

Modélisation et prévision du taux de chômage mensuel aux États-Unis

Yvo LE DOUDIC et Manh Hung NGUYEN

I. INTRODUCTION

Le taux de chômage est un indicateur économique clé lorsqu'il s'agit d'évaluer le dynamisme du marché du travail. Les causes d'un taux de chômage élevé peuvent être diverses : il peut être dû à une contraction conjoncturelle de l'économie ou bien à des frictions structurelles dans la recherche d'emploi, par exemple en raison d'un code du travail très exigeant. L'évolution du taux de chômage est donc très importante pour les institutions politiques et économiques – principalement les états et les banques centrales – pour leur permettre de prendre des décisions pertinentes en terme de politique financière et monétaire.

Le cas du taux de chômage des États-Unis d'Amérique est particulièrement intéressant en ce qu'il s'agit de l'un des plus grands marchés au monde couplé à une très faible régulation étatique. En résulte un taux de chômage structurel très bas, généralement proche du plein emploi, à savoir environ 4,5%. Par opposition et à titre d'exemple, la France, qui figure dans le haut du classement en terme de protection de l'emploi, affiche depuis plusieurs décennies un taux de chômage structurel proche de 8%.

Dans le cadre de ce projet, nous avons voulu modéliser le taux de chômage de sorte à pouvoir effectuer des prévisions mois par mois sur les années à venir. Notre matériau de base est une série temporelle de fréquence mensuelle retraçant l'évolution du taux de chômage aux États-Unis, de 1948 à 2023.

II. UN HISTORIQUE DU CHÔMAGE AUX ÉTATS-UNIS

La série temporelle brute dont nous disposons (cf. figure 1) présente deux obstacles à une modélisation efficace du taux de

chômage.

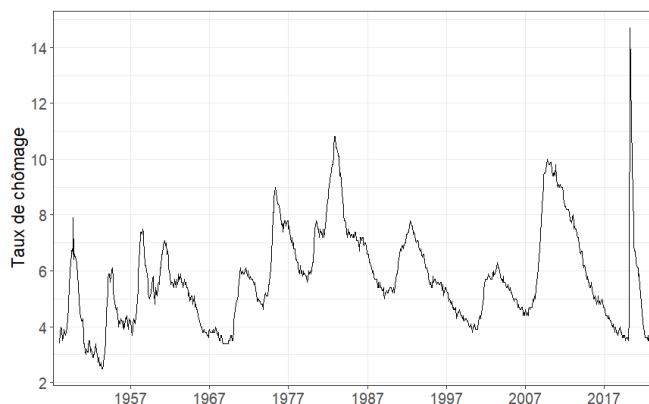


FIGURE 1. Variation mensuelle du taux de chômage (en %) de 1948 à 2023.

Le premier obstacle est que cette série couvre une période temporelle très longue, soixante-quinze ans plus exactement. La remarque que l'on peut faire, du point de vue de l'analyse macro-économique pure, est que chercher à modéliser un indicateur tel que le chômage sur une aussi longue durée n'est pas pertinent : la démographie, le progrès technologique, la productivité ainsi que l'ordre politique et économique mondial ont été bouleversés en moins d'un siècle et n'ont aujourd'hui plus rien à voir avec ce qu'ils étaient en 1948. Des trente glorieuses à la crise des *subprimes* en passant par la stagflation due aux chocs pétroliers des années 1970, de très différents paradigmes économiques se sont succédés en l'espace de soixante-quinze ans.

Il nous faut prendre en compte cette réalité. C'est pourquoi nous avons fait le choix de réduire la durée de cette série en la

faisant commencer en 1989. Cette date correspond à la chute du mur de Berlin et donc peu ou prou à l'effondrement du bloc soviétique et à la fin de la guerre froide. Cette année marque le passage du monde bipolaire issu de l'après-guerre à un monde unipolaire, relativement stable, dont les États-Unis sont la première puissance mondiale.

Le deuxième obstacle est lié à la pandémie de COVID-19 qui a frappé le monde en 2020. Le confinement mis en place en réponse à cette crise sanitaire unique a eu pour conséquence de faire exploser le taux de chômage d'un peu moins de 3,5% en février 2020 à 14,7% en avril 2022 – un niveau inédit aux États-Unis depuis la Grande Dépression – avant de redescendre vers son niveau pré-pandémie début 2022. Un tel événement étant absolument imprévisible et échappant à toute considération économique et financière, nous avons estimé qu'il était préférable d'en faire abstraction. Les modèles que nous avons construits sont donc basés sur une version réduite de cette série temporelle, interrompue en décembre 2019.

En revanche, les prévisions calculées à partir du modèle que l'on aura retenu seront, elles, effectuées sur la série de 1989 jusqu'à aujourd'hui, crise sanitaire comprise.

C'est donc à partir de l'évolution mensuelle du taux de chômage sur une période temporelle de trente ans, allant de 1989 à 2019, que nous allons pouvoir travailler (cf. figure 2).

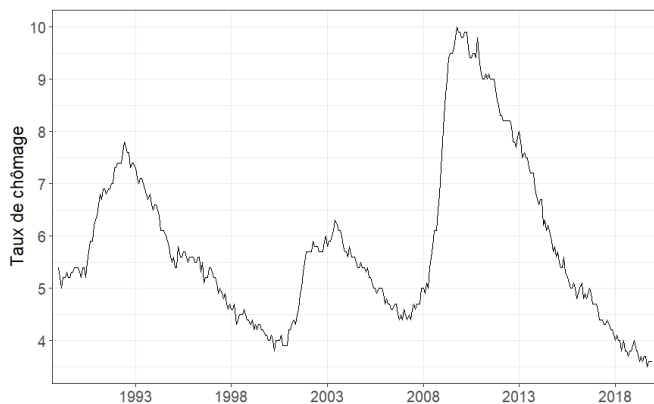


FIGURE 2. Variation mensuelle du taux de chômage (en %) de 1989 à 2019.

III. MODÉLISATION DU CHÔMAGE

Nous avons construit trois modèles différents. Deux de ces modèles sont linéaires – les modèles ARIMA et ARIMAX – tandis que le troisième est non linéaire – le modèle LSTAR. L'idée était de comparer ces trois modèles à l'aide de plusieurs indicateurs et de retenir celui dont les prévisions intra-échantillon étaient les plus pertinentes.

A. Modèle ARIMA

Le premier modèle utilisé est le modèle ARMA (pour *Autoregressive Moving Average*). L'idée générale de ce modèle est de décrire la série temporelle analysée comme une combinaison linéaire de ses valeurs et termes d'erreurs aléatoires passés. Étant donné que la série traitée ici n'est pas stationnaire, c'est le modèle ARIMA (pour *Autoregressive Integrated Moving Average*) qui a en réalité été utilisé. En effet, le *I* pour *Integrated* signifie que le modèle ARIMA prend en compte les différences entre les observations de la série temporelle plutôt que les valeurs brutes, ce qui permet de traiter une série non stationnaire.

Le choix d'un modèle ARIMA a du sens dans la mesure où le taux de chômage est une variable généralement très corrélée à ses valeurs passées : si le chômage est en hausse en mars, il est très probable qu'il le soit encore en avril de la même année et inversement. C'est à partir de ce postulat, sans introduire pour le moment de variable exogène, que nous avons modélisé le chômage.

En l'occurrence, le modèle ARIMA retenu est le suivant : ARIMA (1, 1, 3) (1, 0, 2). Le test de Ljung-Box renvoie une p-value de 0,71 largement supérieure au seuil α de 0,05. Ceci signifie que l'hypothèse nulle n'est pas rejetée, ce qui signifie que les résidus de notre modèle sont indépendants et non autocorrélés : le modèle estimé est de qualité.

Afin d'évaluer la performance de notre modèle, nous avons fait de la prévision intra-échantillon. Nous avons sélectionné une date arbitraire T^* , en l'occurrence le mois de décembre de l'année 2014. De cette façon, nous pouvons comparer les prévisions données par le modèle et les valeurs réelles. Pour ce modèle comme pour les suivants, nous avons déterminé les RMSE (pour *root-mean-square error*) aux horizons de prévision $h = 1$ et $h = 5$, soit un mois et cinq mois plus tard. Nous avons également calculé la statistique du *U* de Theil sur les horizons $h = 1$ et $h = 4$, soit un mois et quatre mois plus tard. Son expression est la suivante :

$$U = \frac{\text{RMSE du modèle retenu}}{\text{RMSE du modèle naïf}}$$

Si U est supérieure ou égale à 1, le modèle retenu est pire que le modèle naïf – qui se contente de répéter la dernière valeur connue en guise de prévision. Si U est inférieure à 1, le modèle retenu est plus prédictif que le modèle naïf.

Pour ce premier modèle ARIMA, les RMSE obtenus sont de 0,169 pour $h = 1$ et 0,157 pour $h = 5$; l'indicateur U de Theil, lui, vaut 1,694 pour $h = 1$ et 0,548 pour $h = 4$. Ce premier modèle ARIMA ne semble pas trop loin de la réalité en terme de prévisions intra-échantillon et, d'après le U de Theil, le modèle ARIMA est plus performant que le modèle naïf à l'horizon $h = 4$.

B. Modèle ARIMAX

S'il est pertinent de modéliser le chômage par rapport à ses propres valeurs passées, la prise en compte de facteurs externes peut améliorer la précision de nos prévisions. Le taux de chômage n'échappe pas à cette règle puisqu'il est complètement dépendant de la conjoncture économique du pays concerné.

William Phillips, un économiste néo-zélandais, est le premier à avoir eu l'intuition de l'existence d'une relation entre taux de chômage et taux d'inflation. L'existence de la courbe de Phillips, nommée en son honneur, a par la suite été vérifiée empiriquement, notamment par Paul Samuelson et Robert Solow dans les années 1960.

L'idée derrière cette courbe est que, lorsque le taux de chômage est faible, les salaires ont tendance à augmenter, ce qui entraîne une augmentation des coûts de production pour les entreprises. Les entreprises répercutent ces coûts supplémentaires sur les prix des biens et services qu'elles vendent, ce qui entraîne une augmentation de l'inflation. Inversement, lorsque le taux de chômage est élevé et que le travail se fait rare, les salaires ont tendance à baisser, ce qui réduit les coûts de production pour les entreprises. Les entreprises peuvent donc maintenir des prix bas pour attirer les consommateurs, ce qui limite l'inflation.

Cette relation négative se retrouve très bien sur nos données (cf. figure 3) : on constate, par exemple, que lors de la crise des subprimes de 2008, le chômage bondit jusqu'à 10% tandis que l'inflation¹, elle, chute à moins de 1%. Le chômage se résorbe ensuite lentement pendant que l'inflation récupère son niveau d'avant crise.

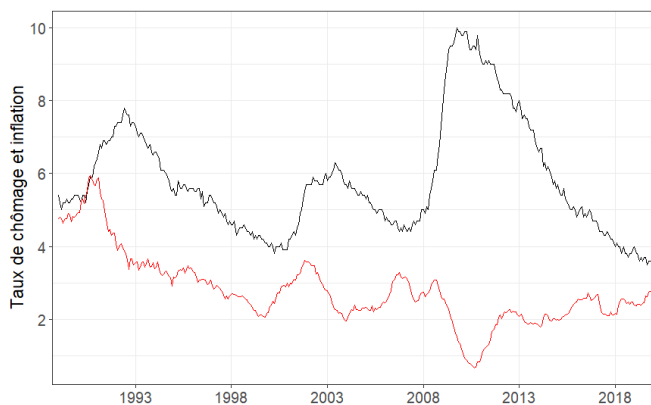


FIGURE 3. Variation mensuelle du taux de chômage en noir (en %) et de l'inflation en glissement annuel en rouge (en %) de 1989 à 2019.

Nous avons donc, dans un deuxième temps, construit un modèle prenant en compte l'évolution mensuelle de l'inflation – en glissement annuel – sur la même période. C'est là que

1. Données issues du site Federal Reserve Bank of St. Louis, Consumer Price Index

le modèle ARIMAX (pour *AutoRegressive Integrated Moving Average with exogenous variables*) entre en jeu. Il s'agit d'une extension du modèle ARIMA qui permet de modéliser les relations à court terme entre une variable endogène et une ou plusieurs variables exogènes. En l'occurrence, notre variable exogène est ici l'inflation.

Le modèle ARIMAX retenu est composé des mêmes ordres que le modèle ARIMA, à savoir : ARIMAX (1, 1, 3) (1, 0, 2). Cette fois, la p-value renvoyée par le test de Ljung-Box est de 0,98 ce qui signifie que l'hypothèse nulle n'est encore une fois pas rejetée et que notre modèle est pertinent.

Pour ce modèle ARIMAX, les RMSE obtenus sont de 0,169 pour $h = 1$ et 0,157 pour $h = 5$: ces valeurs sont très proches de celles obtenues pour le modèle ARIMA sans tenir compte de l'inflation. L'indicateur U de Theil se comporte de manière similaire puisqu'il vaut 1,694 pour $h = 1$ et 0,548 pour $h = 4$. Ce modèle ARIMAX semble assez largement équivalent au modèle ARIMA et ce malgré la présence de l'inflation en tant que variable exogène.

C. Modèle LSTAR

Le modèle LSTAR (pour *Logistic Smooth Transition AutoRegressive*) est un modèle économétrique non linéaire qui permet de modéliser des séries temporelles présentant des régimes de croissance différents. Ce modèle est capable de capturer les phénomènes de seuil, fréquents dans les données économiques et financières. La présence d'un seuil signifie que la relation entre deux variables données est non linéaire et que cette relation change à partir d'un certain seuil.

Il est par exemple possible que le taux de chômage soit très stable en deçà d'un certain seuil – par exemple 5% – mais explose et devienne très volatile au-delà. Cette réalité ne pourrait pas être capturée par les modèles linéaires du type ARMA.

Pour ce modèle LSTAR, les RMSE obtenus sont de 0,170 pour $h = 1$ et 0,216 pour $h = 5$: ces valeurs sont plus élevées que dans le cas des modèles ARIMA et ARIMAX, ce qui signifie que ce troisième modèle serait moins efficace. De la même manière, l'indicateur U de Theil vaut 1,700 pour $h = 1$ et 0,766 pour $h = 4$, ce qui indique un modèle meilleur que le naïf à l'horizon $h = 4$, mais moins bon que nos deux premiers modèles.

D. Bilan des modèles

Voici un tableau récapitulatif des performances de chacun des trois modèles que nous avons construits (cf. table 1).

Modèle	ARIMA			ARIMAX			LSTAR		
	1	4	5	1	4	5	1	4	5
RMSE	0,17	/	0,16	0,17	/	0,16	0,17	/	0,22
U	1,69	0,55	/	1,69	0,55	/	1,70	0,77	/

TABLE I

Afin de compléter cette comparaison des trois modèles, nous avons utilisé le test de comparaison de Diebold-Mariano. La seule condition requise à l'utilisation de ce test est que les modèles concernés ne soient pas emboîtés, ce qui n'est pas le cas pour nos modèles. L'hypothèse nulle testée ici est que la différence des fonctions de pertes associées aux modèles comparés est statistiquement égale à zéro : cela signifie que les deux modèles ont la même capacité prédictive.

Dans notre cas, les tests de Diebold-Mariano exécutés, en raison de p-values beaucoup trop élevées pour rejeter l'hypothèse nulle, n'ont pas permis de départager nos modèles. Ceci confirme la proximité en terme de performance annoncée par les RMSE et les U de Theil, tous très proches.

IV. PRÉVISION DE L'ÉVOLUTION DU TAUX DE CHÔMAGE

Enfin, nous avons procédé à une estimation de l'évolution du chômage sur les deux années à venir. Il faut garder en mémoire que la conjoncture économique mondiale actuelle est fortement marquée par une inflation élevée, atteignant 6,6% en décembre 2022 (cf. figure 4).

Cette hausse de l'inflation, en premier lieu liée à la pandémie, serait la conséquence de la rupture des chaînes d'approvisionnement ainsi que des aides monétaires massives de la part des gouvernements et des banques centrales en 2020 et en 2021. L'invasion russe de l'Ukraine en février 2022 a exacerbé cette situation en affectant à la hausse les prix mondiaux du pétrole et du gaz.

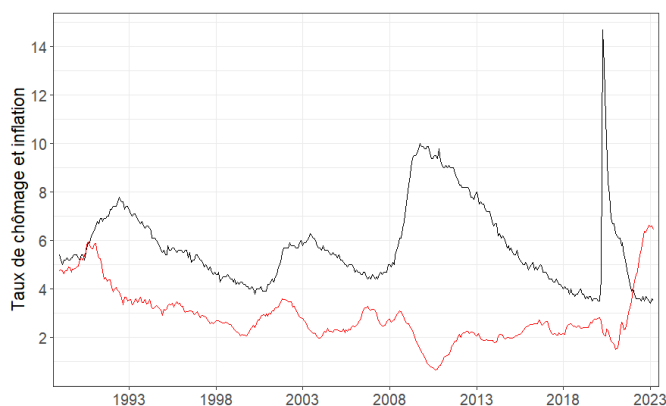


FIGURE 4. Variation mensuelle du taux de chômage en noir (en %) et de l'inflation en glissement annuel en rouge (en %) de 1989 à 2023.

Puisque rien ne semble départager nos modèles en terme de performance brute, nous avons fait le choix, au vu du contexte inflationniste décrit précédemment, d'utiliser notre

modèle ARIMAX qui tient justement compte de ce paramètre exogène.

Pour effectuer de la prévision à partir de ce modèle à variable exogène, il nous faut faire une hypothèse sur l'évolution de l'inflation elle-même. Nous faisons le choix de proposer deux scénarios historiquement réalistes :

- Dans le premier cas, possiblement en réaction à la hausse des taux directeurs par la Réserve fédérale, l'inflation chute et retrouve, en l'espace de 24 mois, un niveau de 2%.
- Dans le deuxième cas, l'inflation aux États-Unis se stabilise à son niveau actuel, autour de 6,5% : le monde occidental entre dans une période prolongée d'inflation élevée, à l'instar de la décennie 1970-1980.

A. Premier scénario : diminution rapide de l'inflation

Dans ce premier scénario, les valeurs futures de l'inflation fournies à notre modèle ARIMAX partent de la dernière valeur connue, soit environ 6,5%, et convergent en l'espace de deux ans vers 2%, soit le taux d'inflation généralement visé par la plupart des économies développées. Cette diminution de l'inflation est plausible puisque souhaitée par la Réserve fédérale qui, en augmentant ses taux directeurs, cherche à faire diminuer la demande et donc à limiter la "surchauffe" actuelle de l'économie.

Ce scénario, souhaitable pour les États-Unis, est a priori le plus probable d'après les analystes financiers.

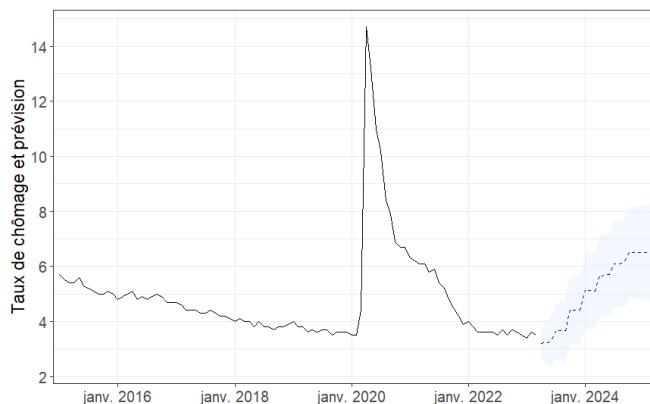


FIGURE 5. Variation mensuelle du taux de chômage (en %) de 2015 à 2023 et prévisions de 2024 à 2025 en cas d'inflation faible.

Dans ce premier cas de figure (cf. figure 5), si l'inflation se calme, le chômage, lui, repart à la hausse. Il finit par se stabiliser à partir d'octobre 2024 autour de 6,5%. Ce taux est deux fois plus élevé qu'aujourd'hui mais reste néanmoins relativement faible, même pour l'économie américaine. Cette estimation est cohérente avec la courbe de Phillips qui annonce une relation négative entre chômage et inflation.

B. Deuxième scénario : maintien d'une inflation élevée

Dans ce deuxième scénario, les valeurs de l'inflation fournies au modèle correspondent à une inflation "en plateau" restant constante à 6,5% sur les deux années à venir. Ce scénario, peu souhaitable, est moins plausible que le premier étant donnée la nature non structurelle de l'inflation.

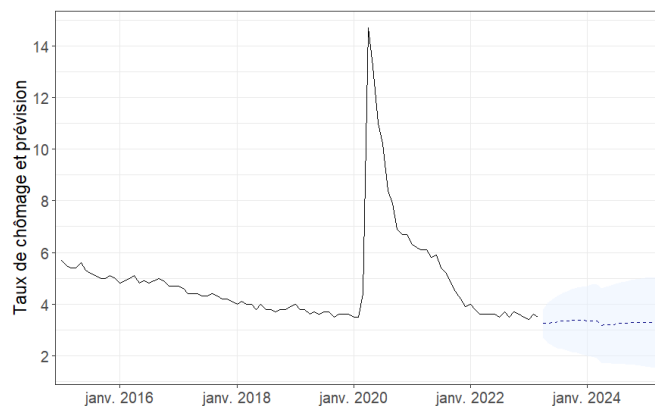


FIGURE 6. Variation mensuelle du taux de chômage (en %) de 2015 à 2023 et prévisions de 2024 à 2025 en cas d'inflation élevée.

Dans ce second cas de figure (cf. figure 6), fortement inflationniste, le chômage se maintient très bas, aux alentours de 3,5%. Encore une fois, ce résultat est cohérent avec la courbe de Phillips.

V. CONCLUSION ET RECOMMANDATION

Les deux scénarios envisagés ci-dessus correspondent à des réalités économiques dont les décideurs publics américains doivent tenir compte pour arbitrer.

La question peut se poser en ces termes : faut-il chercher à contenir l'inflation au prix d'un chômage élevé voire, éventuellement, d'une récession ? Si l'on part du principe qu'il faut prioritairement juguler l'inflation, en dessous de quelle valeur celle-ci serait jugée raisonnable ? Les pouvoirs publics américains sont-ils prêts à composer avec la hausse du taux de chômage qui en résulterait ?

Une fois dit cela, il faut garder en tête que ces modélisations économiques ne sont que des représentations simplifiées de la réalité. Afin de prendre des décisions appropriées, les institutions doivent également tenir compte d'autres facteurs extérieurs, qu'ils soient d'ordre politique ou sociologique. Les décideurs publics doivent envisager une approche plus complète, en tenant compte des contextes historiques, culturels et institutionnels spécifiques aux États-Unis.

Références :

- Christophe Tavera, *Cours de prévisions de séries temporelles de M2 MAS*, 2023
- Christophe Tavera, *Économie & Management*, Éditions Economica, 2021
- Forbes, *Inflation And Unemployment*, 2022
- Forbes, *When Will Inflation Go Down ?*, 2023
- Federal Reserve Bank of St. Louis, *Unemployment - The Economic Lowdown Video Series*