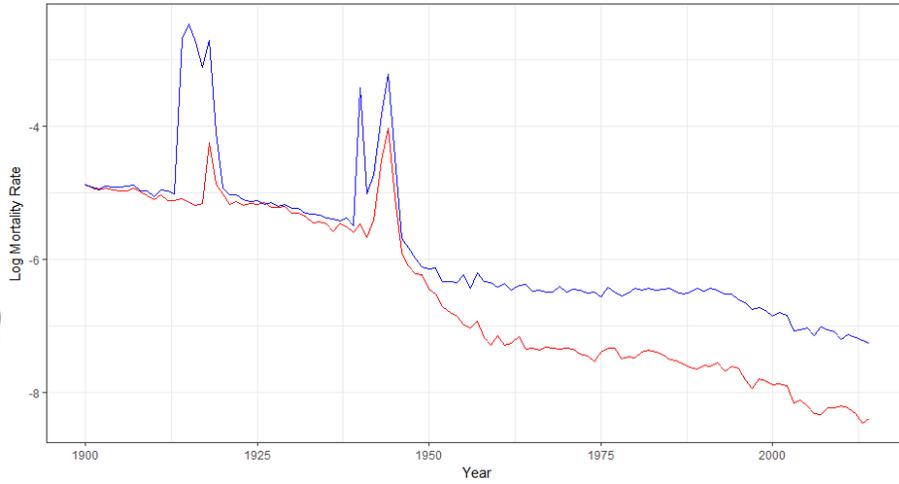
$$\ln\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \hat{\sigma}^2) + \gamma_{t-x}$$

- APC (2009)

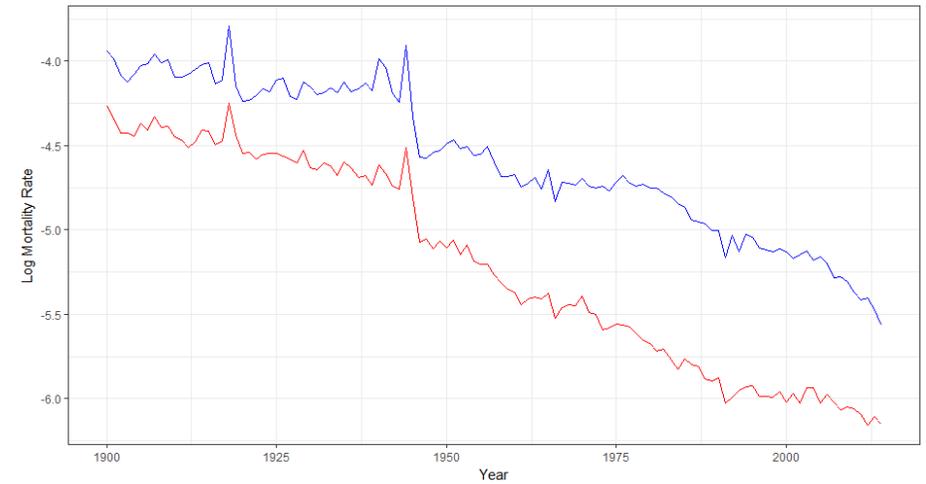
$$\ln(m_{x,t}) = \beta_x + \kappa_t + \gamma_{t-x}$$

- Facilement extensible à l'analyse multi-population
  - Pas de population dominante
  - Pas de limite à deux populations
- Une approche centrée sur les données
  - Imposer le moins d'a priori possible
  - Compatible avec la grande dimension
- Se détacher des modèles à facteurs
  - Pas d'estimation en deux temps

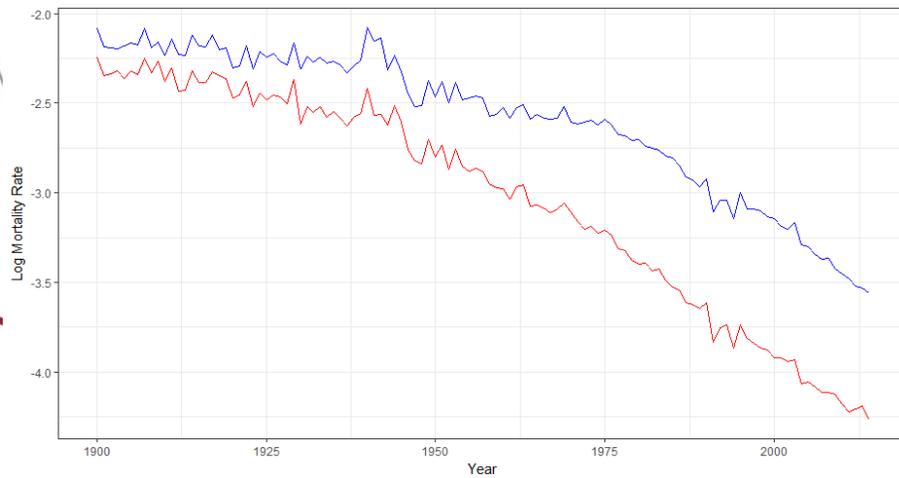
25 ans



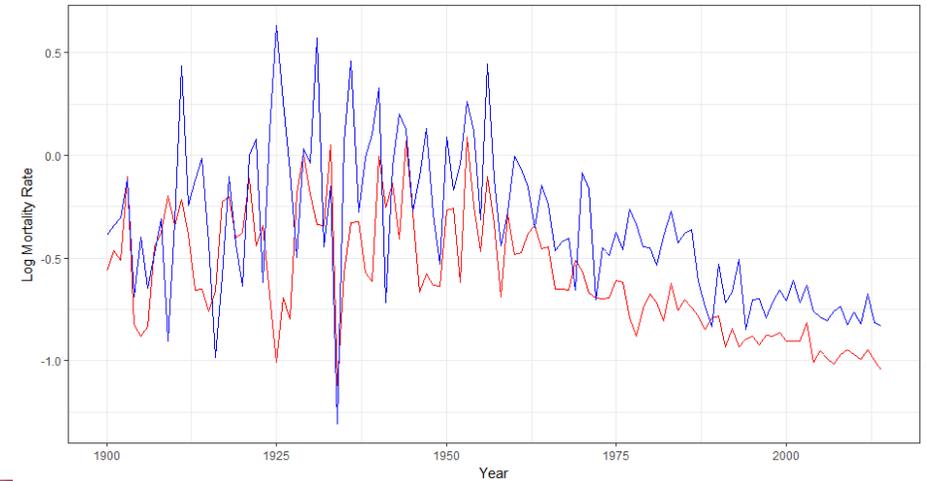
50 ans



75 ans



100 ans



- Introduction
- **Modèle VAR en grande dimension**
- Résultats
- Conclusion

- Les séries de log-mortalité ne sont pas stationnaires, nous étudions la série différenciée, i.e. le taux d'amélioration de la mortalité

$$\Delta y_{i,t} = y_{i,t} - y_{i,t-1} = \ln(m_{i,t}) - \ln(m_{i,t-1})$$

- Projection de la courbe par une approche vectorielle

$$\Delta Y_t = (\Delta y_{i_{\min},t}, \Delta y_{i_{\min}+1,t}, \dots, \Delta y_{i_{\max}-1,t}, \Delta y_{i_{\max},t})^T$$

- Grande liberté dans la structure spatio-temporelle de dépendance

$$\Delta Y_t = C + \sum_{k=1}^p A_k \Delta Y_{t-k} + E_t$$

- L'estimation du modèle est un problème de grande dimension
  - Les  $p$  matrices autorégressives sont de dimension  $d^2$
  - Le bruit Gaussien a une matrice var-covar de dimension  $\frac{d(d+1)}{2}$

$$\Delta Y_t = C + \sum_{k=1}^p A_k \Delta Y_{t-k} + E_t$$

- Particulièrement vrai dans la modélisation de la mortalité
  - Tranche d'âge 0-100 ans avec 70 ans d'historique
  - VAR(3) implique 30 704 coefficients (sans la matrice var-covar)
  - Seulement 7 070 points d'observation
- Besoin d'une approche spécifique pour l'estimation

- Utilisation d'une régression pénalisée, nouveau critère à optimiser:

$$\frac{1}{T-p} \sum_{t=p}^T \left\| \Delta Y_t - C - \sum_{k=1}^p A_k \Delta Y_{t-k} \right\|_2^2 - \alpha \lambda \sum_{k=1}^p \|A_k\|_1 - \frac{(1-\alpha)\lambda}{2} \sum_{k=1}^p \|A_k\|_2^2$$

- L'Elastic-Net (Zou & Hastie, 2005) est double pénalisation
  - Lasso (Tibshirani, 1996): sélection de variable
  - Ridge (Hoerl & Kennard, 1988): effet de groupement
  - Implémentation récente sur les modèles VAR (Basu & Michailidis, 2015; Gefang, 2014)
- Deux hyperparamètres dans l'estimation Elastic-Net
  - $\lambda$  est estimé par cross-validation
  - $\alpha$  est estimé par grid search

- Matrice Variance-Covariance est un problème de grande dimension
  - $d(d + 1)/2$  paramètres
  - Existence de techniques pour la grande dimension (Friedman et al., 2008; Schäfer & Strimmer, 2005)
  - Choix d'une modélisation par fonction de covariance (Spodarev et al., 2015)

- Fonction de corrélation

- Estimation des variances empiriques  $\hat{\sigma}_i^2$  et des corrélations empiriques

$$\hat{r}_{i,j} = \frac{\hat{\sigma}_{i,j}}{\hat{\sigma}_i \hat{\sigma}_j}$$

- Choix d'une fonction de corrélation proche de la famille stable.

$$r_{i,j} = \beta e^{-(\alpha_i + \alpha_j) \times |i-j|} \times \mathbf{1}_{\{i \neq j\}} + \mathbf{1}_{\{i=j\}}$$

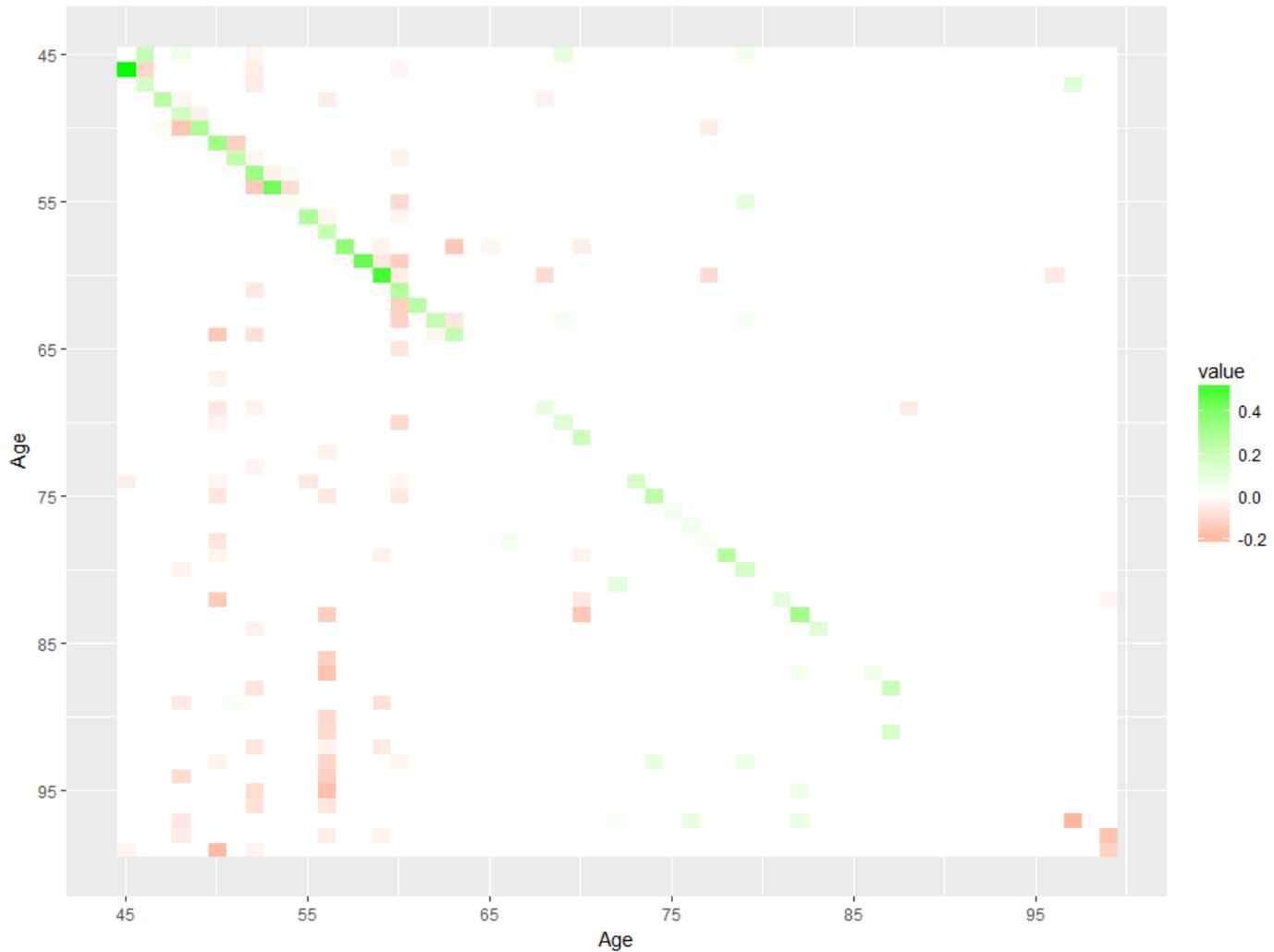
Avec  $\beta > 0$  et  $\alpha_i \geq 0$ .

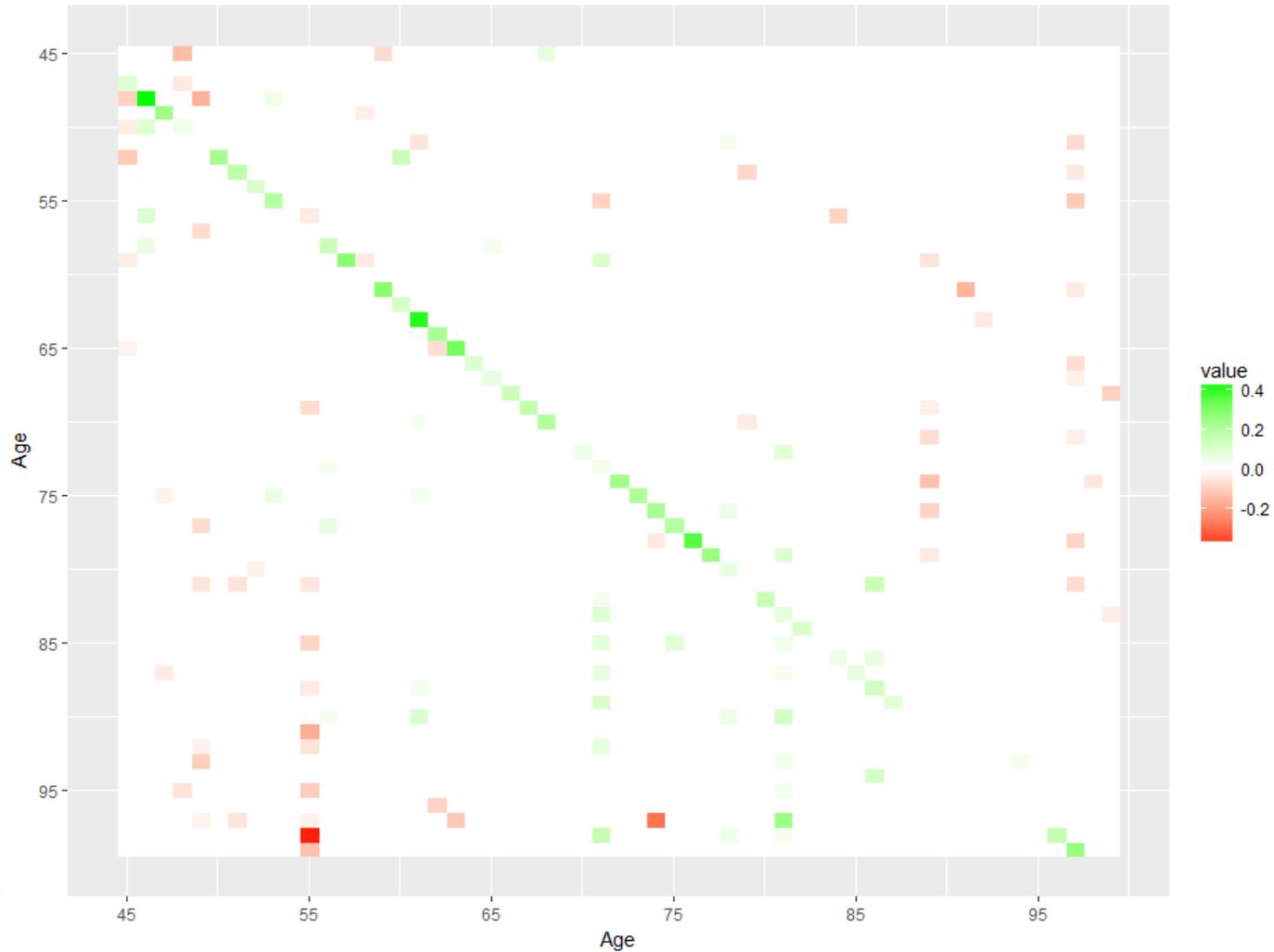
- $2d + 1$  paramètres.

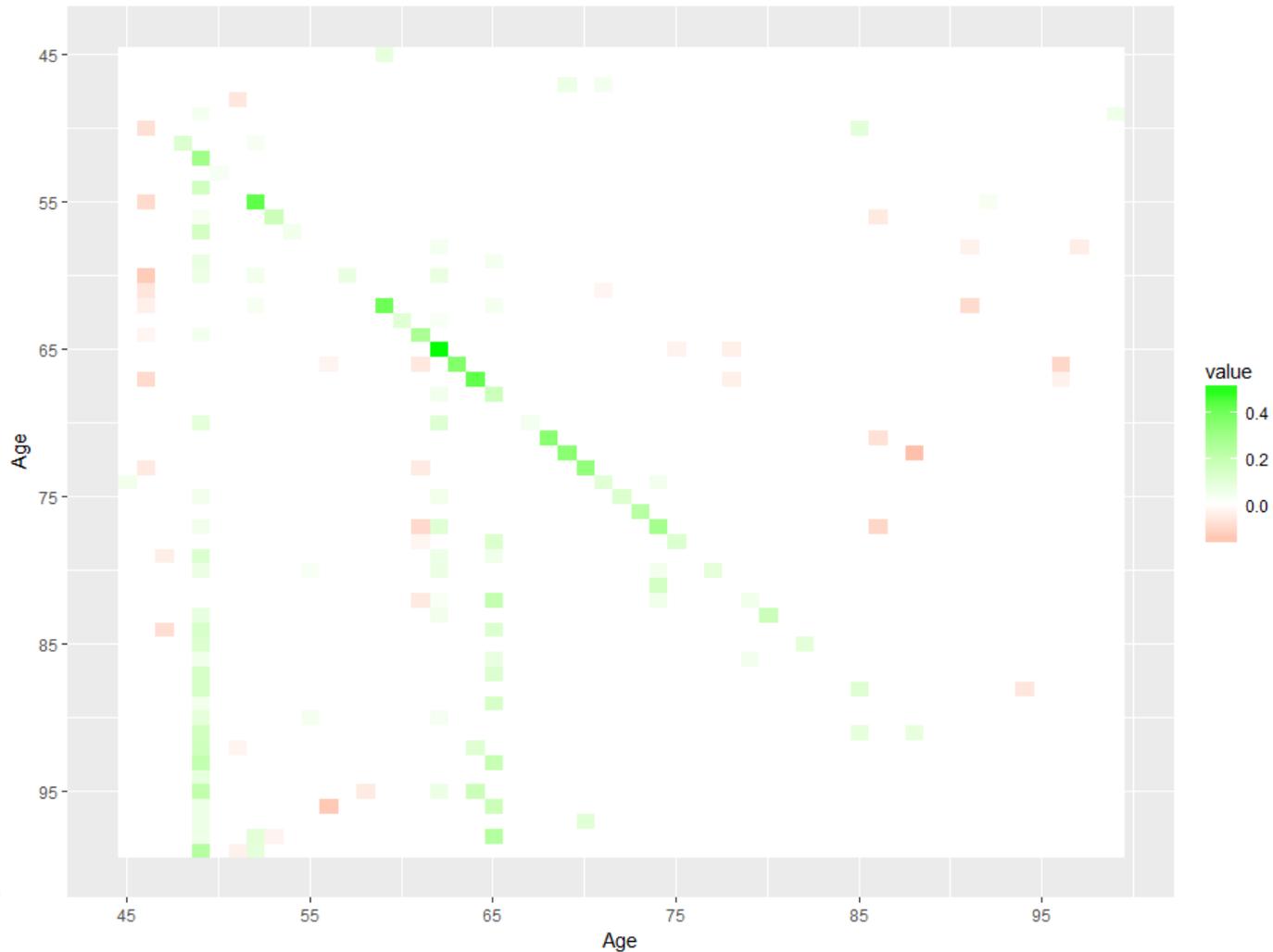
- En général, en série temporelle, utilisation d'un critère pour le choix du lag  $p$ 
  - AIC, BIC
  - Sélection de modèle
- VAR-ENET en grande dimension
  - La pénalisation LASSO est une technique de sélection de modèle
  - Le nombre de paramètres n'est pas complètement déterminé par  $p$
  - Peu de littérature sur VAR-ENET en grande dimension, pas de *best practice*
- Notre choix
  - Choix d'un  $p$  relativement grand, quitte à avoir des matrices quasi-nulles
  - Déterminé par la taille de l'historique
  - $p = 7$  (i.e. entre 10 et 15% de l'historique)

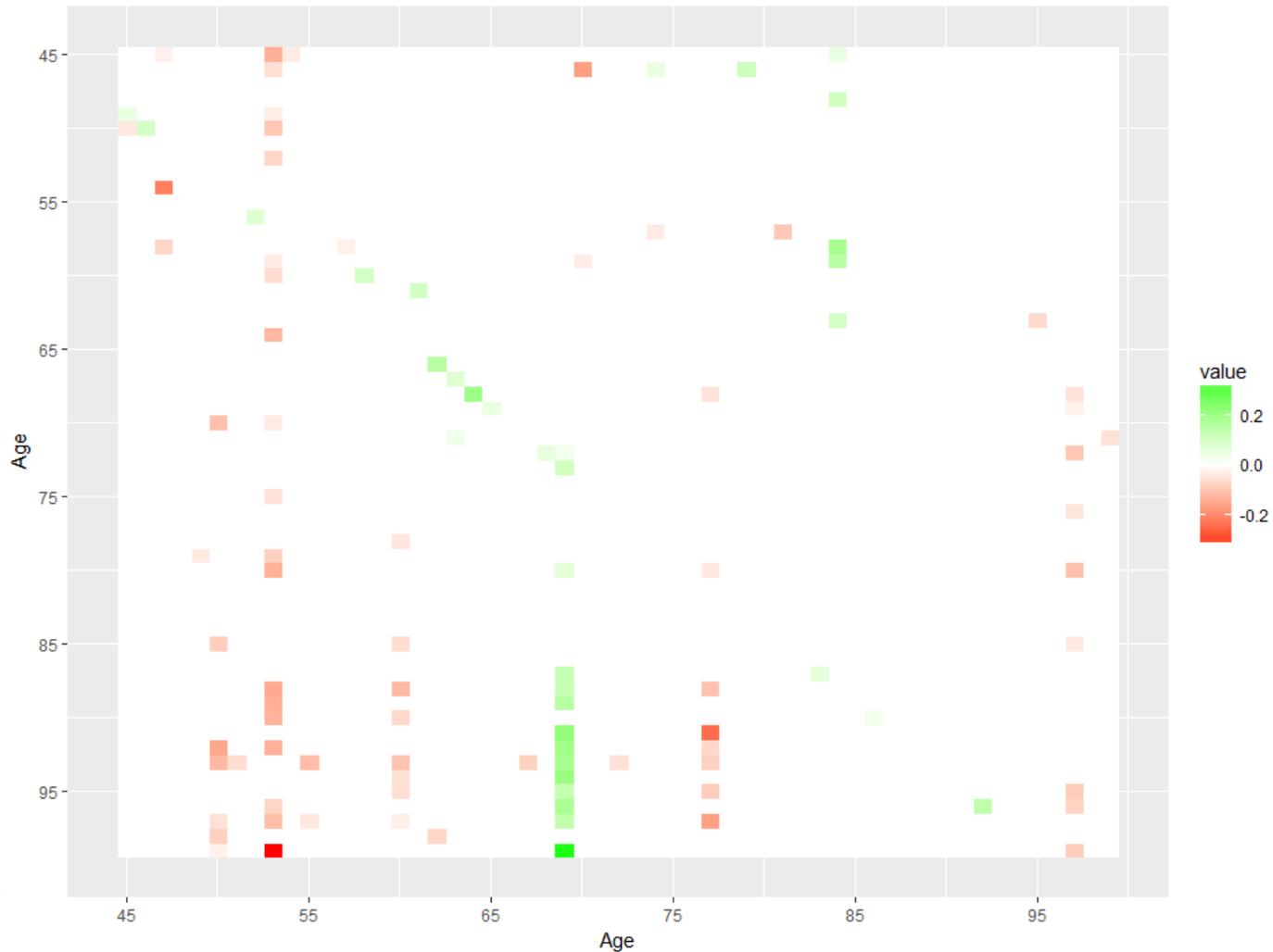
- Introduction
- Modèle VAR en grande dimension
- **Résultats**
- Conclusion

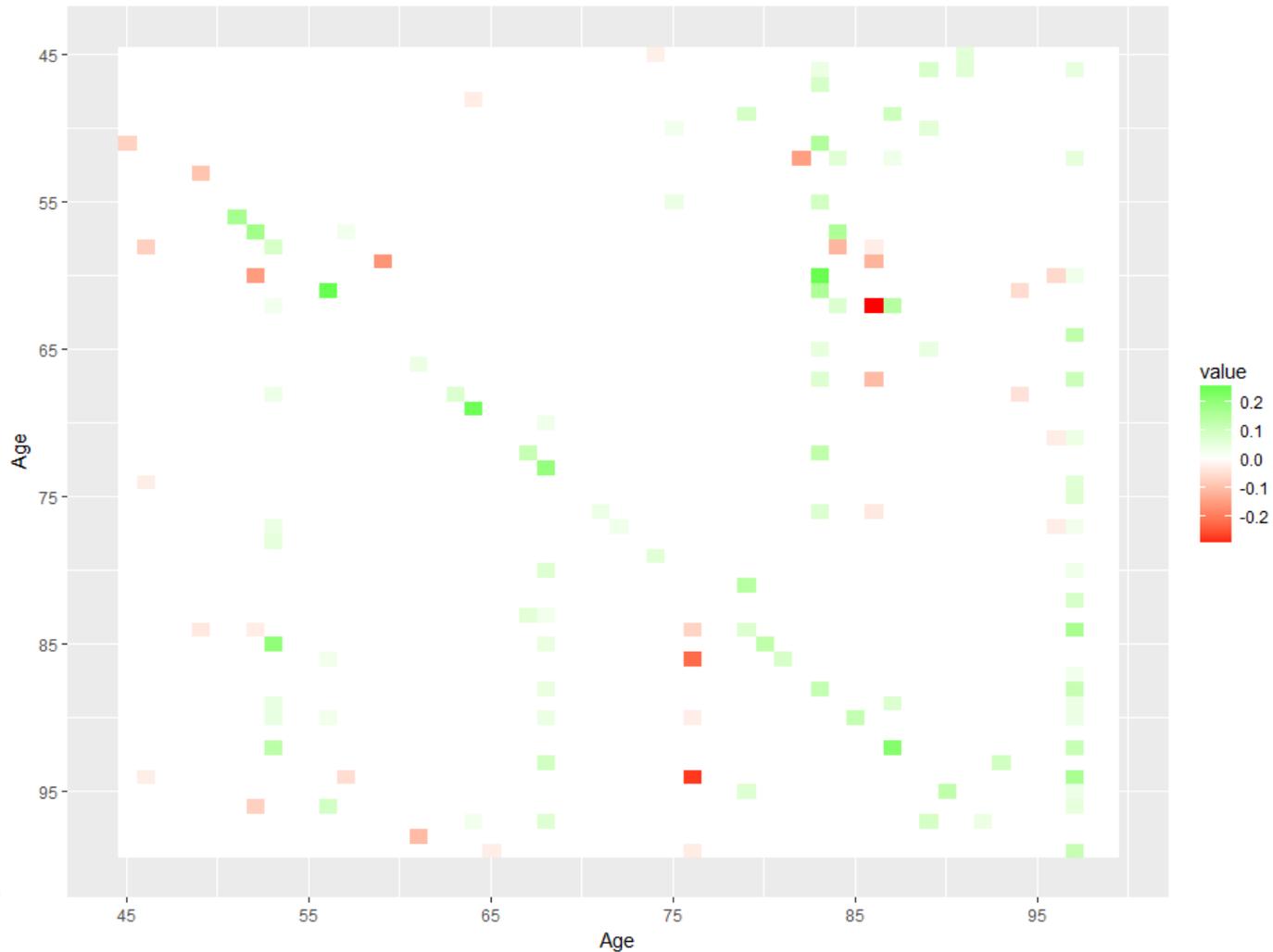
- Base Human Mortality Database
  - [mortality.org](http://mortality.org)
- Trois pays
  - France (FR)
  - États-Unis (US)
  - Angleterre et Pays de Galles (UK)
- Période
  - 1950-2012
- Ages
  - 45-99

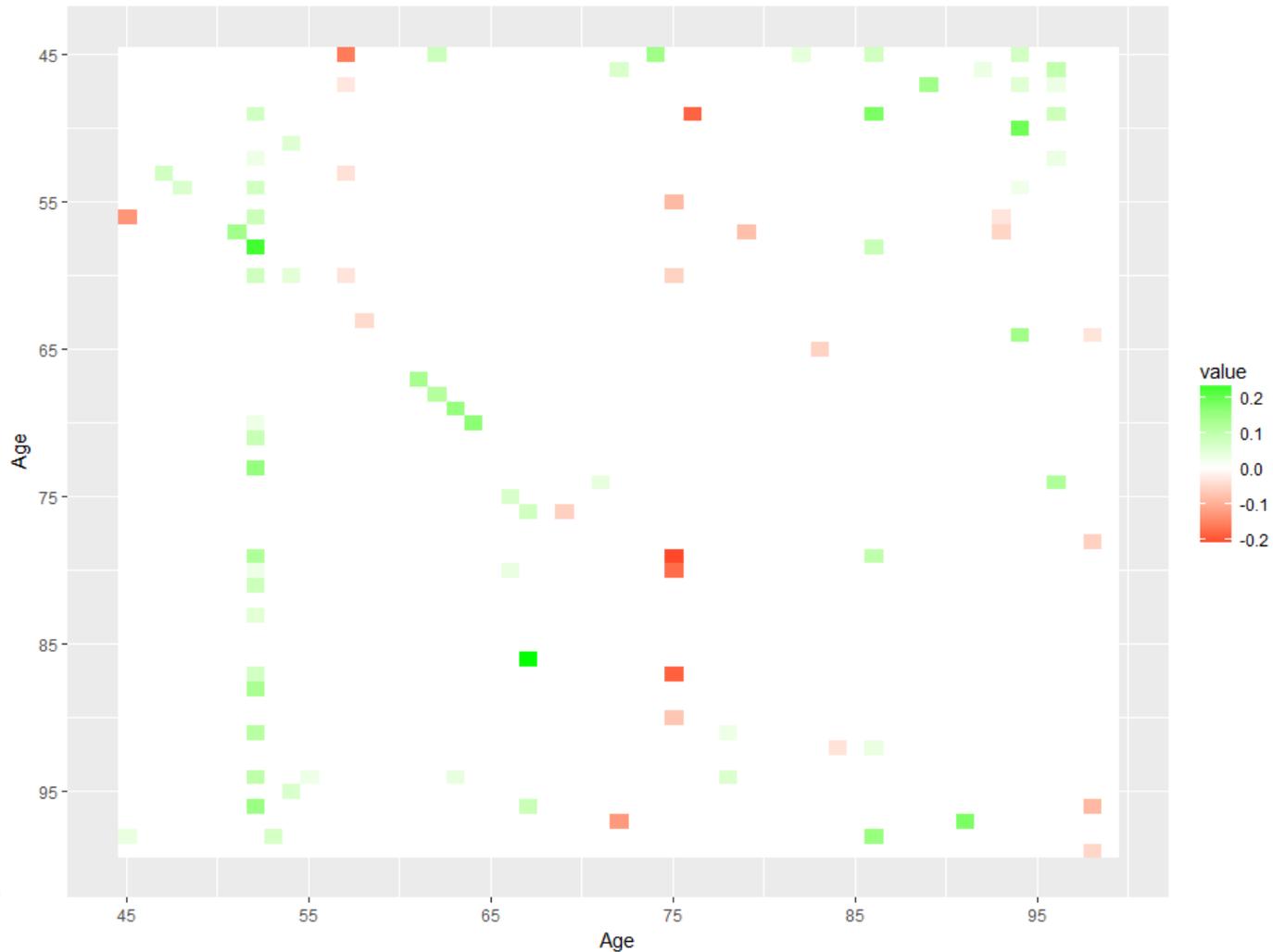


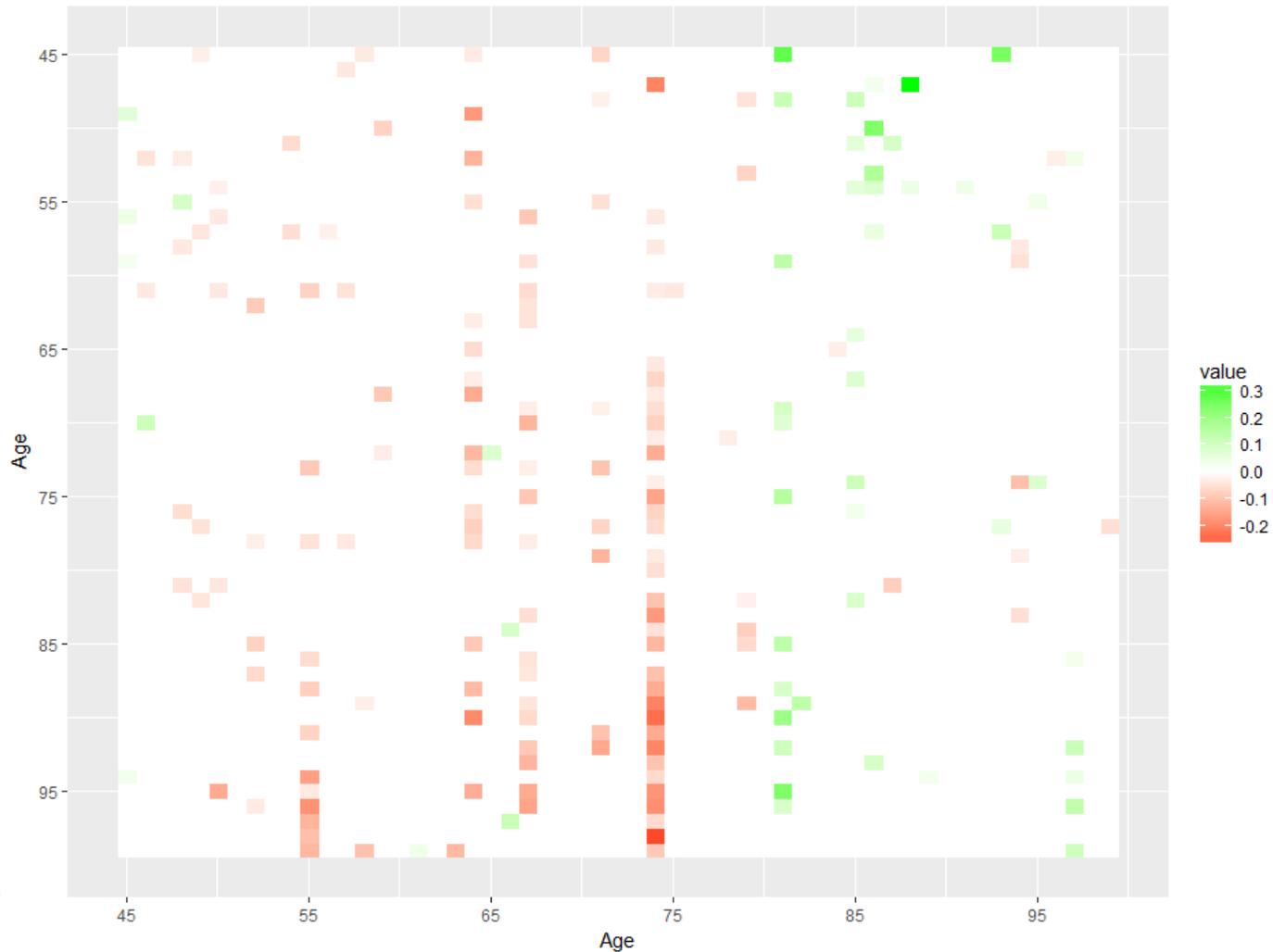












- Métrique de comparaison : Root Mean Squared Error

- Procédure

- Estimation des modèles sur 1950-2000
- Projection sur 2001-2012
- RMSE entre le projeté et l'observé

$$RMSE = \sqrt{\frac{1}{I \times T} \sum_{(i,t)} (\hat{y}_{i,t} - y_{i,t})^2}$$

RMSE out-sample (10<sup>-2</sup>)

		VAR	LC	CBD	M7
US	Total	7.76	8.66	14.98	10.31
	Female	6.65	8.50	14.32	17.20
	Male	8.80	10.40	13.38	10.94
UK	Total	7.56	12.78	9.65	12.59
	Female	9.07	13.32	12.05	17.14
	Male	8.35	12.20	8.40	9.82
FR	Total	6.32	6.58	20.52	20.06
	Female	7.56	10.29	36.18	50.00
	Male	7.94	9.78	15.43	15.89

• Procédure

- Estimation des modèles sur 1970-2000
- Projection sur 2001-2012

RMSE out-sample ( $10^{-2}$ )

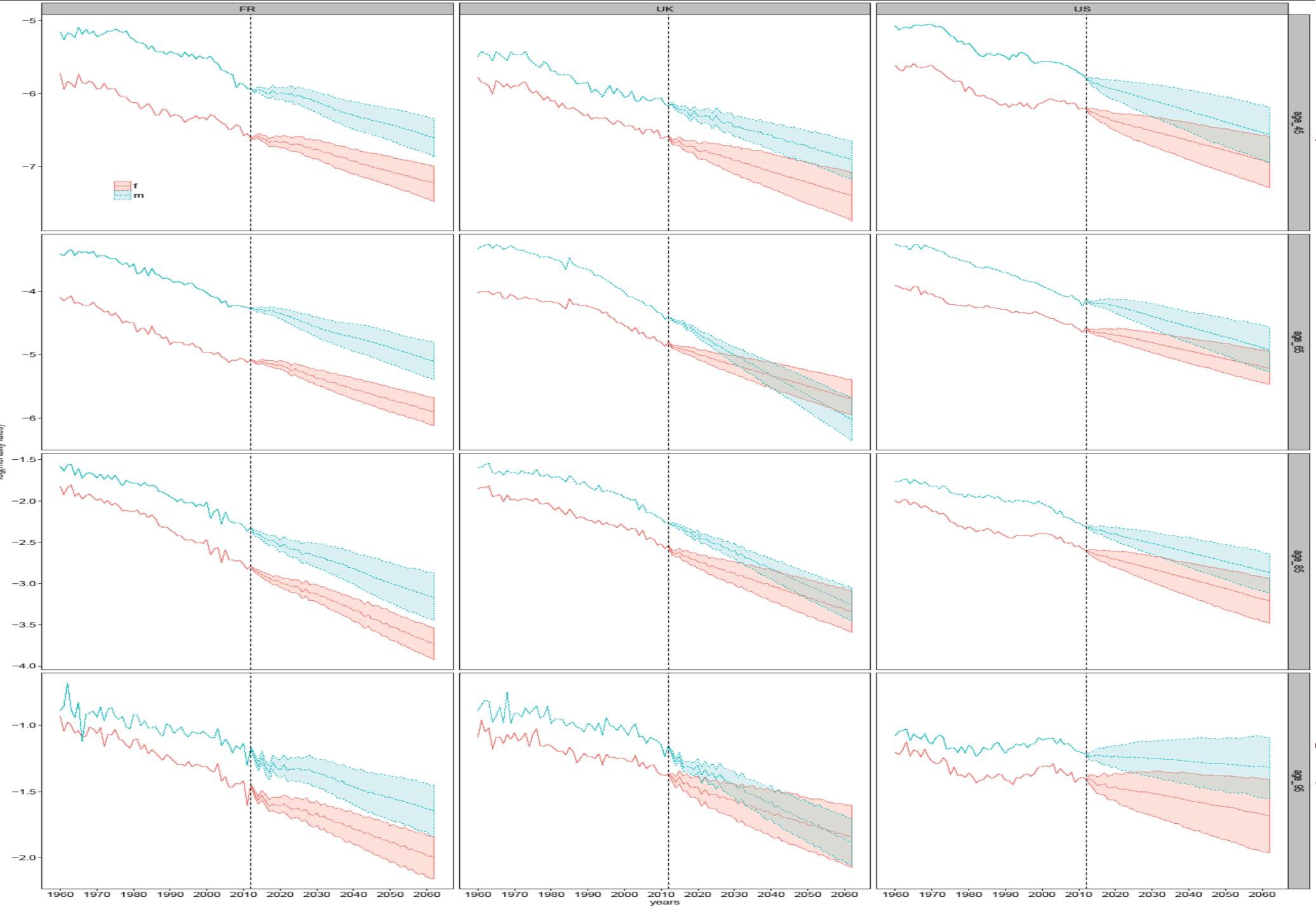
		VAR	LC	CBD	M7
US	Total	7.29	8.11	12.78	10.09
	Female	8.08	12.06	15.19	14.68
	Male	8.56	10.06	12.98	10.59
-----					
UK	Total	6.86	11.76	10.01	9.10
	Female	9.43	12.41	12.37	12.26
	Male	8.10	12.47	8.66	10.07
-----					
FR	Total	5.76	5.59	20.96	11.06
	Female	6.82	7.76	36.32	20.29
	Male	7.22	7.85	15.29	11.96

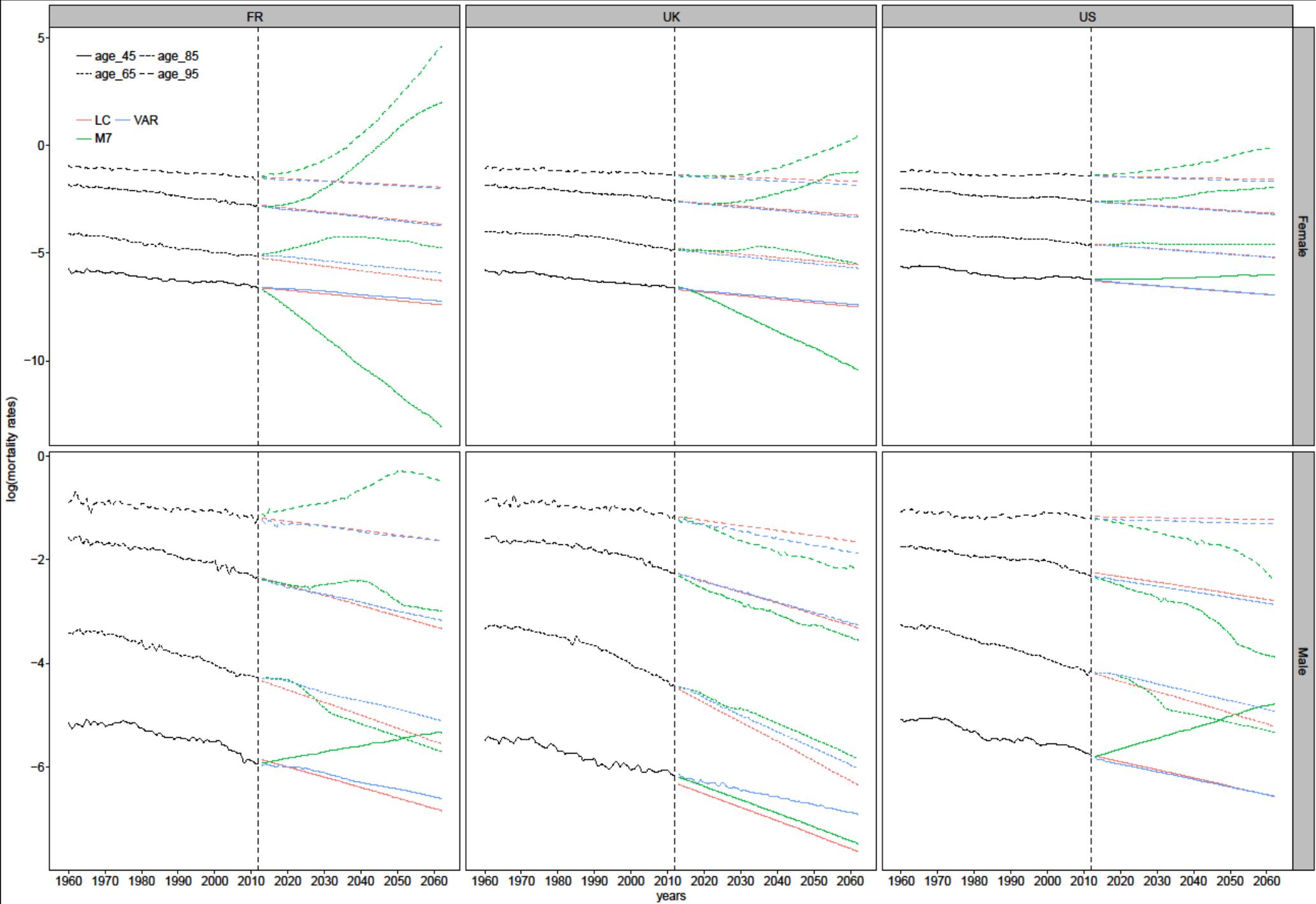
• Procédure

- Estimation des modèles sur 1980-2000
- Projection sur 2001-2012

RMSE out-sample ( $10^{-2}$ )

		VAR	LC	CBD	M7
US	Total	8.21	8.00	13.09	10.66
	Female	9.72	9.28	15.02	12.93
	Male	9.30	9.61	12.85	10.29
UK	Total	6.95	10.26	9.66	8.80
	Female	8.77	10.61	12.73	9.84
	Male	8.08	11.13	10.39	9.17
FR	Total	6.36	6.57	21.61	9.59
	Female	6.88	6.33	35.95	17.30
	Male	8.15	8.48	15.62	10.14





- Introduction
- Modèle VAR en grande dimension
- Résultats
- **Conclusion**

- Approche innovante
  - Grande liberté dans la structure spatio-temporelle de dépendance
  - Estimation en une seule étape
- Amélioration de la prédiction
  - En général par rapports aux modèles standards
  - Stabilité de l'erreur sur les différentes populations
- Mise en lumière de phénomènes démographiques
  - Effet cohorte avec les sous diagonales
  - Phénomènes peu étudiés: effets verticaux
- Adapté à la multi-population
  - Elastic-Net permet de prendre en compte un grand nombre de population

- Approche Vectorielle

- Approche identique sur  $M$  populations: concaténation ddes vecteurs

$$\Delta Y_t = (\Delta y_{1,i_{\min},t}, \dots, \Delta y_{1,i_{\max},t}, \Delta y_{2,i_{\min},t}, \dots, \Delta y_{M,i_{\max},t})^T$$

- Matrice Variance-Covariance

- Matrice var-covar peut être vu comme une matrice bloc  $\Sigma_{m,n}$

- Estimation par fonction des blocs diagonaux  $\tilde{\Sigma}_{m,m}$

- Décomposition par Cholesky  $\tilde{\Sigma}_{m,m} = \tilde{R}_m^T \tilde{R}_m$

- Estimation d'un coefficient de corrélation inter-population  $\tilde{\Sigma}_{m,n} = \rho_{m,n} \times \tilde{R}_m^T \tilde{R}_n$

- $M(2d + 1) + \frac{M(M-1)}{2}$  paramètres.

- Choix du modèle VAR
  - Différenciation pour la stationnarité
  - Perte d'information sur la relation de long-terme (cointégration)
  - VECM-ENET, mais pas encore développé
- Modélisation de l'erreur
  - Autre forme paramétrique, glasso, copule, ARCH, ...
- Choix des hyperparamètres
  - $\alpha$  peut être déterminé par cross-validation
  - Optimisation de  $p$  pour chaque population

Merci pour votre attention 😊

Des questions ?