

Apport des modèles cliométriques composites pour les tables de mortalité prospectives

Pays émergents et pays en développement

Kué Gilles GABA (ISFA, Univ. Lyon 1), Stéphane LOISEL (ISFA, Univ. Lyon 1)

Sommaire

● Cadre historique et conceptuel

- Cliométrie
- Transition démographique

● Limites des approches actuelles et originalités méthodologiques

- Profils de pays et problématiques de données de mortalité
- Problématiques méthodologiques communes à tous les pays
 - ▶ Modèle interne PCR optimal
 - ▶ Modèle interne PLS
- Problématiques spécifiques aux pays émergents et pays en voie de développement
 - ▶ Modèles mixtes cliométriques PCR-Optimal et PLS
 - ▶ Modèles relationnels (à référence externe) cliométriques

● Quelques résultats empiriques

- Modélisation pour la table prospective de l'Inde
- Modélisation pour la table prospective de l'Equateur

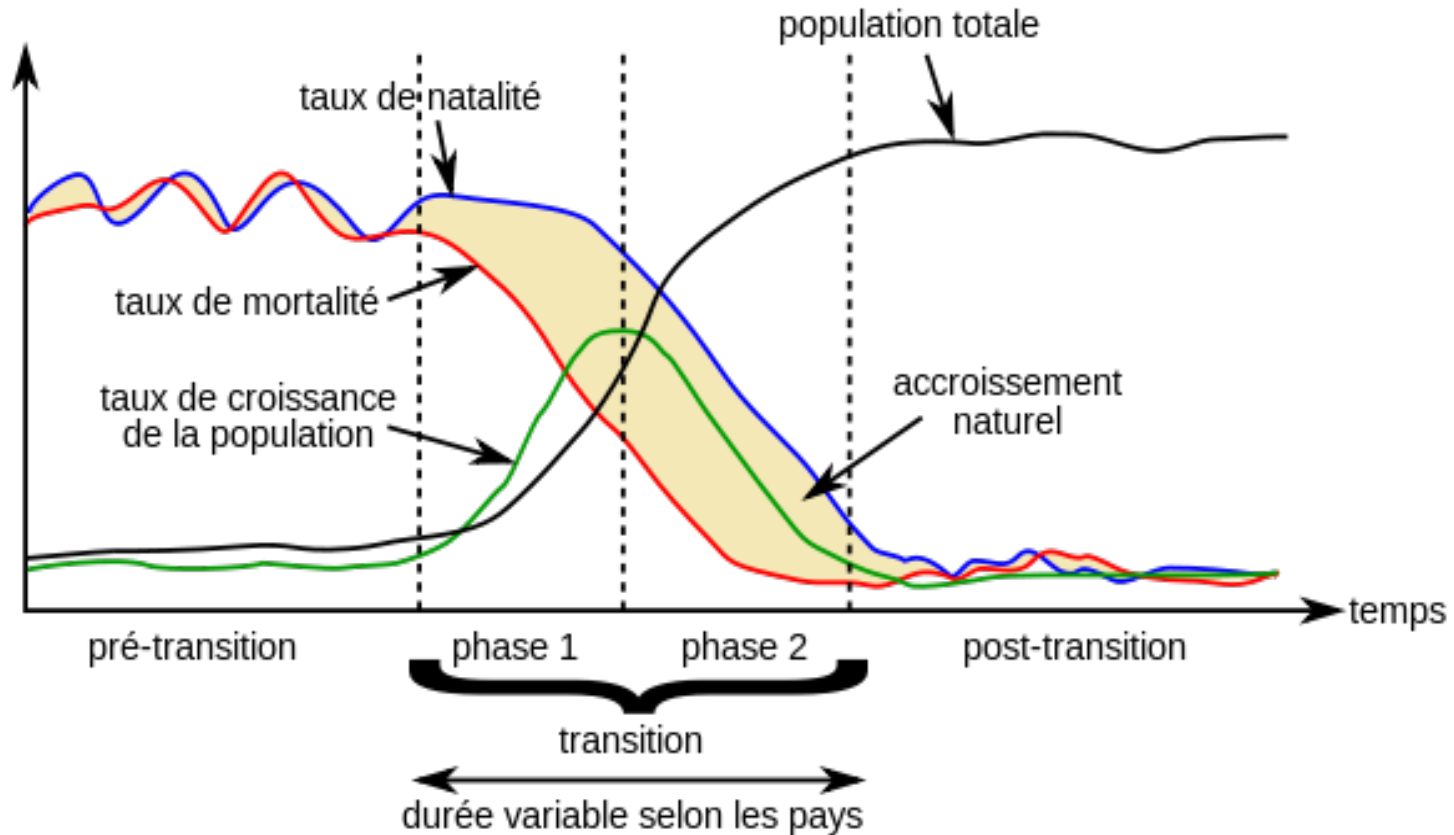
● Perspectives et travaux futurs

Cliométrie

- Clio pour « Histoire », métrie pour « mesure » : La cliométrie se définit donc comme l'étude quantitative de l'Histoire.
- Approche **multidisciplinaire, quantitative et de long terme**
 - ▶ École des Annales française (Lucien Febvre, Marc Bloch)
 - ⊕ Revue : « Annales d'histoire économique et sociale » fondée en 1929
 - ▶ USA :
 - ⊕ Revue 1963 : Explorations in Economic History
 - ⊕ Association 1983 : Cliometrics society,
 - ⊕ « Prix Nobel » d'économie 1993 : Douglass North et Robert Fogel
- Indicateurs mesurables sur le long terme (plusieurs générations) :
 - ▶ Données climatiques
 - ▶ Indicateurs démographiques (survie, reproduction, migrations,...).
 - ▶ Données sociales, économiques
- L'approche quantitative et multidisciplinaire de l'Histoire permet :
 - ▶ Une lecture des événements Historiques par rapport aux tendances de long terme
 - ▶ La détection des tendances lourdes multigénérationnelles et des schémas transitionnels
 - ▶ De faire de la prospective à long terme (multi-générationnelle) aussi bien dans les domaines démographique, sociale ou économique.

Transition démographique

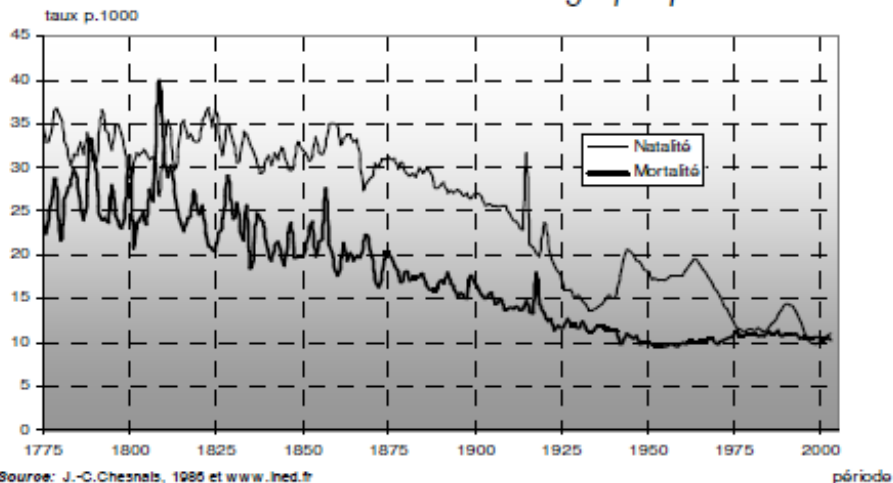
● Schéma général de la transition démographique (Avdeev A., 2014)



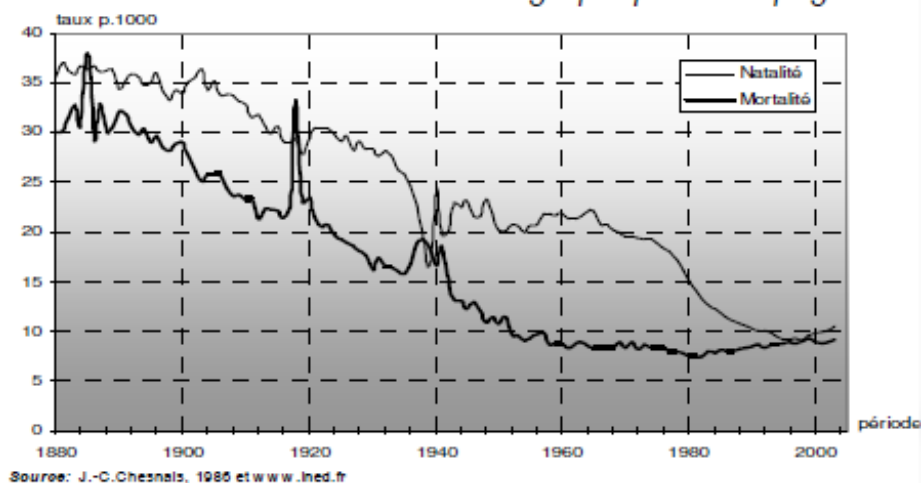
Transition démographique

✦ (Avdeev A., 2014)

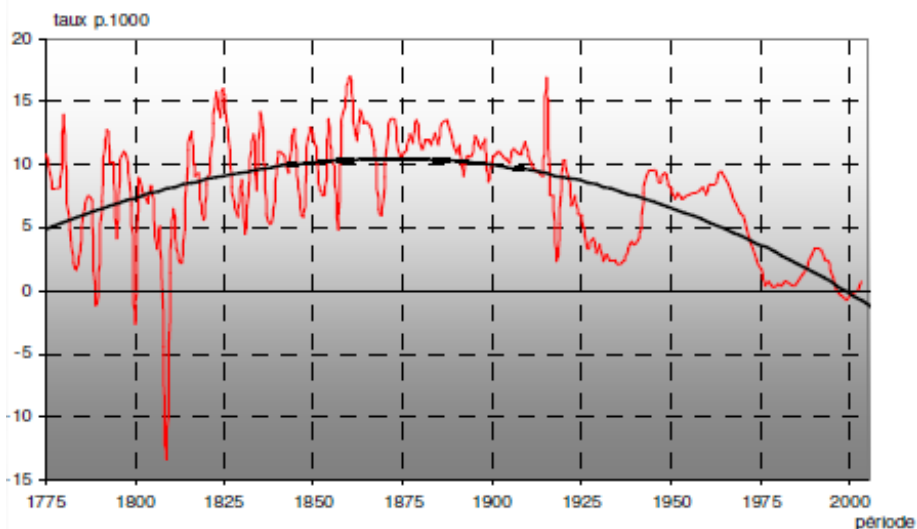
Deux cents ans de la transition démographique en Suède



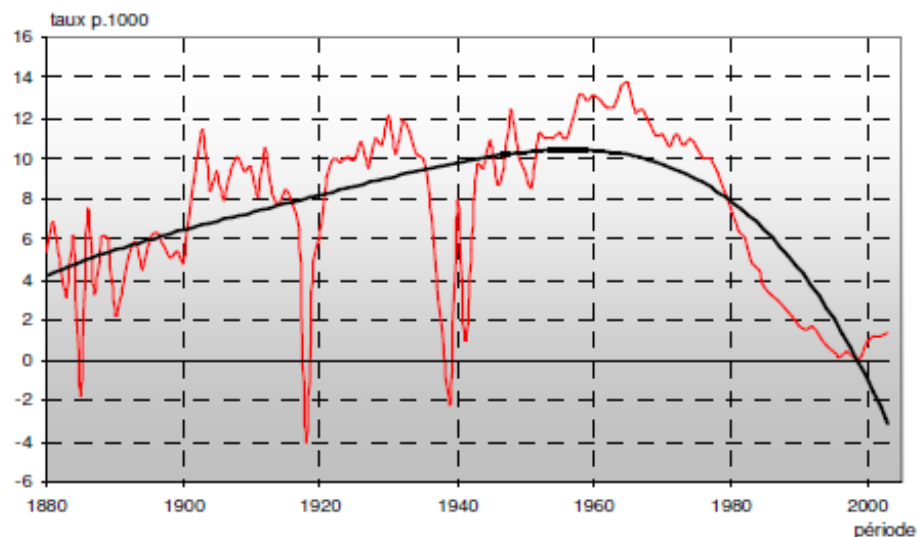
Cent ans de la transition démographique en Espagne



Taux d'accroissement naturel en Suède 1775-2004



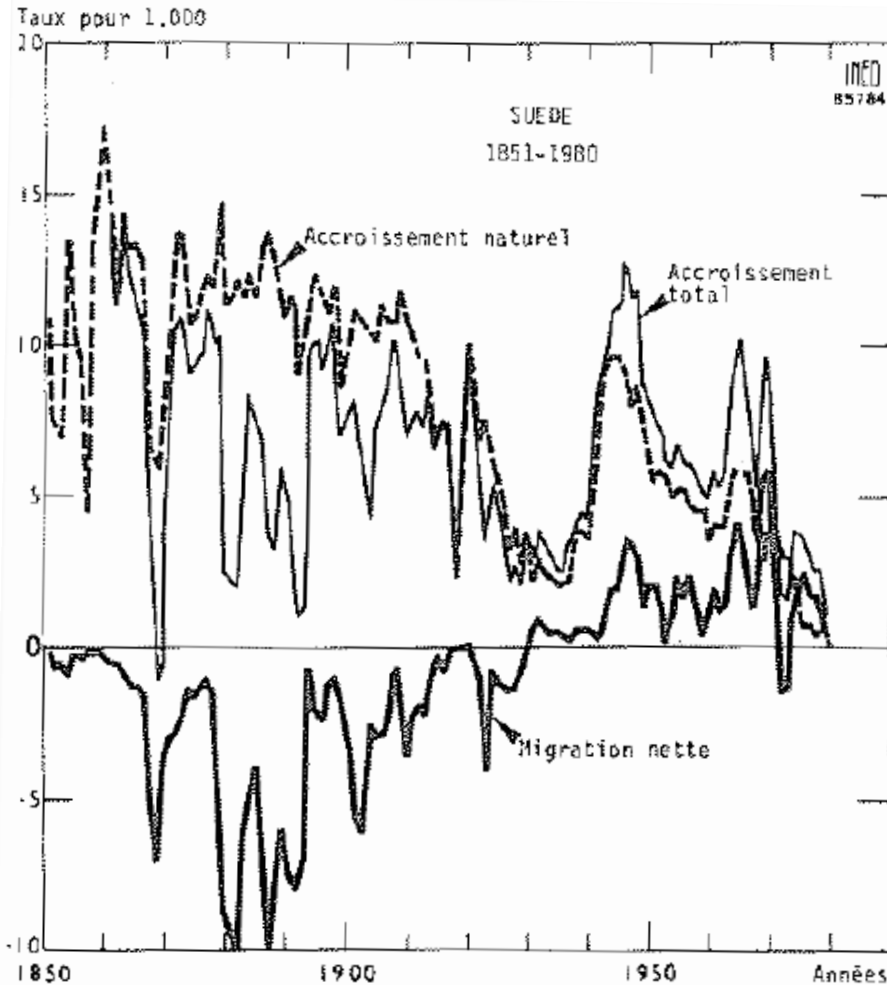
Taux d'accroissement naturel en Espagne 1880-2004



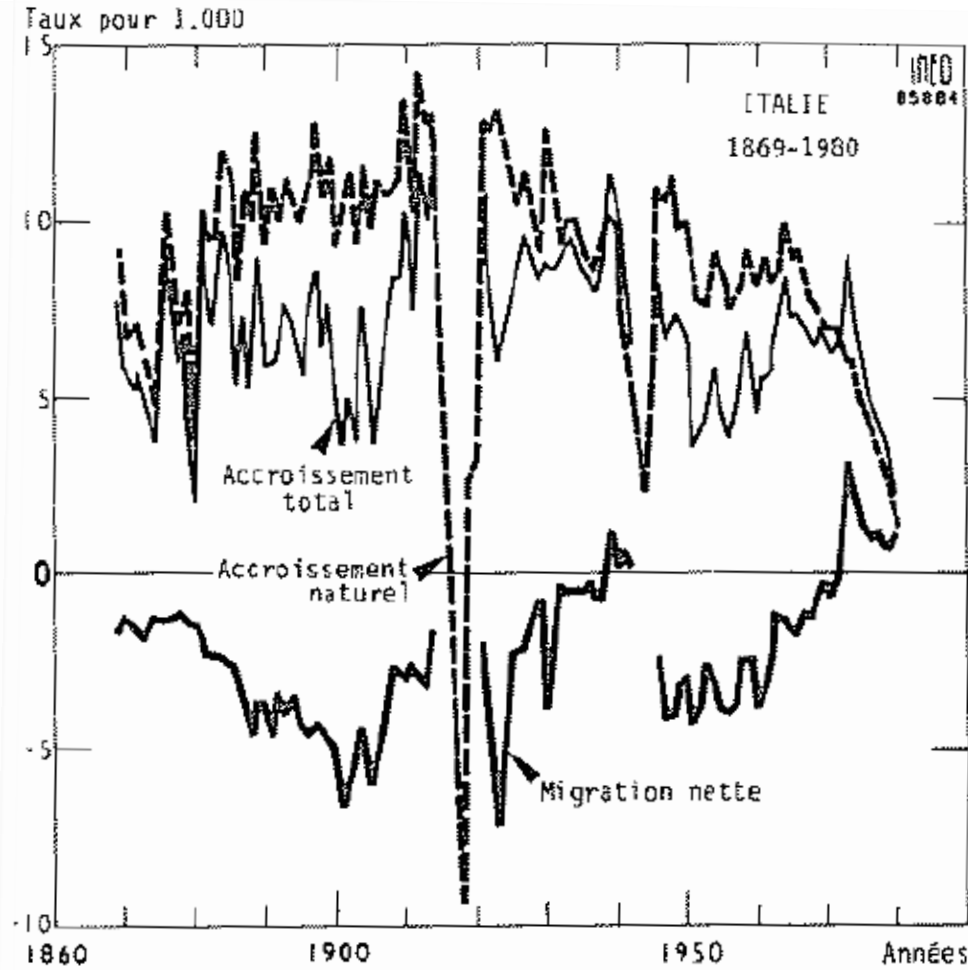
Transition démographique

- Ne pas oublier : La transition migratoire (Chesnais, 1986)

■ Suède



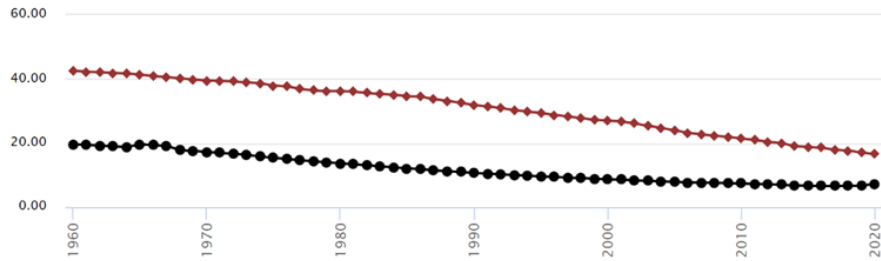
■ Italie



Transition démographique

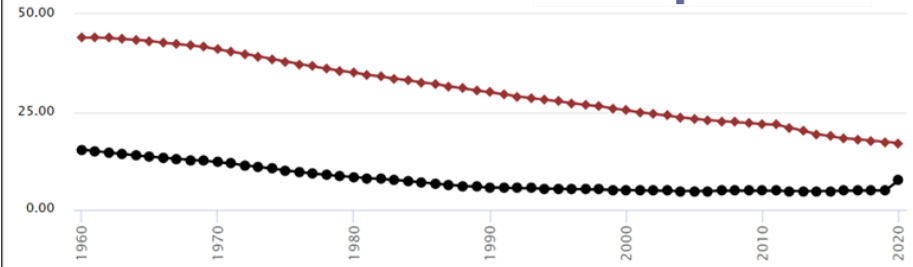
Taux de mortalité brut (par 1000 habitants), Inde
Taux de natalité brut (par 1000 habitants), Inde

Inde



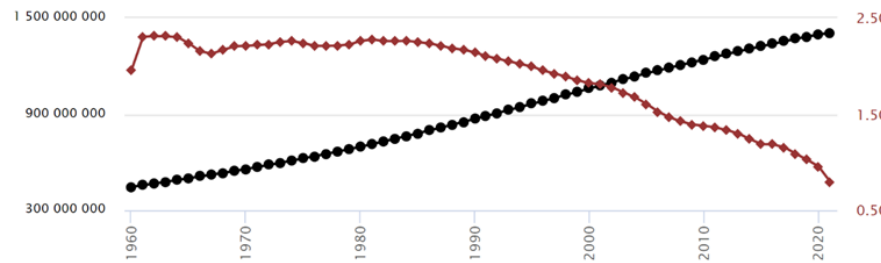
Taux de mortalité brut (par 1000 habitants), Équateur
Taux de natalité brut (par 1000 habitants), Équateur

Équateur



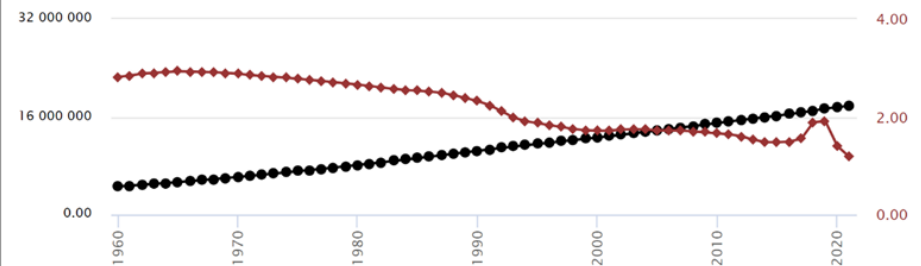
Population totale, Inde

Croissance annuelle de la population (en % de la population totale), Inde

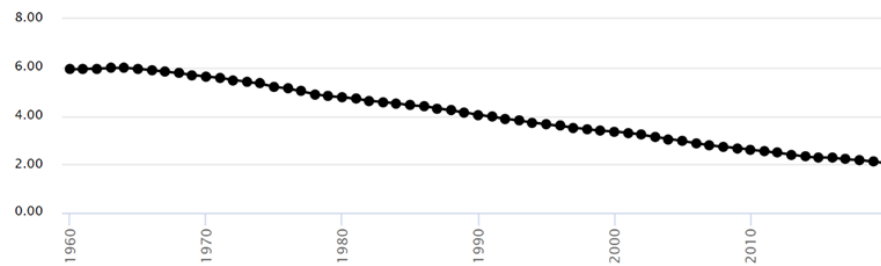


Population totale, Équateur

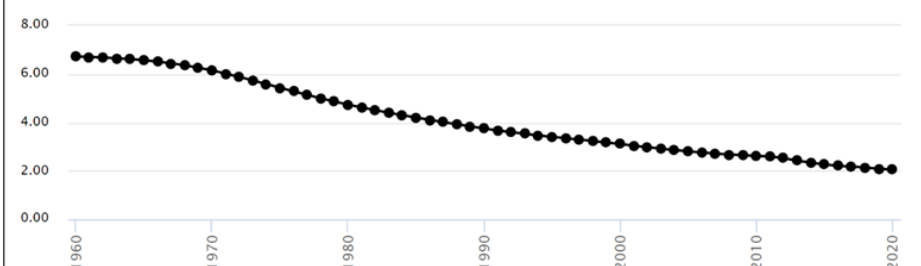
Croissance annuelle de la population (en % de la population totale), Équateur



Taux de fertilité (naissances par femme), Inde



Taux de fertilité (naissances par femme), Équateur

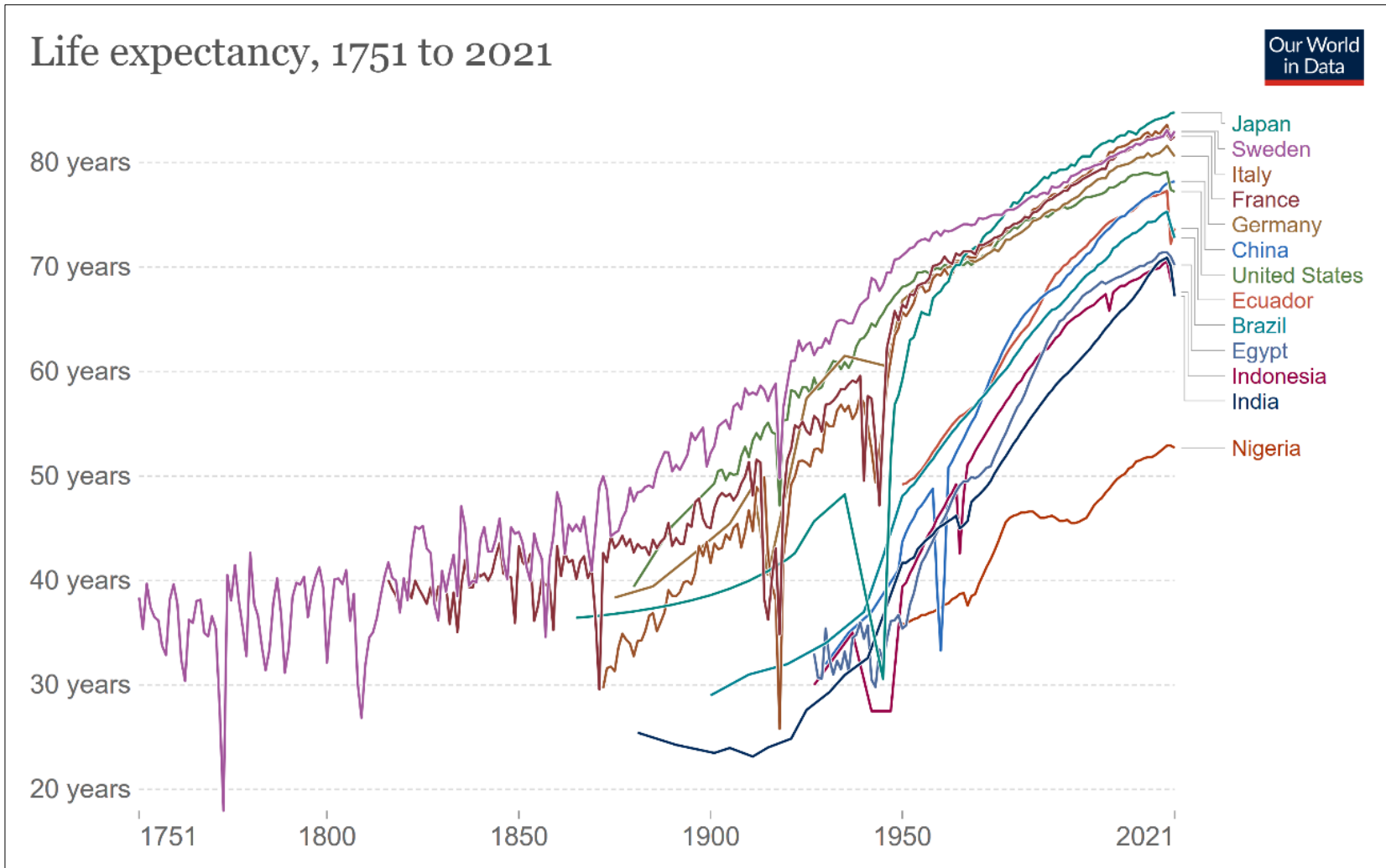


Transition démographique



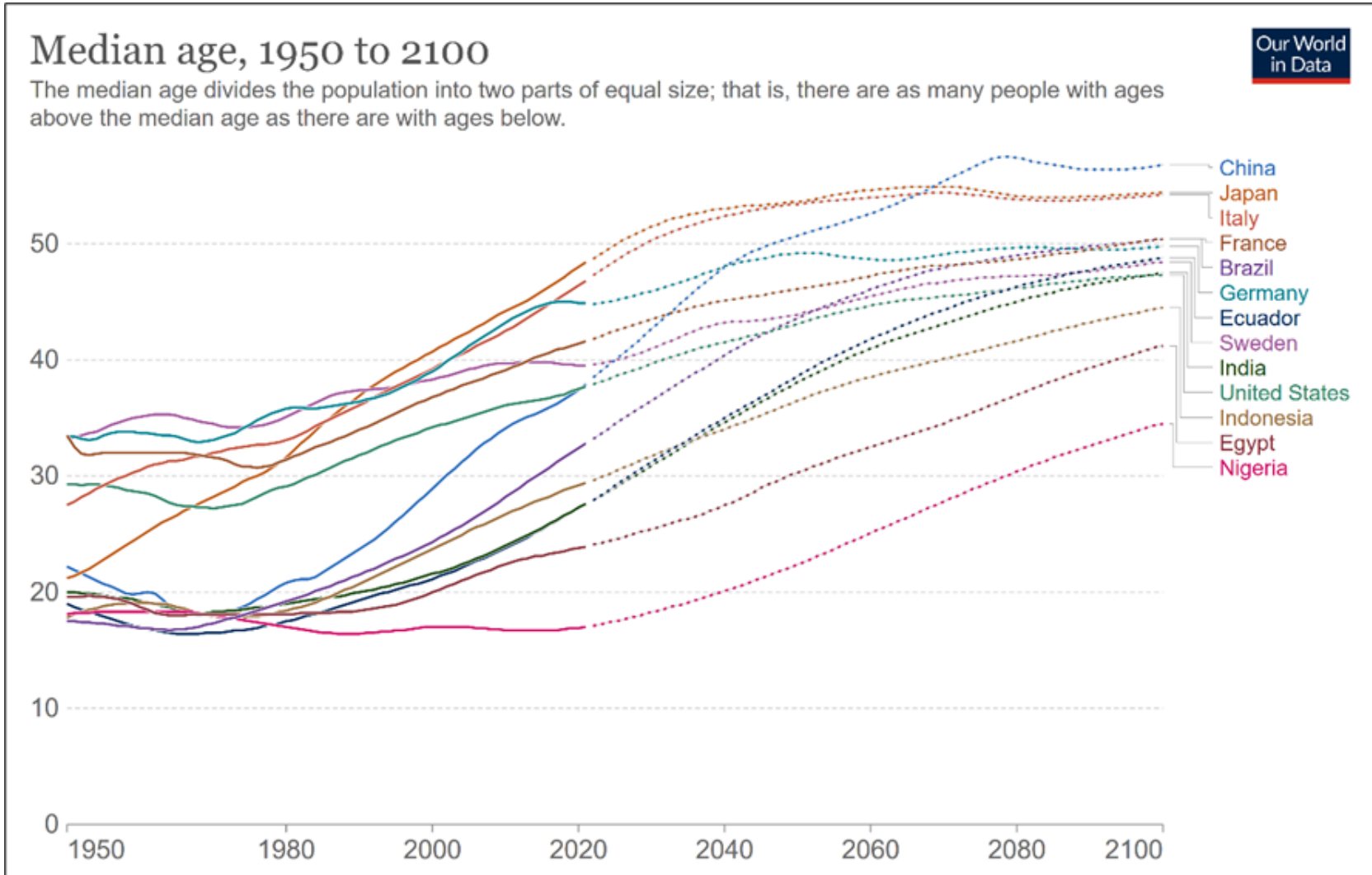
Longévité

■ Espérance de vie à 10 ans



Transition démographique

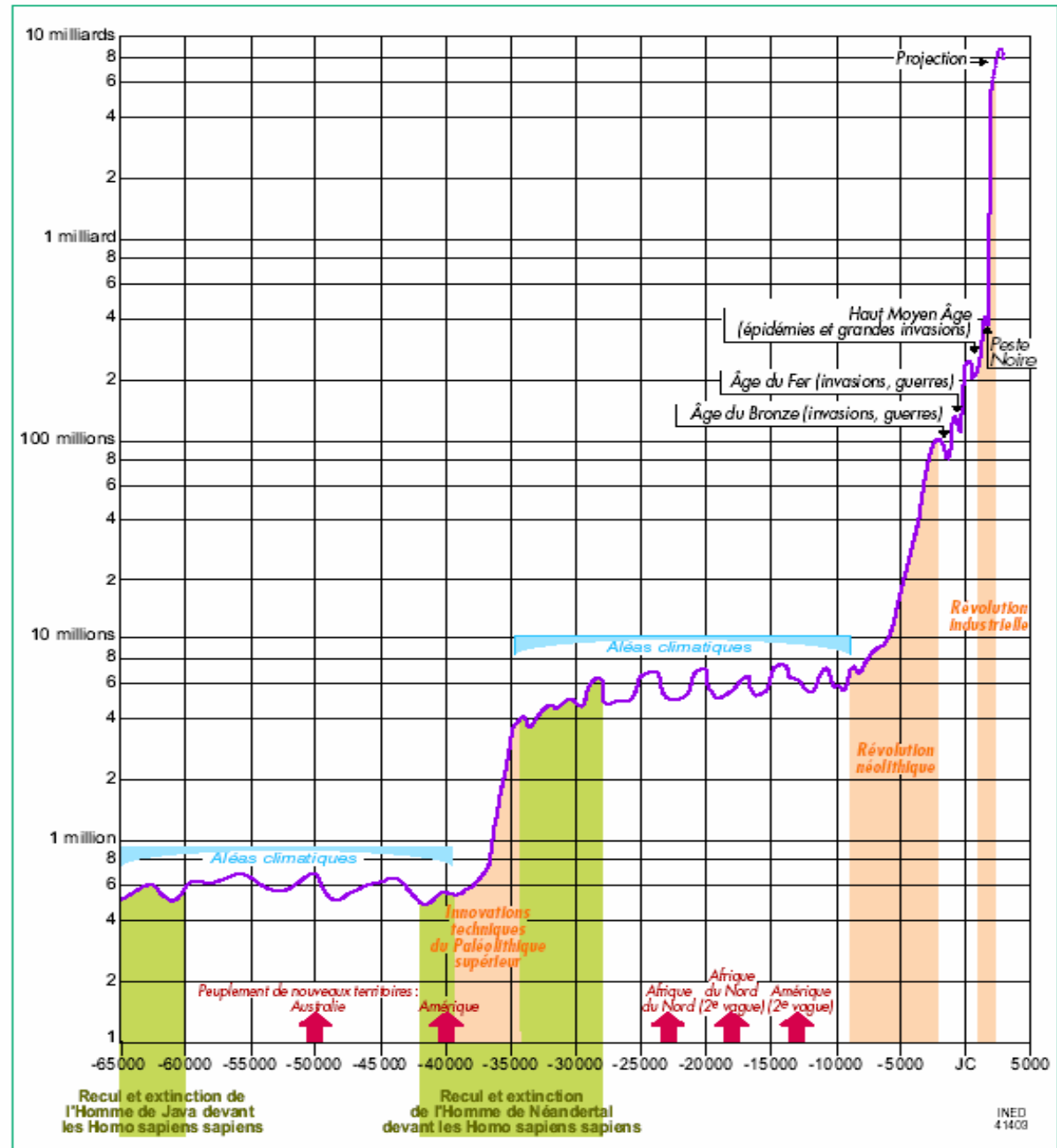
Vieillesse



Transition démographique : Antécédents

Antécédents historiques

D'après « L'évolution du nombre des hommes », Jean-Noel Biraben, Population & Sociétés, n° 394, INED, octobre 2003.



Profils de pays et problématique de données de mortalité

● Pays développés

- ▶ Globalement l'OCDE
- ▶ Historiques de plusieurs générations (souvent plus de 4 générations)
- ▶ Instituts de statistiques produisant des données standardisées et de fiabilité « acceptable »
- ▶ Bases de données internationales : HMD, World Pop. Prospects (WPP)

● Pays émergents

- ▶ Critères du FMI : PIB par habitant, PIB, l'ouverture aux échanges et la stabilité financière
- ▶ BRICS, Indonésie, Malaisie, Mexique, Philippines, Thaïlande, Turquie ...
- ▶ Historique et fiabilité variables selon les pays
- ▶ Bases de données internationales : Human Life-table Database, WPP (?)

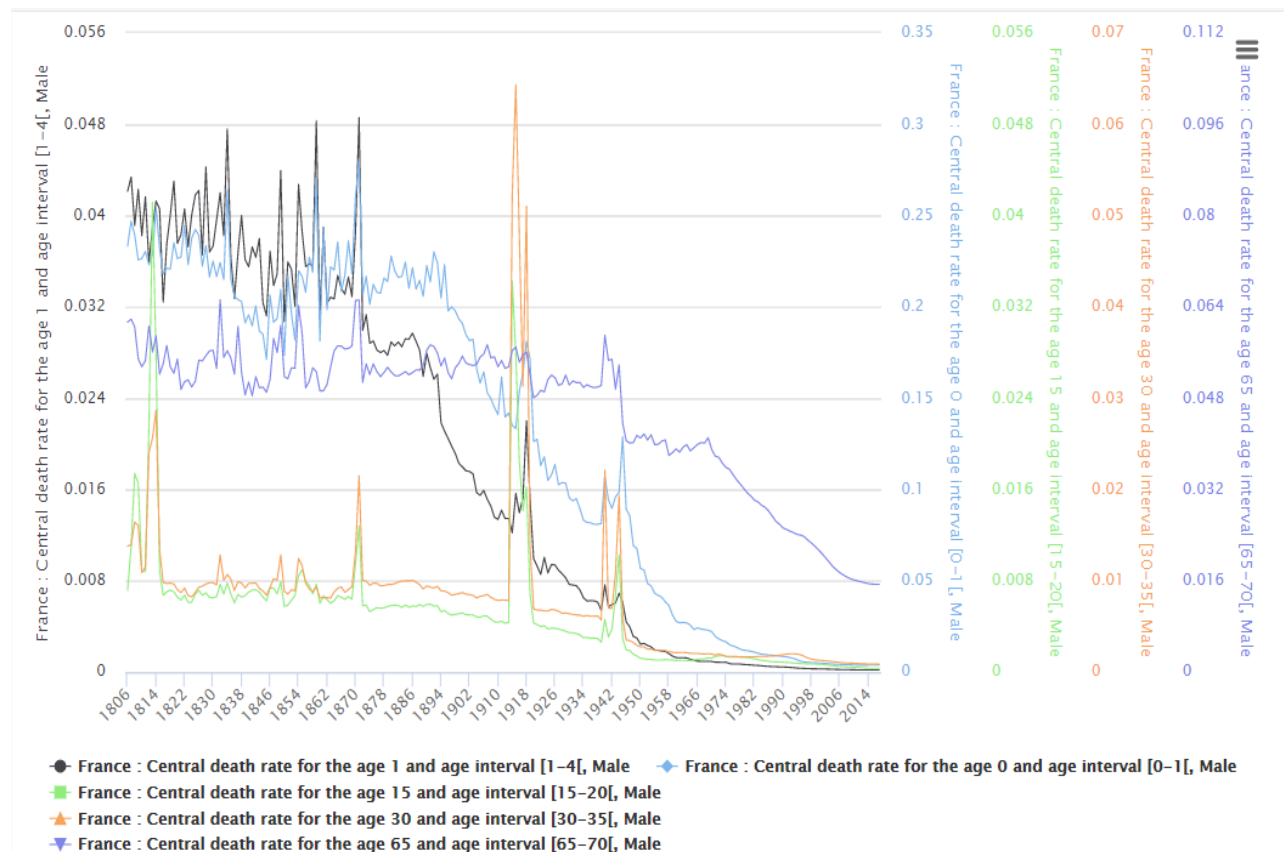
● Pays en voie de développement et pays les moins avancés

- ▶ Historiques réguliers très récents (années 2000)
 - ✦ Exception importante : mortalité infantile
- ▶ Fiabilité et standardisation variables mais en cours d'amélioration
- ▶ Bases de données internationales : HLD, WPP (?)

Problématiques méthodologiques communes à tous les pays

■ Constats empiriques 1

- ▶ **Décalages temporels** et des **différences d'allure** dans les courbes des taux de mortalité selon les âges
- ▶ Si une composante temporelle unique reflète la transition démographique du point de vue de la mortalité globale, la **prise en compte de la transition sanitaire qui affecte les âges de manière différenciée**, nécessite des composantes temporelles différenciées par tranche d'âge



Problématiques méthodologiques communes à tous les pays

■ Constat méthodologique 1

- ▶ Pour chaque âge, le modèle de Lee-Carter est une régression
 - ⊕ **sur la 1ère composante principale** de la table des séries temporelles centrées des mortalités par âge
 - ⊕ **à un facteur multiplicatif près**, selon la contrainte posée sur les paramètres

■ Originalité méthodologique 1

- ▶ Généraliser les modèles de mortalité classiques (Lee-Carter et ses variantes),
 - ⊕ en sélectionnant de manière optimale (minimisation du MSE de validation croisée), une ou plusieurs composantes principales (régression PCR) ou composantes PLS (régression PLS),
 - ⊕ pour chaque âge modélisé
- ▶ Deux types de modèles internes **multifactoriels** :
 - ⊕ Modèles internes à facteurs PCR-optimal
 - ⊕ Modèles internes à facteurs PLS

Problématiques méthodologiques communes à tous les pays

■ Constat méthodologique 2

- ▶ **Même classe de modèles** appliquée à la modélisation de toutes les courbes des taux de mortalité par âge
- ▶ Classe de modèles : ensemble de modèles caractérisés par la même spécification sur les critères suivants
 - ⊕ Ensemble des séries explicatives initiales
 - ⊕ Procédure de sélection des séries explicatives
 - ⊕ Formulation mathématique
 - ⊕ Loi et méthode d'estimation

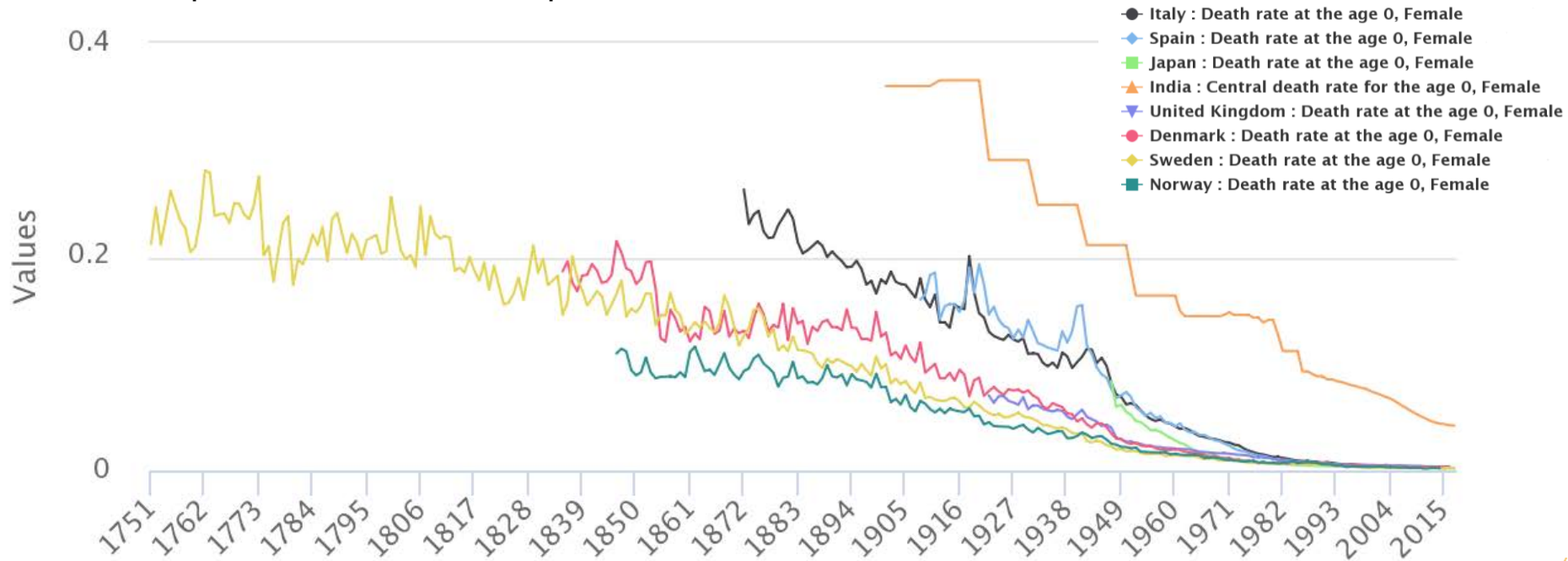
■ Originalité méthodologique 2

- ▶ Séparer méthodologiquement les modélisations des mortalités des âges
 - ⊕ Possibilité d'attribuer différentes classes de modèles à différents âges
- ▶ Modèle développé : modèle interne **composite** à facteurs PCR-Optimal/PLS
- ▶ Stratégie : modélisation en 2 itérations pour calculer les performances sans biais
 - ⊕ Itération 1 : Choix **a posteriori** du Best Model Class pour chaque âge
 - Historique : $t_0 ; t_1$
 - Prévisions : $t_1 + 1 ; t_2$
 - ⊕ Itération 2 : Pour chaque âge, utilisation **a priori** du Best Model Class de l'itération 1
 - Historique : $t_0 ; t_2$
 - Prévisions : $t_2 + 1 ; t_3$

Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

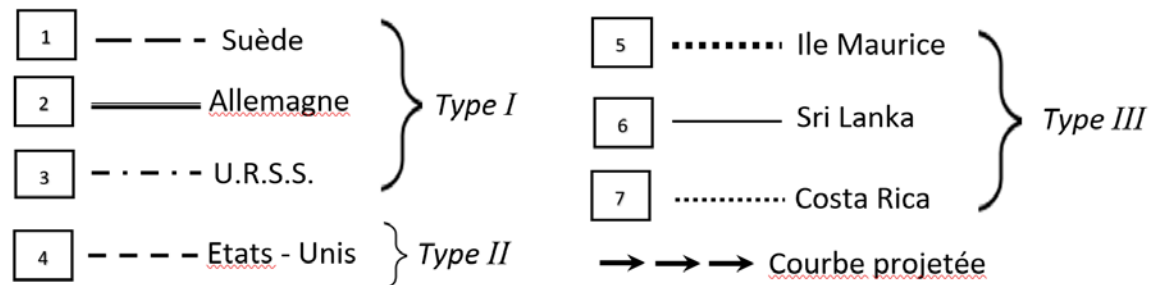
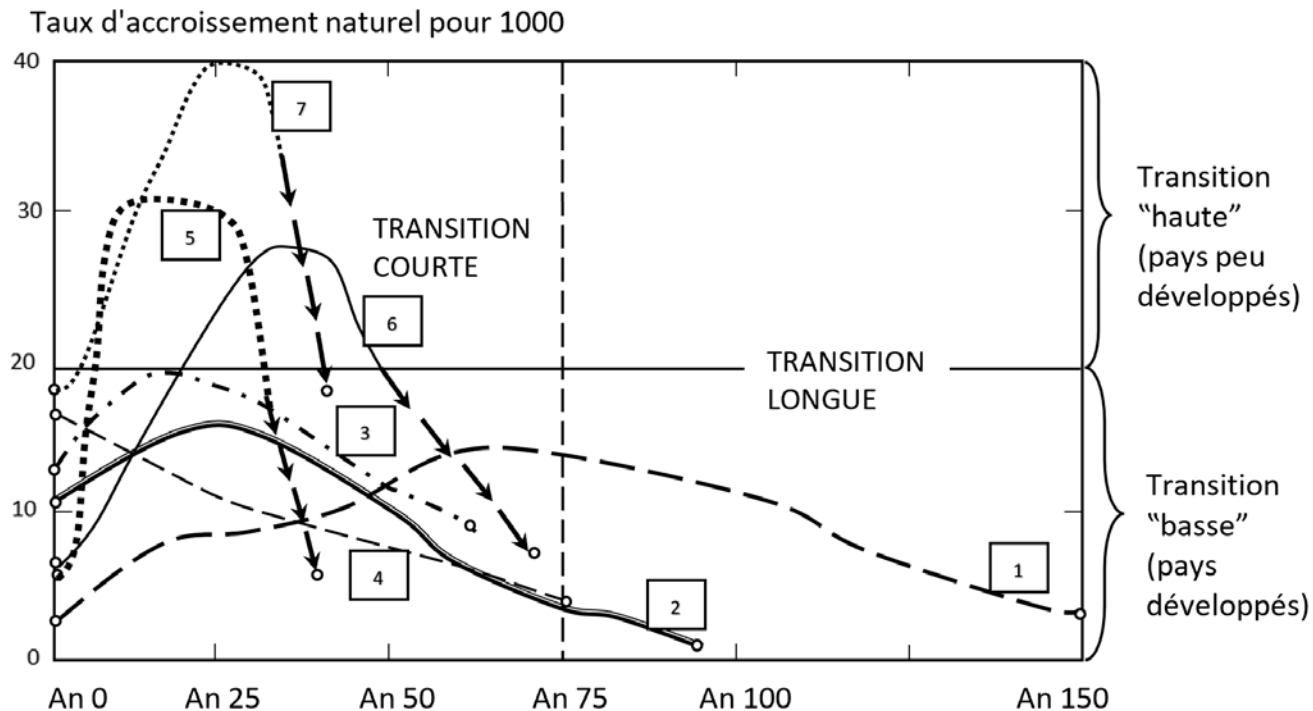
■ Constats empiriques 3

- ▶ Transition démographique universelle, se déroulant par vagues (cohortes) de pays
- ▶ Rythmes des changements transitionnels hétérogènes d'une cohorte de pays à l'autre
 - ⊕ Pays les plus récemment entrés en transition avec changements plus rapides
- ▶ Difficulté à faire des prévisions en utilisant des modèles multivariés classiques de séries temporelles avec 'variable' exogène retardée, car le retard entre les séries n'est pas stable dans le temps



Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

■ Constats empiriques 3 (suite) : Profils de transition démographique



Source : Chesnais (1986), p. 270

Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

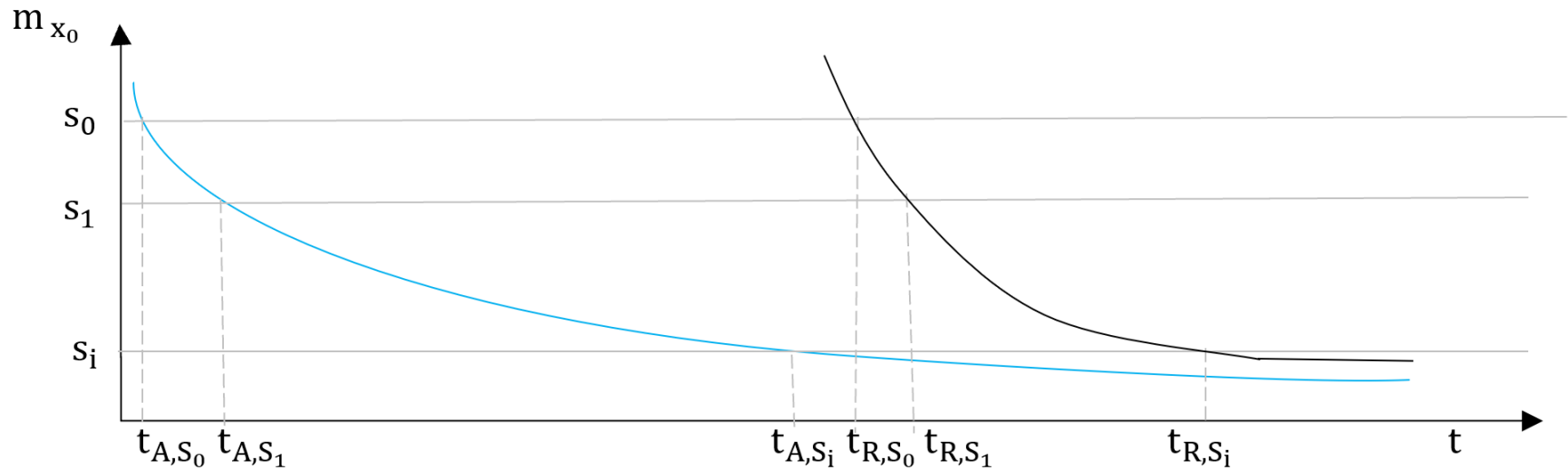
■ Originalité méthodologique 3 : Pays ayant un historique **suffisant** pour les modèles internes de prospective

- ▶ Utiliser l'histoire quantitative des mortalités des **pays transitionnellement plus âgés (A)**, pour tenter d'améliorer les prévisions à long terme de mortalité dans les **pays transitionnellement moins âgés (R)**
- ▶ **Série cliométrique de mortalité à l'âge x du pays R par rapport au pays A ($C_{R,A,x}$)** : série créée à partir des données historiques du pays A (transitionnellement plus âgé que le pays R), **pour jouer un rôle de série exogène dans les modèles de prévisions de mortalité du pays R**
- ▶ Cette série cliométrique exogène de mortalité sera construite de sorte à avoir la **même allure temporelle que dans le pays A mais adaptée au rythme du pays R à modéliser**
- ▶ La série cliométrique sera ainsi utilisée comme série explicative **supplémentaire** dans les modèles prospectifs de mortalité du pays R
- ▶ Modèle développé : modèle **mixte cliométrique** et composite à facteurs PCR-optimal/PLS
 - ⊕ Modèle mixte : **à la fois interne et externe** (via l'utilisation de l'historique du pays A qui joue le rôle de référence externe)
 - ⊕ Modèle composite : les **âges sont modélisés de façon indépendante** et peuvent donc recourir à des classes de modèles diversifiées (choix du ou des pays A, facteurs PCR-O ou PLS)

Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

Originalité méthodologique 3 (suite)

► Courbes de mortalité des pays A (bleue) et R (noire)



Les s_i sont des valeurs seuils du taux de mortalité m_{x_0} ,

$t_{A,s_i} = \inf \{t \in \mathbb{N}, m_{A,x_0,t} \leq s_i\}$ et $t_{R,s_i} = \inf \{t \in \mathbb{N}, m_{R,x_0,t} \leq s_i\}$.

Soient $t_A = (t_{A,s_i}), i \in \mathbb{N}$ et $t_R = (t_{R,s_i}), i \in \mathbb{N}$ deux séries temporelles des dates calendaires d'atteinte des valeurs de mortalités à l'âge x_0 , $(s_i), i \in \mathbb{N}$, respectivement dans les pays A et R.

Les points de la série *explicative cliométrique* à l'âge x du pays R par rapport au pays A ($C_{R,A,x}$) sont définis par les couples $(t_{R,s_i}, m_{A,x,t_{A,s_i}}), i = 0, \dots, r$:

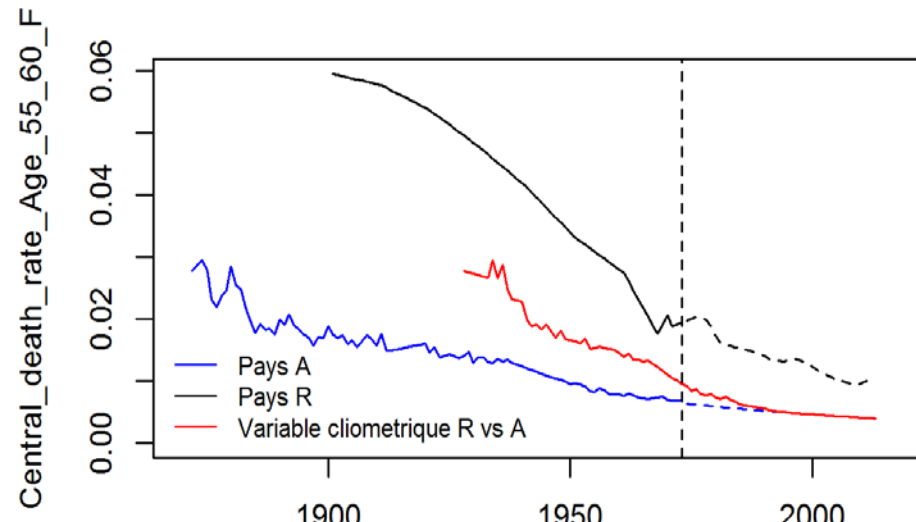
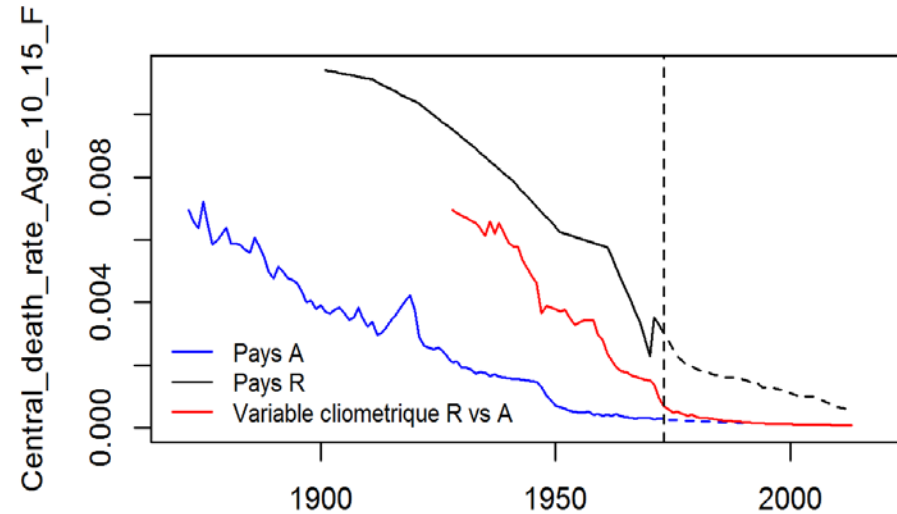
$$C_{R,A,x,t_{R,s_i}} : t_{R,s_i} \mapsto m_{A,x,t_{A,s_i}}$$

Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

■ Originalité méthodologique 3 (suite)

► Pays A : Italie

► Pays R : Inde



Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

■ Originalité méthodologique 4 : Pays ayant un historique **insuffisant** pour les modèles internes de prospective

- ▶ Utiliser l'histoire quantitative des mortalités des **pays transitionnellement plus âgés (A)**, pour tenter d'améliorer les prévisions à long terme de mortalité dans les **pays transitionnellement moins âgés (R)**
 - ⊕ En exploitant le fait que la **mortalité infantile** a un **historique relativement** long dans les pays en développement
- ▶ **L'appariement** entre le **pays d'expérience R** et le **pays de référence A** :
 - ⊕ 1. permet de créer plusieurs séries cliométriques pour chaque âge du pays R (par rapport à un ou plusieurs pays A, avec divers paramétrages de sexe ou d'âge).
 - ⊕ 2. Ces séries cliométriques qui ont par construction des valeurs dans le passé et dans le futur, sont des séries explicatives potentielles dans les modèles **externes cliométriques** pour le pays d'expérience R

Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

■ Originalité méthodologique 4 : Pays ayant un historique **insuffisant** pour les modèles internes de prospective

► Modèles externes cliométriques adaptés des modèles externes classiques

- ⊕ Nous proposons les 3 classes de modèles externes cliométriques suivants, adaptés des modèles classiques externes :

- **Modèle externe cliométrique adapté du modèle externe de BRASS**

$$\text{logit} \left(q_{x,t_{R,s_i}}^{exp} \right) = \theta_1 + \theta_2 \times \text{logit} \left(q_{x,t_{A,s_i}}^{ref} \right)$$

- **Modèle externe cliométrique adapté du modèle externe de COX**

$$\mu_{x,t_{R,s_i}}^{exp} = \theta \times \mu_{x,t_{A,s_i}}^{ref}, \theta > 0$$

- **Modèle externe cliométrique adapté du modèle externe de TGH05-TGF05**

$$\text{logit} \left(q_{x,t_{R,s_i}}^{exp} \right) = a_x \text{logit} \left(q_{x,t_{A,s_i}}^{ref} \right) + b_x$$

- ⊕ Les trois modèles ci-dessus peuvent être combinés dans le modèle composite suivant :

- **Modèle externe cliométrique composite mélangeant les adaptations de BRASS/COX/TGH05-TGF05**

Problématique méthodologique spécifique aux pays émergents et pays en voie de développement

■ Originalité méthodologique 4 : Pays ayant un historique **insuffisant** pour les modèles internes de prospective

▶ Modèles externes cliométriques adaptés des modèles multifactoriels

- ⊕ Nous proposons les 2 classes de modèles externes cliométriques suivants, adaptés des modèles à facteurs PCR-Optimal ou PLS :
 - **Modèle externe cliométrique à facteurs PCR-Optimal** (construits à partir des séries cliométriques)
 - **Modèle externe cliométrique à facteurs PLS** (construits à partir des séries cliométriques)
- ⊕ Les trois modèles ci-dessus peuvent être combinés dans le modèle composite suivant :
 - **Modèle externe cliométrique composite à facteurs PCR-Optimal/PLS**

Quelques résultats empiriques : modèles mixtes cliométriques composites à facteurs PCR-Optimal/PLS

- **Modélisation des mortalités pour la table prospective de l'Inde**
- **Modélisation des mortalités pour la table prospective de l'Equateur**
- **Les modèles dont les performances sont comparées sont :**
 - **Modèle Lee-carter (LC),**
 - **Modèle LC Log-Poisson,**
 - **Modèle CBD Log-Poisson,**

 - **Modèle mixte cliométrique composite à facteurs PCR-Optimal/PLS (Cliometric composite model)**

Modélisation des mortalités pour la table prospective de l'Inde

Paramétrages associés au pays « R » transitionnellement « moins âgé » (ou pays d'expérience) à modéliser

- Pays « R » : Inde
- Source des données : Les données par tranche d'âge de l'Inde et l'Italie sont issues de la base de données Human Life-table Database (HLD) publiées par la Max Planck Institute for Demographic Research (2023).
- Variable étudiée : Probabilité de décès (femmes)
- Liste des transformations préalables (moyennes mobiles) : "Interpolation + CMA5"
- Ages étudiés : Age 0, et tranches d'âge de 1 à 80 ans

Modélisation des mortalités pour la table prospective de l'Inde

Itérations :

- Itération 1 : Détection **a posteriori** de la meilleure classe de modèle (Best Model Class) pour chaque âge modélisé
 - Début de l'historique étudié : 1930
 - Année actuelle supposée : 1965
 - Horizon de prévision : 1990 (25 ans)
- Itération 2 : Utilisation **a priori** de la meilleure classe de modèle (Best Model Class) pour chaque âge modélisé
 - Début de l'historique étudié : 1930
 - Année actuelle supposée : 1990
 - Horizon de prévision : 2015 (25 ans)

Modélisation des mortalités pour la table prospective de l'Inde

Paramétrage et performances du meilleur modèle obtenu

Paramétrages d'exploration associés aux pays de référence A

Lors de l'itération 1 de chaque âge modélisé, l'exploration consiste à tester chaque combinaison du produit cartésien des ensembles ci-dessous. Chaque ensemble désigne un paramètre et ses éléments sont les modalités du paramètre.

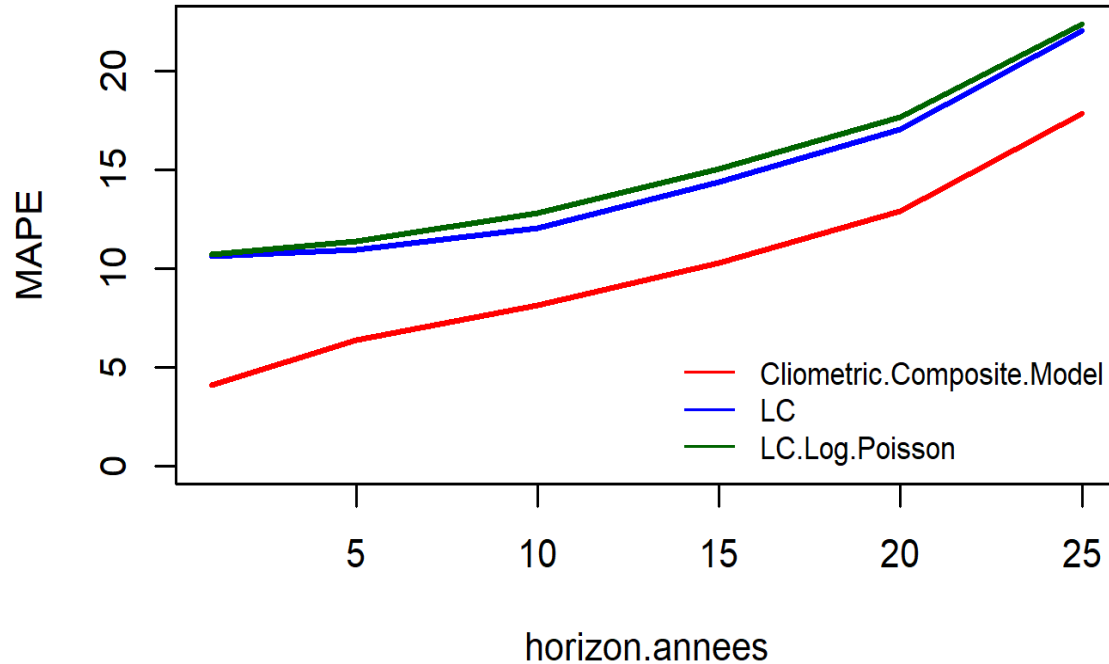
Pour chaque paramètre les modalités à tester sont :

- Liste des vecteurs de pays « A » : ("Italy"), ("Spain")
- Variable étudiée : Probabilité de décès
- Ages étudiés : Age 0, et tranches d'âge de 1 à 80 ans
- Liste de vecteurs de sexes testés : ("M"), ("F"), ("MF")
- Liste de vecteurs d'âges d'appariement testés : Age 0
- Liste des transformations préalables (moyennes mobiles) : "Interpolation + CMA5"

Source des données : Les données par tranche d'âge de l'Inde et l'Italie sont issues de la base de données Human Life-table Database (HLD) publiées par la Max Planck Institute for Demographic Research (2023).

Modélisation des mortalités pour la table prospective de l'Inde

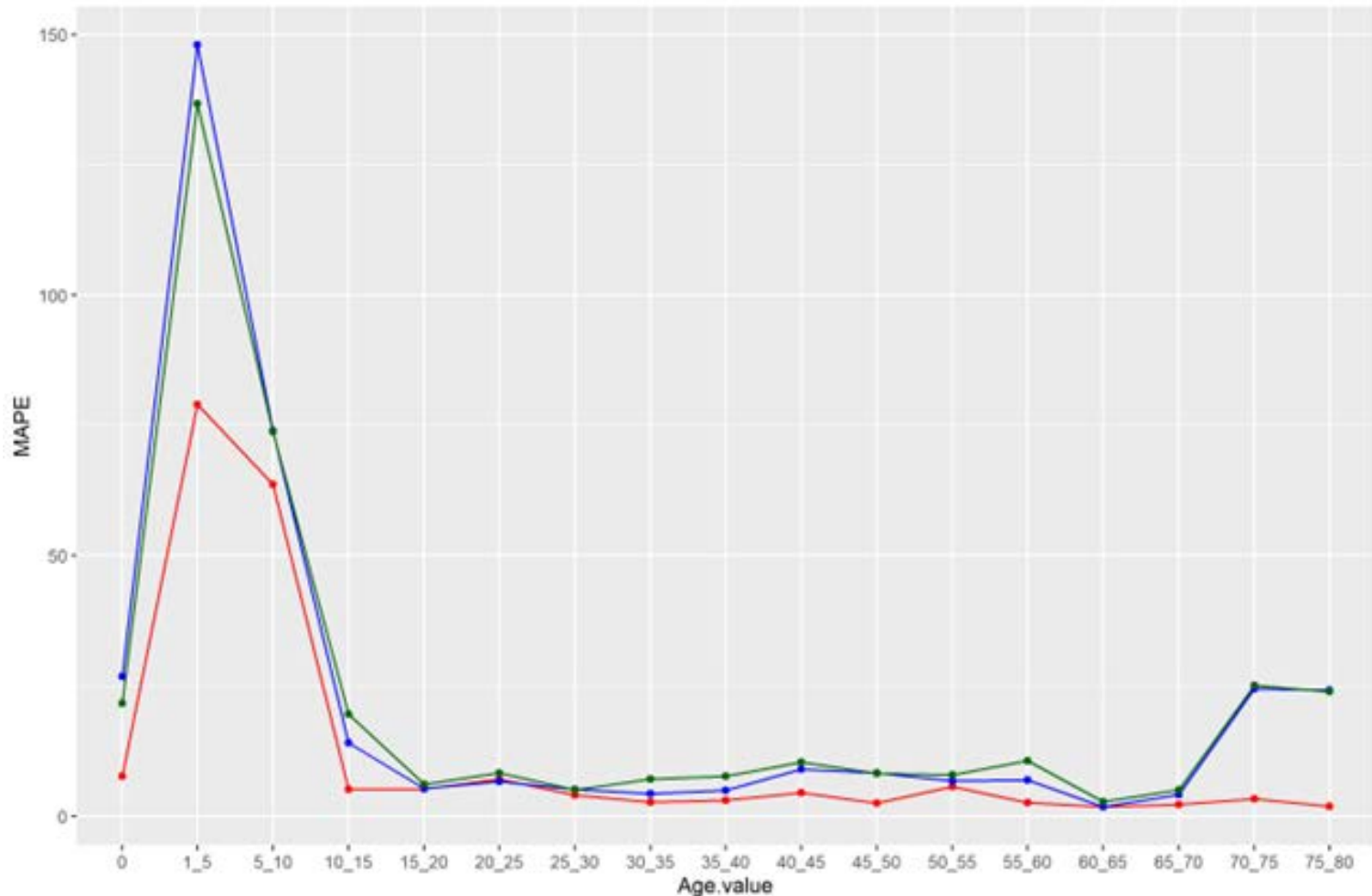
Performance des modèles selon l'horizon de prévision.



MAPE (%)	Horizons					
Modèles	1	5	10	15	20	25
CBD.log.Poisson	103,7	102,6	100,8	102,2	105,1	108,3
Cliometric.Composite.Model	4,1	6,4	8,2	10,3	12,9	17,9
LC	10,6	11,0	12,0	14,4	17,0	22,1
LC.Log.Poisson	10,7	11,4	12,8	15,0	17,6	22,4

Modélisation des mortalités pour la table prospective de l'Inde

MAPE tous horizons confondus, selon les différents âges modélisés (sans CBD)



Modélisation des mortalités pour la table prospective de l'Inde

MAPE tous horizons confondus, selon les différents âges modélisés

MAPE (%)	Modèles			
Âges	CBD.log.Poisson	Cliometric.Composite.Model	LC	LC.Log.Poisson
0	97,0	7,7	26,8	21,7
1_5	63,3	79,0	148,0	136,7
5_10	59,1	63,7	74,0	73,8
10_15	201,5	5,2	14,1	19,6
15_20	123,5	5,2	5,4	6,1
20_25	113,1	7,1	6,7	8,3
25_30	156,3	4,1	5,2	5,0
30_35	196,6	2,7	4,3	7,2
35_40	214,2	3,1	4,9	7,7
40_45	211,6	4,5	9,1	10,4
45_50	152,0	2,5	8,4	8,3
50_55	98,2	5,7	6,8	8,0
55_60	53,9	2,6	6,9	10,7
60_65	12,1	1,8	1,8	2,8
65_70	13,4	2,3	4,2	5,1
70_75	32,6	3,4	24,4	25,1
75_80	43,4	1,9	24,2	23,9

Modélisation des mortalités pour la table prospective de l'Equateur

Paramétrages associés au pays « R » transitionnellement « moins âgé » (ou pays d'expérience) à modéliser

- Pays « R » : Equateur
- Source des données : Les données par tranche d'âge de l'Inde et l'Italie sont issues de la base de données Human Life-table Database (HLD) publiées par la Max Planck Institute for Demographic Research (2023).
- Variable étudiée : Probabilité de décès (femmes)
- Liste des transformations préalables (moyennes mobiles) : Aucune
- Ages étudiés : Age 0, et tranches d'âge de 1 à 80 ans

Modélisation des mortalités pour la table prospective de l'Equateur

Itérations :

- Itération 1 : Détection **a posteriori** de la meilleure classe de modèle (Best Model Class) pour chaque âge modélisé
 - Début de l'historique étudié : 1950
 - Année actuelle supposée : 1969
 - Horizon de prévision : 1994 (25 ans)
- Itération 2 : Utilisation **a priori** de la meilleure classe de modèle (Best Model Class) pour chaque âge modélisé
 - Début de l'historique étudié : 1950
 - Année actuelle supposée : 1994
 - Horizon de prévision : 2019 (25 ans)

Modélisation des mortalités pour la table prospective de l'Equateur

Paramétrage et performances du meilleur modèle obtenu

Paramétrages d'exploration associés aux pays de référence A

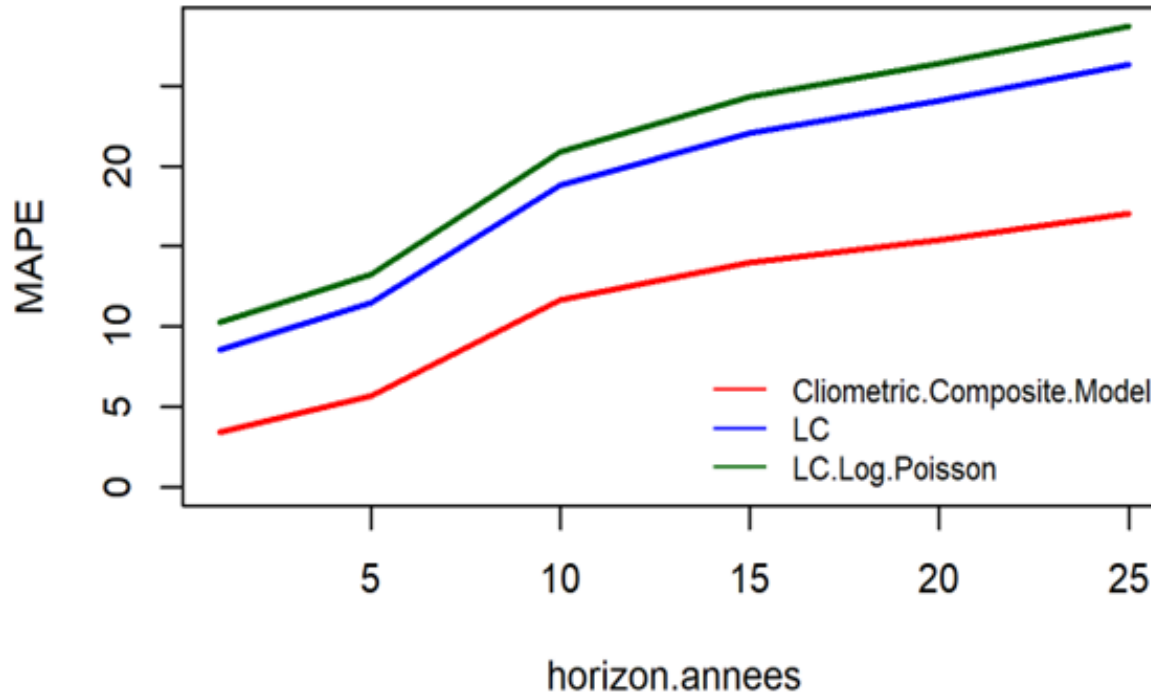
Lors de l'itération 1 de chaque âge modélisé, l'exploration consiste à tester chaque combinaison du produit cartésien des ensembles ci-dessous. Chaque ensemble désigne un paramètre et ses éléments sont les modalités du paramètre.

Pour chaque paramètre les modalités à tester sont :

- Liste des vecteurs de pays « A » : ("Danemark"), ("Spain"), ("France"), ("Italy"), ("Sweden")
- Variable étudiée : Probabilité de décès
- Ages étudiés : Age 0, et tranches d'âge de 1 à 80 ans
- Liste de vecteurs de sexes testés : ("M"), ("F"), ("MF")
- Liste de vecteurs d'âges d'appariement testés : ('0'), ('1-5'), ('5-10'), ('0', '1-5', '5-10')
- Liste des transformations préalables (moyennes mobiles) : ("Interpolation", "Interpolation + CMA5", "Interpolation + CMA7", "Interpolation + CMA9")

Modélisation des mortalités pour la table prospective de l'Equateur

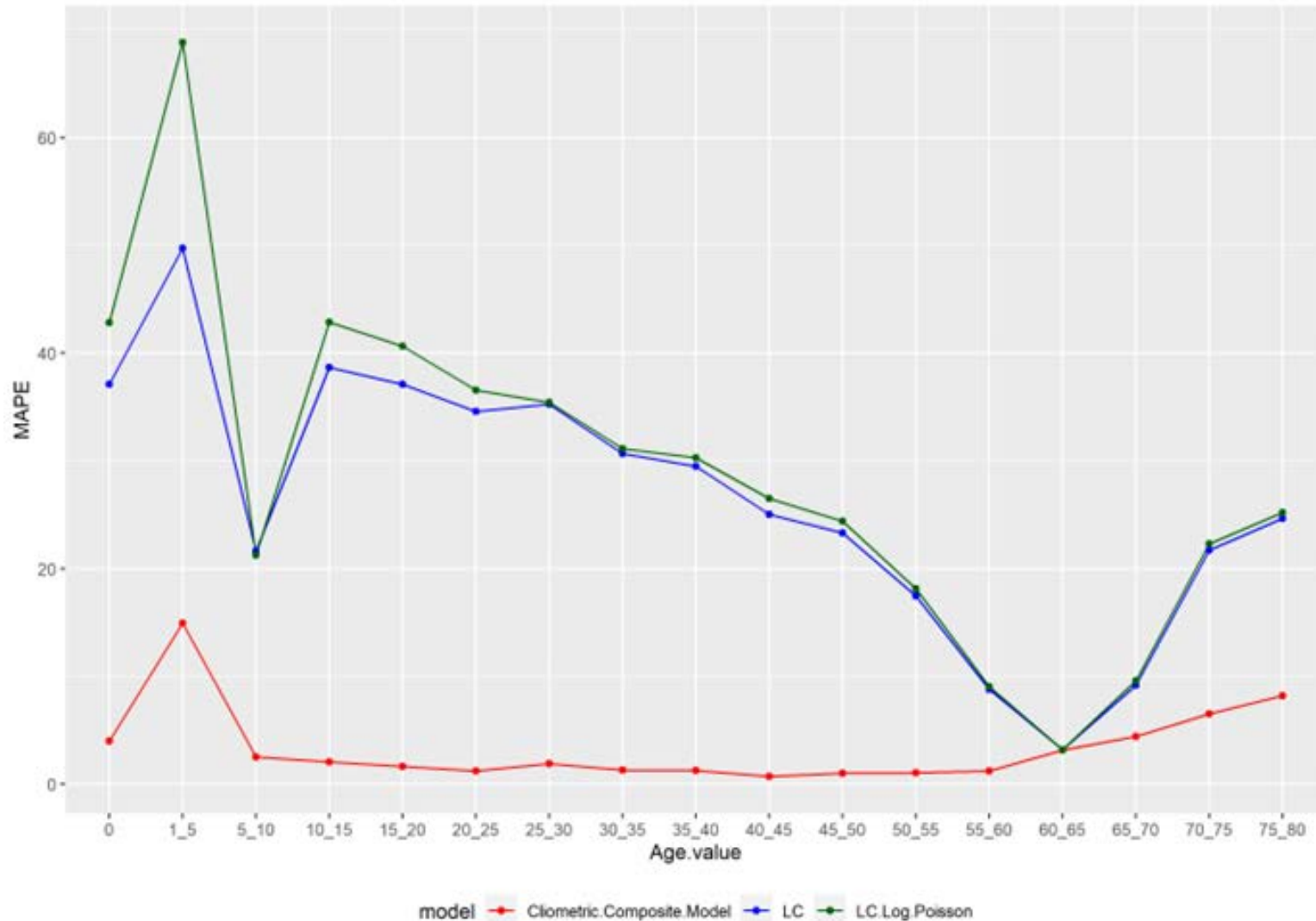
Performance des modèles selon l'horizon de prévision



MAPE (%)	Horizons					
Modèles	1	5	10	15	20	25
CBD.log.Poisson	66,1	61,1	55,7	50,8	46,9	44,0
Cliometric.Composite.Model	3,4	5,7	11,7	14,0	15,4	17,0
LC	8,6	11,5	18,8	22,1	24,1	26,3
LC.Log.Poisson	10,2	13,2	20,9	24,3	26,4	28,7

Modélisation des mortalités pour la table prospective de l'Equateur

MAPE tous horizons confondus, selon les différents âges modélisés (sans CBD)



Modélisation des mortalités pour la table prospective de l'Equateur

MAPE tous horizons confondus, selon les différents âges modélisés

MAPE (%)	Modèles			
Ages	CBD.log.Poisson	Cliometric.Composite.Model	LC	LC.Log.Poisson
0	98,4	4,0	37,1	42,8
1_5	70,6	14,9	49,7	68,8
5_10	19,1	2,5	21,6	21,2
10_15	32,2	2,1	38,7	42,9
15_20	22,0	1,7	37,1	40,7
20_25	27,7	1,2	34,6	36,6
25_30	47,1	1,9	35,3	35,4
30_35	65,5	1,3	30,7	31,1
35_40	73,1	1,3	29,5	30,3
40_45	69,5	0,7	25,0	26,5
45_50	62,4	1,0	23,3	24,4
50_55	50,6	1,0	17,5	18,1
55_60	39,2	1,2	8,8	9,0
60_65	27,8	3,1	3,2	3,2
65_70	18,9	4,4	9,2	9,6
70_75	13,2	6,5	21,7	22,3
75_80	11,3	8,2	24,7	25,2

Résultats des tests de sensibilité par rapport au paramétrage exploratoire

- ▶ Face à un modèle performant et riche en options de paramétrages, mais ne présentant aucun signe de surapprentissage, la réduction des choix de paramétrage entraîne un risque de réduction des performances.
- ▶ Il n'existe pas un dimensionnement paramétrique optimal « absolu » pour les modèles mixtes cliométriques composites.
- ▶ Le dimensionnement paramétrique optimal semble plus dépendre de la structure interne des données traitées que de leur volume : en effet nous constatons que l'Equateur bien que disposant d'un historique plus court (remontant à 1950) a pu supporter un dimensionnement paramétrique plus élevé que l'Inde qui dispose d'un historique plus long (depuis 1930).
- ▶ Le seuil de surapprentissage est variable selon le pays et ne peut être déterminé qu'après exploration d'un certain nombre de scénarios de dimensionnement paramétrique

Perspectives et travaux futurs

- ▶ **Des travaux empiriques supplémentaires seront nécessaires pour tester et améliorer les autres modèles théoriques que nous avons développés, en les appliquant dans les contextes adaptés à leur usage. Cela concerne notamment les modèles externes cliométriques composites destinés spécifiquement aux pays les moins avancés**
- ▶ **Le calcul des intervalles de confiance est un autre aspect des modèles développés qui nécessite des recherches et des travaux empiriques complémentaires, notamment via une approche Monte-Carlo**
- ▶ **Poursuite des travaux auprès des assureurs, organismes gouvernementaux, autorités de régulation et de contrôle dans la zone CIMA (Conférence interafricaine des marchés d'assurance)**