

---

## Nowcasting PIB : Imputation de variables non encore publiées

Marion CABROL, Kevin FERNANDES

Société Générale CIB, Core Data Scientists

marion.cabrol@sgcib.com, kevin.fernandes@sgcib.com

**Mots-clés** : PIB français, *machine learning*, imputation, séries temporelles, *nowcasting*

**Domaines** : *Nowcasting*, apprentissage statistique

---

Afin d'anticiper la position de l'économie dans son cycle, les macro-économistes doivent se doter d'outils robustes, permettant de fournir des prédictions immédiates ("*nowcasting*") de l'état de l'économie et plus spécifiquement de son agrégat macro-économique de référence : le Produit Intérieur Brut (PIB).

Dans ce contexte, l'article s'intéresse à l'élaboration de méthodes de prévision du PIB français pour le trimestre en cours via l'utilisation de données officielles issues d'institutions publiques (INSEE, Banque de France etc.) et privées (Markit). En d'autres termes, l'objectif est de prévoir le PIB en se fondant sur un ensemble de données qui inclut à la fois des données d'activité ("*hard-data*"), des données d'enquêtes ("*soft-data*") ainsi que des données des marchés financiers. Les premières données sont utilisées dans la construction du PIB (production industrielle, ventes au détail, etc.) alors que les secondes sont généralement des enquêtes de confiance. Ces variables à multifréquence sont pour la plupart publiées à différentes dates durant le trimestre qu'elles caractérisent. Elles peuvent également être revues au fur et à mesure du temps, tout comme la valeur du PIB est régulièrement réestimée pendant une période allant jusqu'à trois ans.

Néanmoins, l'objectif final de prédire le PIB de manière quasi-continue implique de pouvoir délivrer des prédictions sans pour autant attendre que l'ensemble des variables ait été publié. Ainsi, au début d'un trimestre, l'ensemble des variables explicatives portant sur ce même trimestre est partiellement vide. Généralement, seulement quelques enquêtes ont été réalisées. A contrario, en fin de trimestre, l'ensemble des variables explicatives est quasi-complet. Ainsi, plus nous avançons dans le trimestre et plus l'information portée par l'ensemble des variables explicatives s'enrichit.

Si les méthodes économétriques comme les modèles à facteurs dynamiques (Banbura et al., 2010 [1] et Bok et al., 2018 [3]) sont largement appliquées aujourd'hui et permettent d'ajuster les prévisions au fur et à mesure du trimestre, l'utilisation des méthodes d'apprentissage statistique ("*machine learning*") y est plus rare. En effet, l'emploi classique de ces méthodes requiert un ensemble prédéfini de variables observées et bien souvent une profondeur de données importante.

Afin d'utiliser le pouvoir prédictif des algorithmes d'apprentissage statistique tout en contournant les contraintes susmentionnées, cet article se propose d'étudier une démarche de modélisation en deux temps.

1. **Imputation des variables non encore publiées** : cette étape consiste à prévoir les données non encore publiées afin d'avoir un ensemble de variables explicatives complet à tout moment.
2. **Prédiction du PIB** : il s'agit de prédire le PIB à l'aide des variables déjà publiées et des variables explicatives imputées à l'étape précédente.

Cette approche se retrouve dans les travaux de Miller et al., 1996 [6] et de Zheng et al., 2006 [8].

Notre papier se concentre essentiellement sur la première étape de la modélisation d'"imputation des variables non encore publiées" et cherche à déterminer la méthode la plus adaptée. Pour ce faire, l'étude complète les techniques issues de la prévision des séries temporelles, par des méthodes non spécifiques au "*nowcasting*". L'article établit ainsi, une cartographie récente et d'autant plus exhaustive en s'inspirant, entre autres, de recherches liées à l'apprentissage statistique (Bertsimas et al., 2017 [2] ; Buuren et al., 2010 [4] ; Khan et al, 2020 [5]) et de recueils relatifs à l'imputation de données manquantes (Van Buuren, 2018 [7]). C'est d'ailleurs pour cette raison que nous avons fait le choix de nommer cette étape "**imputation de données non encore publiées**".

L'apport de chacune des méthodes est examiné par une procédure d'évaluation reposant sur deux types de mesure :

- Métriques intrinsèques : elles permettent de quantifier la qualité de l'imputation.
- Métriques extrinsèques : elles permettent de mesurer la tâche de prédiction du PIB.

On accorde cependant plus d'importance à l'objectif final de prédiction du PIB. On privilégie donc les méthodes qui surpassent sur les mesures extrinsèques tout en contrôlant la cohérence de leurs mesures intrinsèques.

Bien que le papier se concentre sur cette modélisation en deux étapes, il détaille également des solutions concrètes à d'autres défis imposés par la prédiction immédiate. Un de ces défis porte sur les données d'entraînement qui se doivent d'être fidèles à cette notion de "temps réel". En effet, la modélisation "*nowcasting*" se construit sur une chronologie réaliste du flux de données. L'article traite alors des méthodes de reconstruction de calendrier, de corrections de changements de nomenclature ainsi que des méthodes de traitement de séries temporelles adaptées aux révisions macro-économiques. Ces mêmes révisions qui, pour rappel, sont caractérisées à la fois par une période de valorisation et une date de publication.

## Références

- [1] Banbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting.
- [2] Bertsimas, D., Pawlowski, C., & Zhuo, Y. D. (2017). From predictive methods to missing data imputation : an optimization approach. *The Journal of Machine Learning Research*, 18(1), 7133-7171.
- [3] Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., & Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10, 615-643.
- [4] Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). mice : Multivariate imputation by chained equations in R. *Journal of statistical software*, 1-68.
- [5] Khan, S. I., & Hoque, A. S. M. L. (2020). SICE : an improved missing data imputation technique. *Journal of big data*, 7(1), 1-21.
- [6] Miller, P. J., & Chin, D. M. (1996). Using monthly data to improve quarterly model forecasts. *Federal Reserve Bank of Minneapolis Quarterly Review*, 20, 16-28.
- [7] Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.
- [8] Zheng, I. Y., & Rossiter, J. (2006). Using monthly indicators to predict quarterly GDP.