



STATISTIQUE BAYÉSIENNE

Dating and forecasting turning point by Bayesian clustering with dynamic structure

Romain Lesauvage et Alain Quartier-la-Tente

Table des matières

1	Introduction	2
2	Cadre théorique	2
2.1	Spécification du modèle	2
2.2	Estimation bayésienne	3
3	Application aux données françaises	4
4	Discussion	7
4.1	Prévision	7
4.2	Comparaison avec l'indicateur de retournement de l'Insee	7
5	Conclusion	9
A	Annexe	11

L'ensemble du projet est disponible à l'adresse <https://github.com/ARKEnsae/BayesianTurningPoint>.

1 Introduction

Dans son ouvrage *Les vagues longues de la conjoncture*, Nikolai Kondratiev mettait en évidence l'existence de cycles économiques formés de deux périodes, une phase ascendante puis une phase descendante. Bien que contestée et complétée par la suite par d'autres analyses, cette découverte des cycles économiques a très vite mené les chercheurs à essayer de savoir dans quelle phase l'économie se trouvait et, de fait, savoir déterminer et prévoir le moment où le cycle s'inverse est un enjeu majeur. C'est sur ce sujet que nous allons travailler ici, à partir de l'article intitulé "Dating and forecasting turning points by Bayesian clustering with dynamic structure: a suggestion with an application to Austrian data. Journal of Applied Econometrics" Kaufmann (2010).

Ce papier poursuit un premier travail introduit par Frühwirth-Schnatter et Kaufmann (2008). L'idée est de travailler sur un ensemble de plusieurs séries temporelles que l'on va essayer de regrouper en fonction de leur dynamique. Nous travaillerons alors sur deux périodes, selon que l'on est inférieur ou supérieur à la moyenne de croissance et on cherchera à déterminer le point de retournement dans le cycle.

Dans ce travail, nous commençons par détailler les spécifications théoriques du modèle et l'estimation bayésienne associée. Nous appliquons cette méthode à la manière de ce qui a été fait par Kaufmann (2010) mais sur des données françaises.

2 Cadre théorique

2.1 Spécification du modèle

Dans la suite, nous travaillerons avec des séries centrées, réduites et corrigées des variations saisonnières. Notons y_{it} la série correspondant à un taux de croissance, avec $i \in \{1, \dots, N\}$ représentant les différentes séries à notre disposition et $t \in \{1, \dots, T\}$ représentant les différentes périodes d'études. Nous supposons que nos séries suivent des processus autorégressifs, c'est-à-dire que l'on a

$$y_{it} = \mu_{I_{it}}^i + \phi_1^i y_{i,t-1} + \dots + \phi_p^i y_{i,t-p} + \epsilon_{it}$$

avec $\epsilon_{it} \sim \mathcal{N}(0, \frac{\sigma}{\lambda_i^2})$. L'indice I_{it} précisera à quel moment du cycle nous nous trouverons, ainsi on aura $I_{it} \in \{1, 2\}$.

Théoriquement, les séries étudiées sont supposées être indépendantes les unes des autres. Cependant, en pratique, il existe des séries qui suivent des évolutions proches et il peut donc être utile d'exploiter cette information pour notre estimation. Nous allons donc chercher à créer des groupes de séries de manière intelligente. Pour cela, nous introduisons la variable $S_i \in \{1, \dots, K\}$ qui indique dans quel groupe est classée la variable i . Dans la suite, nous supposons que toutes les séries d'un groupe vont avoir des paramètres distribués selon une même loi mais différente selon la position dans le cycle :

$$\mu_{I_{S_i,t}}^i \sim \begin{cases} \mathcal{N}(\mu_1^k, q_1^k) & \text{si } S_i = k \text{ et } I_{kt} = 1 \\ \mathcal{N}(\mu_2^k, q_2^k) & \text{si } S_i = k \text{ et } I_{kt} = 2 \end{cases}$$

$$(\phi_1^i, \dots, \phi_p^i) \sim \mathcal{N}(\phi^k, Q_\phi^k) \text{ si } S_i = k$$

Pour la suite, nous considérons que la période 1 ($I_{kt} = 1$) comme celle où l'on se trouve la croissance est supérieure à la moyenne (en haut du cycle) et la période 2 comme celle où la croissance est inférieure à la moyenne (en bas de cycle). Nous ferons également l'hypothèse que I_{kt} suit un processus de Markov, c'est-à-dire $\mathbb{P}(I_{k,t} = j | I_{k,t-1} = j) = \xi_{j,l}^k$ avec $j, l \in \{1, 2\}$. Par propriété sur les chaînes de Markov, on sait donc que $\forall j \in \{1, 2\}, \sum_{l=1}^2 \xi_{j,l}^k = 1$. Cette spécification permet de repérer en temps réel les points où le cycle change de période.

Nous supposons que la variable S_i suit un modèle logit multinomial, c'est-à-dire $\mathbb{P}(S_i = k | \gamma_1, \dots, \gamma_K, \gamma_{z,1}, \dots, \gamma_{z,K-1}) = \frac{\exp(\gamma_k + Z_k \gamma_{z,k})}{1 + \sum_{l=1}^{K-1} \exp(\gamma_l + Z_l \gamma_{z,l})}$ où l'on prendra le groupe K comme référence ($\gamma_K = \gamma_{z,K} = 0$), Z_i est un vecteur

permettant de repérer l'appartenance à un groupe et les paramètres γ sont inconnus mais spécifiques aux groupes.

Enfin, il reste à ajouter une spécification pour en compte la dynamique du cycle. Pour cela, nous supposons avoir deux groupes spécifiques pour la suite de l'analyse et $K - 2$ groupes indépendants. Les séries telles que $S_i = 2$ seront les séries qui mènent le cycle (indicateurs avancés), tandis que les séries pour lesquelles $S_i = 1$ seront les séries qui coïncident avec le cycle (tels que le PIB). Nous définissons donc une nouvelle variable I^* de la façon suivante:

$$\begin{cases} I^* = 1 := (I_{1t} = 1, I_{2t} = 1) \\ I^* = 2 := (I_{1t} = 1, I_{2t} = 2) \\ I^* = 3 := (I_{1t} = 2, I_{2t} = 1) \\ I^* = 4 := (I_{1t} = 2, I_{2t} = 2) \end{cases}$$

En cas d'incertitude sur la structure dynamique des nos séries, c'est-à-dire sur l'identification des groupes 1 et 2, on peut introduire un paramètre ρ^* , caractérisant la structure entre les groupes. Il sera donc la réalisation d'une des $K(K - 1)$ permutations de $\{1, 2, 0_{K-2}\}$. On considérera que $\rho^* = 1$ pour le groupe des séries coïncidentes, $\rho^* = 2$ pour celles qui mènent le cycle, et toute autre valeur renverra aux séries aux comportements autonomes. Le groupe d'appartenance de chaque série peut donc soit être basée sur un a priori (ce qui est fait dans notre cas) ou être estimée par une méthode statistique.

2.2 Estimation bayésienne

2.2.1 Notations

Dans la suite, notons $\forall i \in \{1, \dots, N\}$, $y_i^t = \{y_{i,t}, y_{i,t-1}, \dots, y_{i,1}\}$, $Y_t = \{y_{1,t}, \dots, y_{N,t}\}$ et $Y^T = \{Y_t, \dots, Y_1\}$. De même, notons $S^N = \{S_1, \dots, S_N\}$ l'ensemble des indicatrices d'appartenances aux groupes, $I^T = \{I_1^T, \dots, I_K^T\}$ où $I_k^T = \{I_{k,T}, \dots, I_{k,1}\}$ les indicatrices d'états. Enfin, nous notons $\lambda^N = \{\lambda_1, \dots, \lambda_N\}$ qui regroupe les poids qui seront utilisés.

Par commodité, nous noterons également $\theta = (\mu_1^1, \mu_2^1, \dots, \mu_1^K, \mu_2^K, \phi^1, \dots, \phi^K, Q^1, \dots, Q^K, \sigma^2, \xi^*, \xi^{\rho^*(k)=0}, \gamma, \gamma_z)$ où $Q^k = \begin{pmatrix} q_1^k & 0 & 0 \\ 0 & q_2^k & 0 \\ 0 & 0 & Q_\phi^k \end{pmatrix}$ représente l'hétérogénéité intra-groupe, $\xi^{\rho^*(k)=0} = \{\xi_{11}^k, \xi_{12}^k, \xi_{21}^k, \xi_{22}^k\}$, $\gamma = (\gamma_1, \dots, \gamma_{K-1})$ et $\gamma_z = (\gamma_{z,1}, \dots, \gamma_{z,K-1})$.

Nous pouvons passer à l'estimation par MCMC pour obtenir une posteriori sur le paramètre $\psi = (\theta, S^N, I^T, \lambda^N, \rho^*)$.

2.2.2 Estimation par MCMC

Nous partons de $\pi(\psi|Y^T) \propto L(Y^T|\psi)\pi(\psi)$ où $L(Y^T|\psi)$ est la vraisemblance, que l'on peut réécrire. En effet, $L(Y^T|\psi) = \prod_{t=p+1}^T \prod_{i=1}^N f(y_{i,t}|y_i^{t-1}, \mu_{I_{S_i,t}}^{S_i}, \phi^{S_i}, Q^{S_i}, \lambda_i, \sigma^2)$, avec

$$f(y_{i,t}|y_i^{t-1}, \mu_{I_{S_i,t}}^{S_i}, \phi^{S_i}, Q^{S_i}, \lambda_i, \sigma^2) = \frac{1}{\sqrt{2\pi\nu_{it}^{S_i}}} \exp\left(-\frac{1}{2\nu_{it}^{S_i}}\left(y_{it} - \mu_{I_{S_i,t}}^{S_i} - \sum_{j=1}^p \phi_j^{S_i} y_{i,t-1}\right)^2\right)$$

Avec $y_{it} = X_{it}^{S_i} \beta^{S_i} + \epsilon_{it}^*$, où $\epsilon_{it}^* \sim N(0, \nu_{it}^{S_i})$, $X_{it}^{S_i} = (D_{1t}^{I(S_i)}, D_{2t}^{I(S_i)}, y_{i,t-1}, \dots, y_{i,t-p})$, $D_{jt}^{I(S_i)} = \mathbb{1}(I_{S_i,t} = j)$, $\beta^{S_i} = (\mu_1^{S_i}, \mu_2^{S_i}, \phi^{S_i})$, et $\nu_{it}^{S_i} = X_{it}^{S_i} Q^{S_i} X_{it}^{S_i'} + \frac{\sigma^2}{\lambda_i}$.

Nous supposons dans la suite que la structure ρ^* est connue (déterminée par un a priori), nous pouvons alors réécrire la distribution a priori sur ϕ ainsi :

$$\pi(\phi) = \pi(I^{*T}|\rho^*, \xi^*) \prod_{\rho^*(k)=0} \pi(I_k^T|\rho^*, \xi^*) \pi(S^N|\gamma, \gamma_z, Z^N) \pi(\lambda^N) \pi(\theta)$$

où les densités $\pi(I^{*T}|\rho^*, \xi^*)$, $\pi(I_k^T|\rho^*, \xi^*)$ et $\pi(S^N|\gamma, \gamma_z, Z^N)$ sont connues. Nous supposons que les poids (λ^N) sont indépendants et suivent une loi de Gamma. Il reste à spécifier l'a priori sur $\pi(\theta)$: pour cela, nous décomposons θ en blocs sur lesquels nous faisons des a priori standards.

Nous cherchons ensuite à générer à partir de la loi a posteriori $\pi(\phi|Y^T)$, en se basant sur le procédé détaillé dans [Frühwirth-Schnatter et Kaufmann \(2008\)](#). Nous pouvons raisonner en 4 étapes différentes :

1. $\pi(S^N|Y^T, I^T, \rho^*, \lambda^N, \theta)$
2. $\pi(I^T|Y^T, S^N, \rho^*, \lambda^N, \theta)$
3. $\pi(\lambda^N|Y^T, I^T, S^N, \theta)$
4. $\pi(\theta|Y^T, S^N, I^T)$

Dans la première étape, nous générerons les indicatrices de groupes pour chaque série puis, à partir de ρ^* , nous obtenons une réalisation pour l'indicatrice d'état de l'étape 2. Dans l'étape 4, toutes les distributions a posteriori sont conjuguées aux a priori, sauf pour γ et γ_z , nous utiliserons alors l'agorithme de Metropolis-Hastings pour générer selon leur loi.

3 Application aux données françaises

L'article [Kaufmann \(2010\)](#) a utilisé la spécification théorique présentée dans la partie précédente sur des données autrichiennes. Nous avons décidé dans le cadre de ce travail d'appliquer cela sur des données françaises. Nous avons donc récupéré un certain nombre de données liées au PIB ou aux enquêtes de conjoncture¹, dont le détail est expliqué dans le tableau 1.

Nous avons retenu ici $K = 2$ groupes pour les séries : celles qui coïncident avec le cycle ($S_i = 1$) et celles qui le mènent ($S_i = 2$). Pour l'analyse, nous nous limitons aux hypothèses suivantes :

1. La structure (ρ^*) est connue en avance : nous avons réparti chaque série dans un des groupes.
2. La classification est totale : on suppose qu'il n'y a pas de groupes de séries indépendantes.
3. L'a priori sur S_i est de type logit.

Nous avons adapté les programmes de [Kaufmann \(2010\)](#) à nos données et nous analysons dans ce rapport les résultats obtenus.

Le modèle étudié ici se caractérise entre autres par la prise en compte de la spécification de l'hétérogénéité entre séries. Ceci est illustré par le graphique 1. La figure 1a représente la moyenne de la somme des paramètres autorégressifs. Comme dans l'article de [Kaufmann \(2010\)](#), sur les données françaises, nous pouvons noter l'importance de la prise en compte de l'hétérogénéité. La figure 1b représente la moyenne des variances $\frac{\sigma^2}{\lambda_i}$. Cette fois-ci, on note une différence avec ce que donnaient les données autrichiennes. En effet, notre distribution est plus étirée et plus aplatie que celle de l'article. Cela montre que la dispersion de l'erreur de variance est plus faible, ce qui s'explique sans doute par le processus de standardisation des séries préalable et par le fait que nous travaillons sur moins de séries. Le graphique 2 représente la moyenne non-conditionnée de chaque groupe de série dans les deux états, selon qu'elles sont au-dessous ou au-dessus de la croissance moyenne. Cela permet d'illustrer que, contrairement à l'article, nos deux groupes de séries n'ont pas exactement le même comportement puisque les séries sont décalées les unes par rapport aux autres.

Le graphique 3 permet de juger la significativité de l'information *a priori* utilisée pour la classification. Nous obtenons des conclusions légèrement différentes de celles de l'article : pour le groupe 1, la distribution *a posteriori* des effets du PIB est éloignée de zéro, tandis que pour le groupe 2, on n'observe cela que pour les carnets de commandes (variable *manuf-osc*). Cela s'explique par la façon dont la classification *a priori* a été effectuée, selon que la série allait être en avance ou non sur le PIB.

1. Nous avons décidé de ne retenir que les soldes d'opinion utilisés pour construire l'indicateur de retournement France publié par l'Insee (voir section 4.2) en supposant que, pour sa construction, les soldes les plus pertinents pour détecter les points de retournement avaient été retenus par l'Insee. Nous avons également retenu deux indicateurs synthétiques publiés par l'Insee : le climat des affaires France, qui retrace le cycle conjoncturel, et le climat de l'emploi France, qui retrace le cycle conjoncturel spécifique de l'emploi salarié.

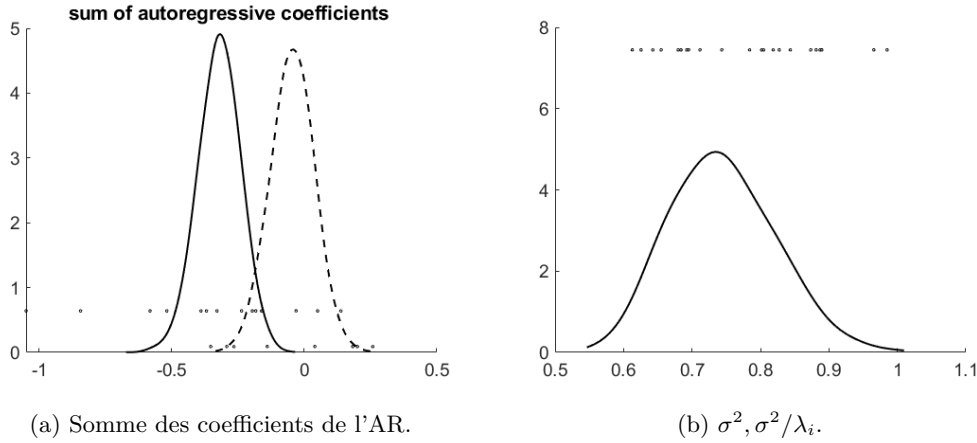


FIGURE 1 – Hétérogénéité des séries - groupe 1 (trait plein) et 2 (pointillés).

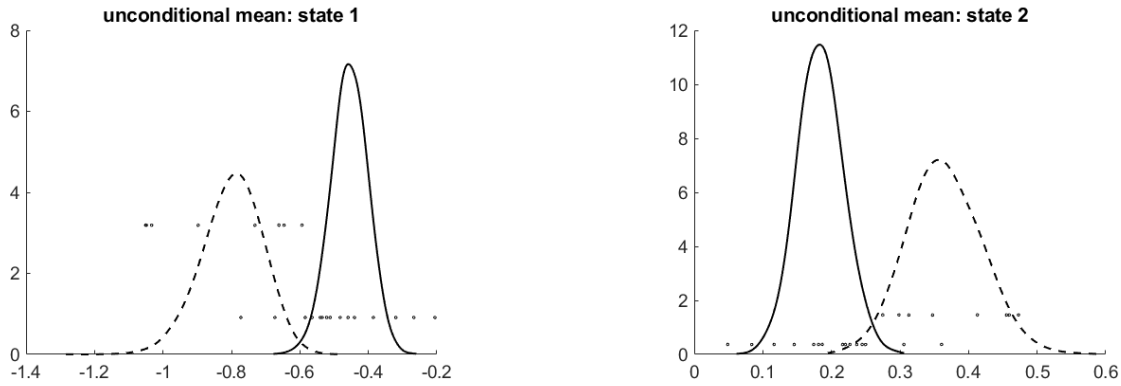


FIGURE 2 – Moyenne non-conditionnée dans les différents états.

Le graphique 4 permet de visualiser l'effet de la corrélation avec le PIB et le carnet de commandes entre les groupes 1 (figure 4a) et 2 (figure 4b). La courbe représente $P(S = j|Z, Y^T) = \int P(S = j|Z, \gamma, \gamma_z)\pi(\gamma, \gamma_z|Y^T)d\gamma d\gamma_z$, la probabilité conditionnelle sur le PIB et le carnet de commandes. Cette probabilité est la plus élevée pour les séries du groupe 1 qui sont fortement corrélées avec le PIB et corrélées positivement avec le carnet de commandes, la corrélation avec le PIB semblant jouer un rôle plus important puisque la probabilité varie plus selon la valeur de cette dernière. Nous retrouvons ici la définition du premier groupe de variables qui contient celles qui sont coincidentes avec le cycle et donc corrélées avec le PIB. Pour le groupe 2, la valeur de la probabilité est maximale pour les séries fortement corrélées avec le carnet de commandes et cette variable semble être plus importante (la corrélation avec le PIB n'ayant pas une grande influence). Cela conforte les analyses tirées du graphique 3.

La hauteur des lignes verticales du graphique 4 représente la probabilité $P(S_i = j|Z_i, Y^T)$ *a posteriori* pour une série d'appartenir aux groupes 1 et 2. Dans la grande majorité des cas, cette probabilité a été mise à jour à 0 ou à 1, grâce à l'information supplémentaire obtenue. Cela permet de vérifier la qualité de notre pré-classification : elle semble cohérente car les séries classées dans le groupe coïncident ont une forte probabilité *a posteriori* d'appartenir à ce groupe, et de même pour les séries classées dans le groupe avancé.

Le graphique 5 est central puisqu'il permet de repérer les points de retournement de notre cycle. Il représente les probabilités $P(I_{k,t} = 1|Y^T)$ pour les séries des deux groupes, obtenues en moyennant les résultats sur M simulations. Les résultats semblent significatifs puisque les probabilités ont des valeurs très proches de 0 ou de 1 dans une majorité de cas. À partir de ce graphique, on définit la date t comme un point de retournement si $P(I_{k,t-1} = 1, I_{k,t} = 1|Y^T) < 0,5$ et $P(I_{k,t+1} = 1, I_{k,t+2} = 1|Y^T) > 0,5$ (pour un pic), k désignant le groupe coïncident, c'est-à-dire une faible probabilité de ne pas changer d'état à ce moment et, une fois qu'on a passé ce moment, une forte probabilité de rester dans le même état. Inversement, on considère qu'il y a un creux lorsque $P(I_{k,t-1} = 1, I_{k,t} = 1|Y^T) > 0,5$ et $P(I_{k,t+1} = 1, I_{k,t+2} = 1|Y^T) < 0,5$. On peut donc repérer ici différents points de retournement qui correspondent à des réels événements sur les marchés financiers en 1996 et 1999 (crise financière

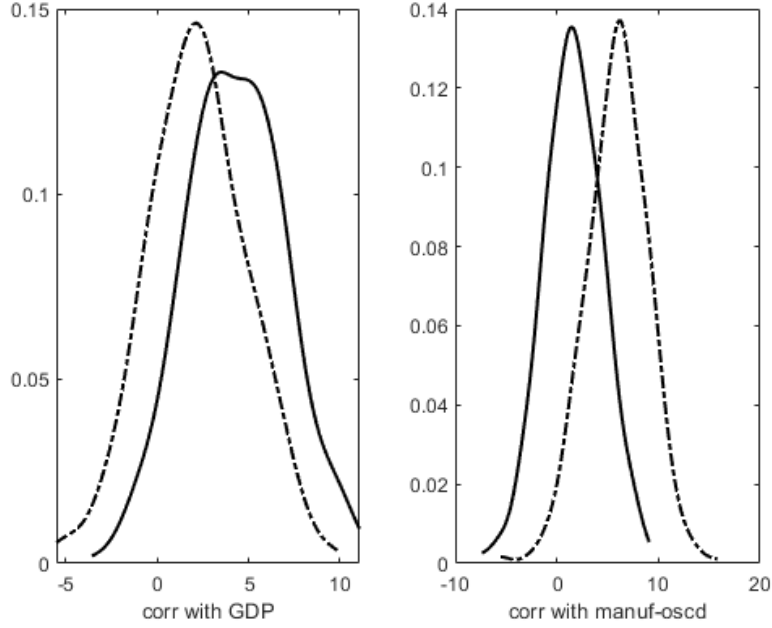


FIGURE 3 – $\pi(\gamma, \gamma_z | Y^T)$ - groupe 1 (trait plein) et 2 (pointillé).
Note : GDP = PIB et manuf-oscd = solde d'opinion sur le niveau des carnets de commandes.

asiatique), 2002 et 2003 (krach boursier), 2009 (crise économique mondiale) ou encore 2013 (fin de la crise de 2008 avec encore des répercussions en France).

Afin de juger de la robustesse de notre modèle, nous avons regardé ce qui changeait lorsque l'analyse commençait soit en 1992, soit en 2004², et quand elle se terminait soit en 2019, soit en 2020 (pour prendre en compte ou non les données de la crise du Covid-19). Les résultats sont résumés dans le graphique 6. On remarque que la date de début de l'analyse semble avoir un rôle dans la détection de certains points de retournement. En effet, quand l'analyse commence en 2004, le point de retournement autour de 2013 n'est plus détecté, peut-être est-ce dû à l'influence de la crise de 2009 sur les estimations des modèles autorégressifs. En revanche, que l'on termine l'analyse en 2019 ou en 2020, on note assez peu de différences : la crise du Covid-19, bien qu'ayant une influence extrême sur les séries, ne change pas le diagnostic sur les estimations passées des points de retournement.

2. Car certains soldes d'opinion dans l'enquêtes de conjoncture dans les services étaient trimestriels avant 2004 et mensuels après 2004 : la dynamique de ces variables ont donc pu changer à cette date.

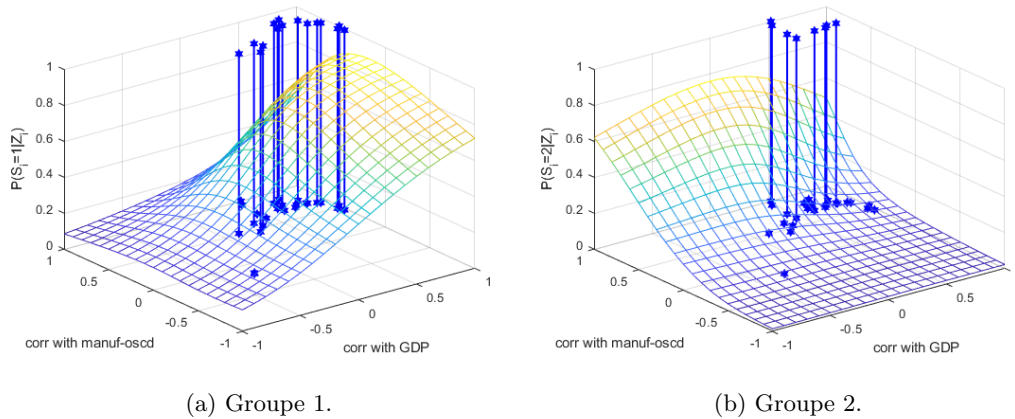


FIGURE 4 – Probabilités *a priori* et *a posteriori* $P(S = j | Z, Y^T)$ et $P(S_i = j | Z_i, Y^T)$.
Note : GDP = PIB et manuf-oscd = solde d'opinion sur le niveau des carnets de commandes.

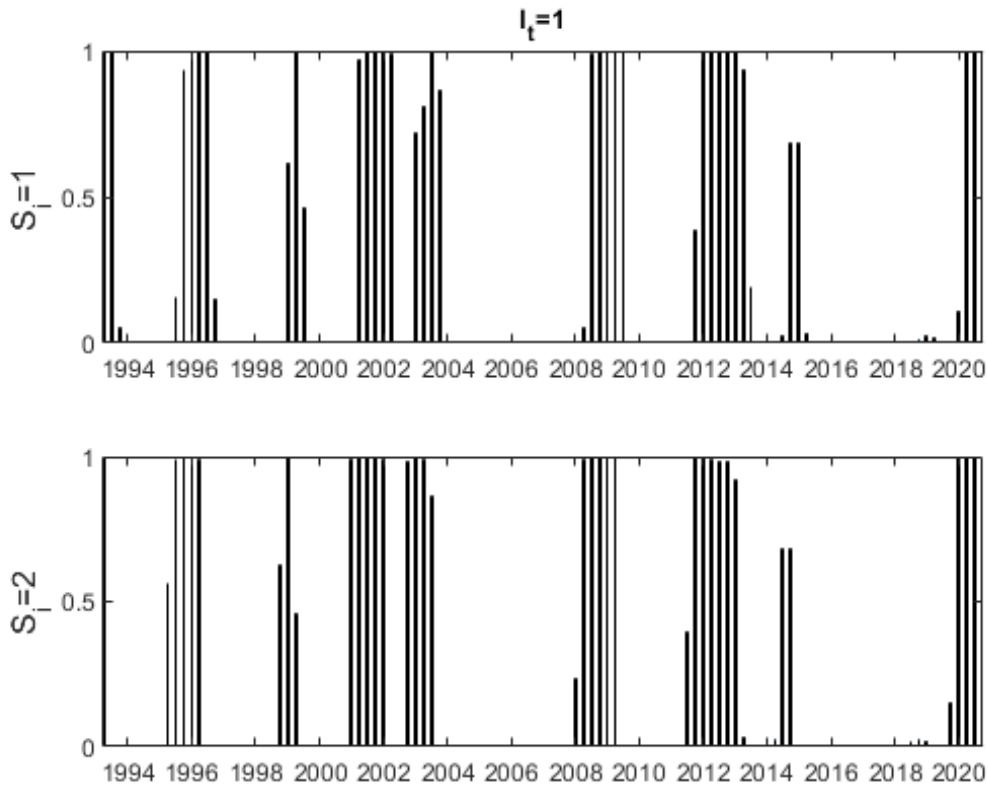


FIGURE 5 – Probabilités *a posteriori* $P(I_{kt} = 1|Y^T)$ des séries coïncidentes ($S_i = 1$) et qui mènent le cycle ($S_i = 2$) sur la période 1992-2020 avec une classification a priori.

Note : la hauteur des lignes verticales représente la probabilité $P(S_i = j|Z_i, Y^T)$ a posteriori d'appartenir aux groupes 1 et 2.

Une fois des données disponibles entièrement pour la période entière du Covid-19, il sera intéressant de voir ce qui a été capté par notre modèle. De manière générale, une réflexion pourrait être menée sur la gestion des points des points atypiques.

4 Discussion

4.1 Prévision

Nous l'avons vu, notre modèle permet de repérer les points de retournement de l'économie. Cependant, s'il est toujours intéressant de pouvoir le remarquer après coup, une utilisation plus immédiate serait appréciée. Le modèle permet cela puisqu'il peut être utilisé dans une optique de prévision. L'idée est d'utiliser la densité prédictive *a posteriori* $\pi(I_{T+h}^*|Y^T) = \pi(I_{T+h}^*|I^*)$ et de simuler récursivement les valeurs futures à partir de cela. En moyennisant nos résultats sur un grand nombre de simulations, on peut alors déterminer la probabilité d'être dans les différents états.

Du fait du caractère non conjoncturel de la crise actuelle, nous n'avons pas effectué d'exercice de prévision qui ne pourrait détecter de manière efficace un point de retournement.

4.2 Comparaison avec l'indicateur de retournement de l'Insee

Afin de pouvoir juger de l'efficacité de notre modèle, il est intéressant de le comparer à des méthodes déjà existantes et utilisées par les instituts nationaux de statistiques. En France, l'Insee publie un indicateur de retournement de

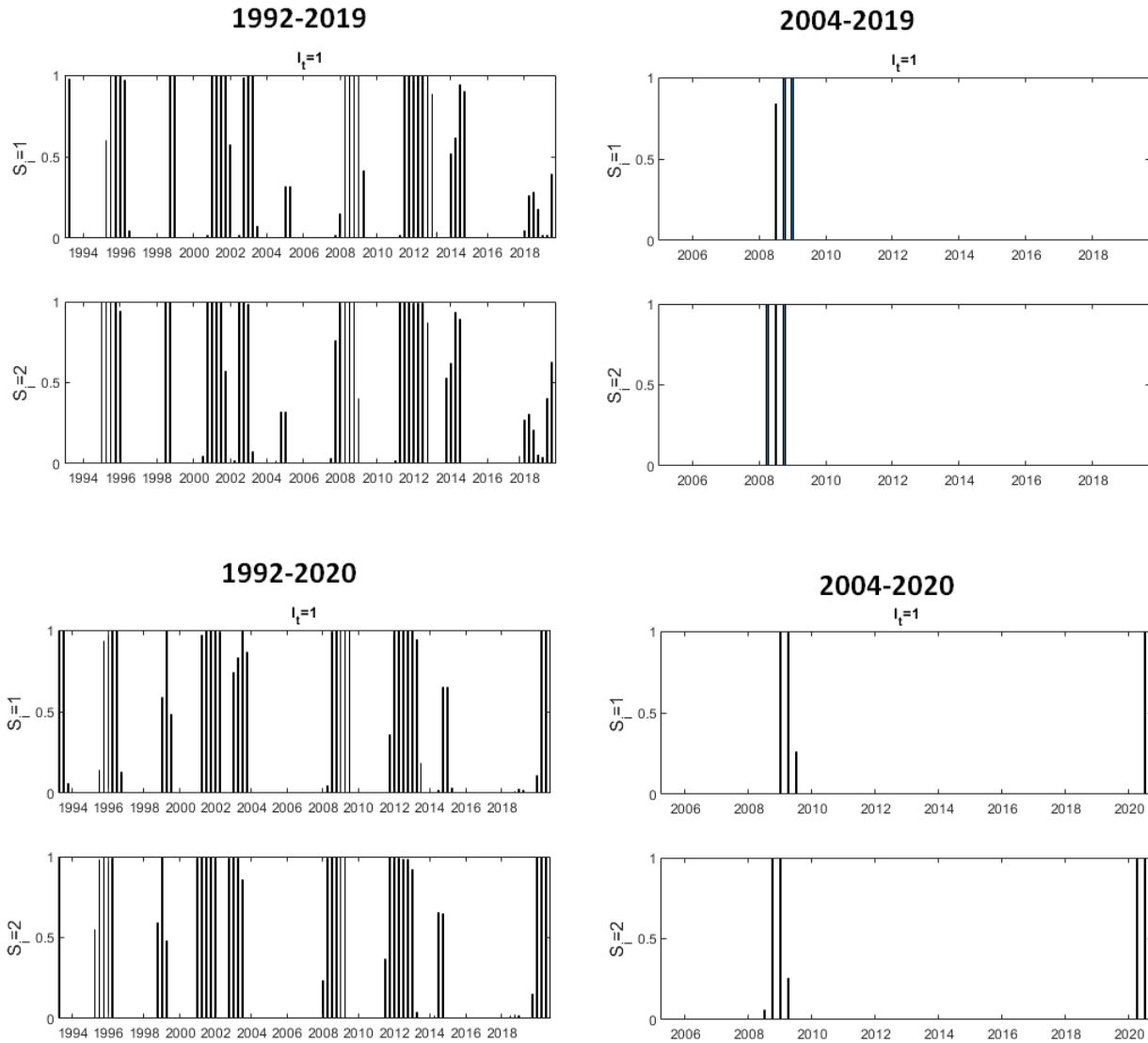


FIGURE 6 – Probabilités *a posteriori* $P(I_{kt} = 1|Y^T)$ des séries coïncidentes ($S_i = 1$) et qui mènent le cycle ($S_i = 2$) sur différentes périodes avec une classification a priori.

conjuncture estimé à partir d'un modèle markovien à variables cachées sur des soldes d'opinion recodés en deux modalités (voir [Bortoli et al \(2015\)](#)). Cet indicateur est mensuel, contrairement à notre modèle bayésien qui se base sur des séries trimestrialisées.

La valeur de cet indicateur correspond à la différence entre la probabilité que la phase conjoncturelle soit favorable et la probabilité qu'elle soit défavorable. Il évolue donc entre $+1$ et -1 : un point très proche de $+1$ (respectivement de -1) signale que l'activité économique est en période de nette accélération (respectivement de nette décélération). Les moments où l'indicateur est proche de 0 sont assimilés à des phases de stabilisation, c'est-à-dire de retour du rythme de croissance de l'activité vers sa moyenne de long terme.

Pour juger de la pertinence de la détection des points de retournement, il est nécessaire de disposer d'une datation des cycles d'expansion ou de contraction de l'activité. Pour avoir une datation de référence, nous appliquons la méthode décrite par [Bardaji et al \(2009\)](#) : à partir du PIB trimestriel en volume, nous extrayons la composante cyclique de la série autour de sa tendance grâce à un filtre de Christiano-Fitzgerald en retenant des cycles d'une durée de 1,5 à 10 ans et une règle de censure pour éliminer les phases de moins de 4 trimestres. Cela permet de déterminer les phases de ralentissement et d'accélération du PIB par rapport à sa tendance. Les comptes nationaux étant encore provisoires à partir de 2018, la datation l'est également à partir de cette date.

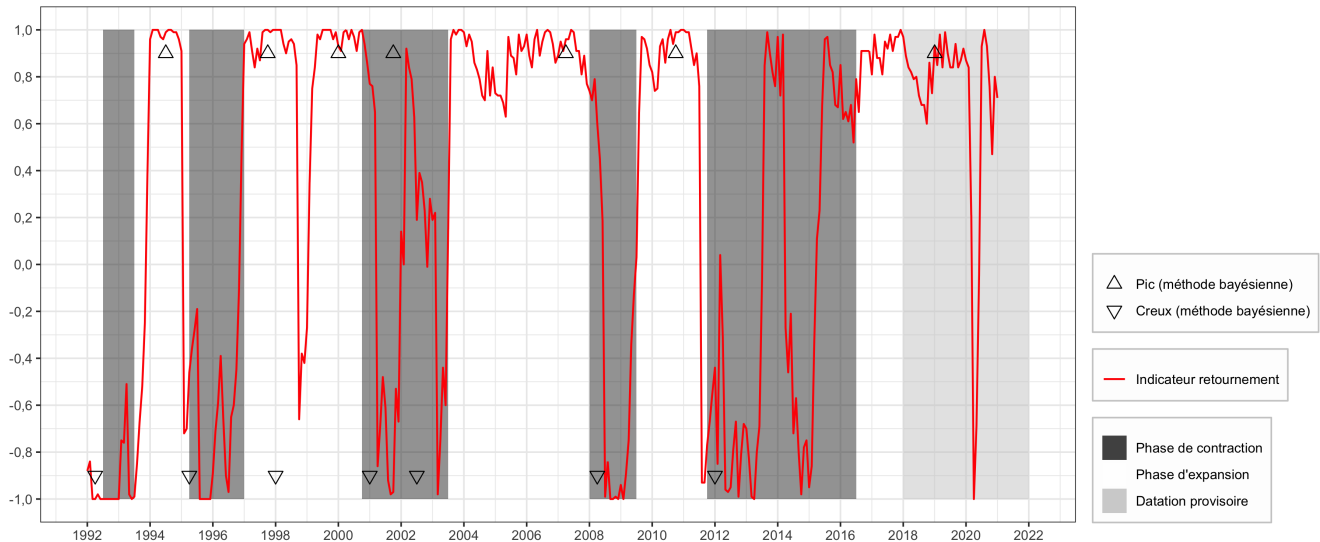


FIGURE 7 – L'indicateur de retournement France, la datation des cycles de croissance et points de retournement déterminés par approche bayésienne.

Lecture : l'indicateur de retournement est en zone favorable quand il est proche de 1, défavorable quand il est proche de -1. Avec la méthode bayésienne présentée dans cette article, les points de retournement détectés sont les suivants :

- Creux : 1992T2, 1995T2, 1998T1, 2001T1, 2002T3, 2008T2, 2012T1 ;
- Pics : 1994T3, 1997T4, 2000T1, 2001T4, 2007T2, 2010T4, 2019T1.

Ces résultats sont confrontés à une datation de référence des cycles de croissance : les zones noires correspondent à un ralentissement du PIB (par rapport à sa tendance) et les zones blanches à une accélération. La datation est provisoire à partir de 2018 (zone grise).

Le graphique 7 compare l'indicateur de retournement France, les points de retournement détectés par notre modèle et la datation de référence. Les points de retournement détectés sont généralement avancés par rapport à ceux de la datation officielle. Par exemple un creux est détecté en 1995T2 alors que la datation officielle détecte ce point de retournement en 1997 (passage d'une phase de contraction à une phase d'accélération). L'indicateur de retournement semble plus proche de la datation officielle mais donne en revanche un signal très bruité à certaines dates.

Le décalage dans la détection des points de retournement pourrait également s'expliquer par le faible nombre de séries utilisées dans notre application par rapport à l'article de Kaufmann (2010) (environ 25 dans notre application contre une centaine). Plus d'investigations pourraient également être faites pour comprendre les contributions des différentes variables à la détection des points de retournement.

5 Conclusion

Dans Kaufmann (2010), l'auteure propose une méthode de détection des points de retournement qu'elle applique à des données autrichiennes. Le modèle est basé sur l'idée de regrouper les séries temporelles de même type en ayant l'idée qu'elles auront alors la même dynamique. Pour prendre en compte que certaines séries sont en avance sur le cycle économique, deux groupes sont principalement étudiés avec les séries en avance sur les cycles et celles qui lui sont coincidentes, comme les PIB et ses composantes. Les distributions *a posteriori* permettent ensuite de repérer les points de retournement.

En appliquant cette méthode aux données françaises, nous avons ainsi pu repérer plusieurs points de retournements. Afin de juger de la qualité de nos résultats, nous les avons comparé avec l'indicateur de retournement produit par l'Insee, ce qui a mis en lumière un léger décalage de notre modèle, mais qui peut par exemple s'expliquer par le nombre de points utilisés pour calibrer le modèle. Les résultats sont tout de même apparus assez proches.

En conclusion, l'analyse des cycles économiques revêt depuis plusieurs années un enjeu majeur. Dans le contexte de crise sanitaire tel que nous le connaissons, il est important d'améliorer notre connaissance de l'activité économique afin de pouvoir juger de la relance post-Covid le moment venu. Par ailleurs, pouvoir anticiper la détection de point de retournement ou d'inflexion de l'économie pourrait être un outil essentiel pour aider les politiques à calibrer au mieux les aides apportées aux entreprises.

Dans le but d'effectuer des estimations en temps réel, le points faible de ce modèle est qu'il ne prend pas directement en compte les valeurs manquantes et donc les *ragged-data*. En effet, dans ce modèle toutes les séries doivent avoir la même fréquence et la trimestrialisation implique une perte d'informations. Par ailleurs, en pratique les différentes séries macroéconomiques ne sont pas publiées en même temps (c'est ce qu'on appelle *ragged-data*). Par exemple en France, les résultats des enquêtes de conjoncture sont disponibles dès la fin du mois alors que les comptes trimestriels ne le sont que 30 jours après la fin du trimestre. Pour prendre en compte ces deux points, le modèle autorégressif utilisé pourrait être amélioré utilisant par exemple un modèle espace-état.

A Annexe

Variable	Commentaire	S_i	Trimestrialisation
YER	PIB	1	G_T
MTR	Imports	1	G_T
PCR	Consommation	1	G_T
ITR	Investissements	1	G_T
XTR	Exports	1	G_T
manuf-oscd	Enquête de conjoncture industrie - carnet de commandes généraux	2	D_T
manuf-oscd	Enquête de conjoncture industrie - carnet de commandes étrangers	2	D_T
manuf-ossk	Enquête de conjoncture industrie - niveau des stocks de produits finis	1	D_T
manuf-tpa	Enquête de conjoncture industrie - tendance passée de production	1	D_T
manuf-tppe	Enquête de conjoncture industrie - tendance prévue de production	2	D_T
manuf-pgp	Enquête de conjoncture industrie - perspectives générales	2	D_T
ser-cap	Enquête de conjoncture services - évolution passée du chiffre d'affaires	1	D_T
ser-capre	Enquête de conjoncture services - évolution prévue du chiffre d'affaires	2	D_T
ser-dem	Enquête de conjoncture services - demande prévue	2	D_T
bat-jcc	Enquête de conjoncture bâtiment - niveau des carnets de commandes	2	D_T
bat-epa	Enquête de conjoncture bâtiment - évolutions passée des effectifs	1	D_T
bat-tuc	Enquête de conjoncture bâtiment - taux d'utilisation des capacités de production	1	D_T
bat-apa	Enquête de conjoncture bâtiment - évolution passée de l'activité	1	D_T
bat-apre	Enquête de conjoncture bâtiment - évolution prévue de l'activité	2	D_T
CLIMAT-FR	Climat des affaires France	1	D_T
CLIMAT-FR-EMPL	Climat de l'emploi France	1	D_T
HIPC-FO	Indice des prix - nourriture	1	G_T
HIPC-IG	Indice des prix - produits industriels	1	G_T
HIPC-E	Indice des prix - énergie	1	G_T
HIPC	Indice des prix	1	G_T
IPI-CZ	Indice de production industrielle - industrie manufacturière	1	G_T

TABLE 1 – Variables utilisées dans l'analyse.

Lecture : $S_i = 1$ désigne les séries appartenant au groupe coïncident avec le cycle et $S_i = -1$ celles appartenant au groupe en avance sur le cycle.

À part les indicateurs de la comptabilité nationale et certains soldes d'opinion de l'enquête de conjoncture dans les services avant 2003, toutes les variables sont mensuelles. Pour les trimestrialiser on calcule leur moyenne sur le trimestre puis on calcule la différence entre deux trimestres (D_T) ou l'évolution trimestrielle (G_T).

Toutes les séries sont centrées-normées.

Références

- Kaufmann S. (2010). Dating and forecasting turning points by Bayesian clustering with dynamic structure: a suggestion with an application to Austrian data. *Journal of Applied Econometrics*, **25**(2): 309-344
- Frühwirth-Schnatter S, Kaufmann S. (2008). Model-based clustering of multiple time series. *Journal of Business and Economic Statistics* **26**(1): 78
- Bortoli C., Gorin Y., Olive P.-D. et Renne C. (2015), « De nouvelles avancées dans l'utilisation des enquêtes de conjoncture de l'Insee pour le diagnostic conjoncturel », *Note de Conjoncture*, mars, p. 21-41.
- Bardaji J., Clavel L. et Tallet F.,(2009), « Constructing a Markov-Switching Turning Point Index Using Mixed Frequencies with an Application to French Business Survey Data », *OECD Journal: Journal of Business Cycle Measurement and Analysis*, Vol. 2009/2.