

# **Mémoire de Master 2**

Estimation de PIB par synthétisation économétrique de PMI

Paris 1 - Panthéon Sorbonne

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Importance d'estimer son PIB . . . . .	3
1.2	Présentation du PIB . . . . .	4
1.3	Présentation du PMI . . . . .	5
1.4	Objectif du mémoire . . . . .	6
<b>2</b>	<b>Revue de littérature</b>	<b>7</b>
2.1	Synthétisation économétrique d'Abadie . . . . .	7
2.2	Différence avec la méthode du <i>Diff and diff</i> . . . . .	7
2.3	Revue des applications de la méthode de contrôle synthétique . . . . .	8
<b>3</b>	<b>Méthodologie</b>	<b>9</b>
3.1	Fonctionnement de la synthétisation . . . . .	9
3.2	L'algorithme de synthétisation . . . . .	10
3.3	Estimer le PIB grâce au PMI . . . . .	11
<b>4</b>	<b>Données</b>	<b>12</b>
4.1	La base IHS Markit . . . . .	12
4.2	La base de l'INSEE . . . . .	12
4.3	La base de la FRED . . . . .	12
<b>5</b>	<b>Résultats</b>	<b>12</b>
5.1	Synthétisation de la France à partir du PMI composite . . . . .	12
5.2	Synthétisation des États-Unis à partir du PMI composite . . . . .	15
5.3	Estimation du PIB à l'aide du PMI . . . . .	19
5.3.1	Estimation du PIB français avec le PMI réel . . . . .	19
5.3.2	Estimation du PIB français avec le PMI synthétique . . . . .	23
5.3.3	Estimation du PIB américain avec le PMI réel . . . . .	26
5.3.4	Estimation du PIB américain avec le PMI synthétique . . . . .	28
<b>6</b>	<b>Conclusion</b>	<b>30</b>

## Abréviations

- FMI : Fonds Monétaire International
- INSEE : Institut National des Statistiques et des Études Économiques
- PGF : Productivité Globale des Facteurs de production
- PIB : Produit Intérieur Brut
- PMI : Purchasing Manager Index
- PSC : Pacte de Stabilité et de Croissance
- RMSE : Roots Mean Square Error (Racine de l'Erreur quadratique Moyenne)

# 1 Introduction

## 1.1 Importance d'estimer son PIB

Le Produit Intérieur Brut (ci après "PIB") s'est imposé comme la mesure de référence afin d'évaluer la santé d'une économie. Il permet d'observer le niveau de production sur le long terme au regard d'une absence de tensions inflationnistes ou déflationnistes. En d'autres termes, c'est la valeur totale de la production de richesses d'une économie. Bien qu'il soit toujours employé par ceux qui l'ont créé à savoir les économistes, il est aussi utilisé par les décideurs publics, les médias ou encore les institutions internationales. C'est notamment à sa lumière que sont construits les classements des États (FONTANEL et GUILHAUDIS, 2019). L'ensemble de ces éléments en fait donc un indicateur incontournable à la fois dans le champ académique que dans le champ politique. Néanmoins il est important de rappeler que le PIB ne permet pas de mesurer le bien être d'une population, le niveau de vie de ses habitants ou encore les inégalités présentes, pour plusieurs raisons (JANY-CATRICE et MÉDA, 2011). C'est d'ailleurs afin d'être en mesure de capter toutes les facettes de l'économie que la Commission sur la mesure des performances économiques et du progrès social dite "Stiglitz" s'est penchée sur les mesures des performances économiques à la demande du président Sarkozy (STIGLITZ et SEN, 2009).

La capacité à estimer le PIB d'un pays est primordiale. En effet, les programmes de consolidation budgétaire s'appuient sur le PIB pour être en mesure de fixer une cible de déficit conjoncturel puis structurel. Par exemple le Pacte de stabilité et de croissance (PSC), adopté à Amsterdam le 17 juin 1997 qui permet de coordonner les politiques budgétaires nationales et d'éviter l'apparition de déficits budgétaires trop importants au sein de la zone Euro, repose en grande partie sur le PIB (LARCH, 2014). Le PSC impose aux Etats membres de maintenir leur déficit et leur dette publique en dessous des seuils fixés respectivement à 3% et 60% du PIB. L'importance d'une bonne estimation du PIB est donc évidente en Europe. Il en va de même pour les pays se trouvant à l'extérieur de la zone Euro. Les institutions internationales telles que la Banque Mondiale ou le Fonds monétaire international (FMI), regarde de très près le PIB des pays à qui ils viennent en aide dans le cadre du développement et du maintien de la stabilité du système monétaire international. Au-delà de ces aspects, c'est sur cet indice que repose une grande partie de l'analyse macroéconomique, entraînant parfois des décisions politiques. Ce qui explique la nécessité d'être en mesure de le calculer de manière précise. De plus, il est important de pouvoir obtenir les chiffres du PIB rapidement pour être en mesure d'observer l'état actuel de l'économie.

En France c'est l'Institut National des Statistiques et des Études Économiques (INSEE) qui est chargée de calculer le PIB. Tout d'abord le PIB est calculé à une fréquence trimestrielle, à savoir quatre fois par an. Néanmoins afin d'obtenir la valeur définitive du PIB d'un Etat il faut trois ans. En effet, le temps de récolte des données provenant de l'ensemble des acteurs économiques ne permet pas de connaître la valeur du PIB en temps réel. Ces données, nécessaires au calcul du PIB, récupérées auprès des entreprises, des banques, des administrations publiques, etc ... permettent d'observer la valeur ajoutée des différents organes de l'économie. Ces données s'appellent les *hard data* en opposi-

tion aux *soft data*. Les *soft data* sont les données récoltées grâce à des enquêtes auprès des entreprises ou encore des sondages. Elles ont l'avantage de pouvoir être obtenues rapidement et ainsi d'estimer de manière provisoire le PIB à l'instant  $t$ . Le premier chiffre provisoire du PIB est publié un mois et demi après la fin du trimestre concerné. L'estimation provisoire du PIB est essentielle à la décision publique. L'élaboration et la mise en place de politiques publiques dépendent de ces estimations ce qui exige un niveau de précision et une vitesse de publication suffisante. Les estimations dont il sera question dans ce mémoire sont ces estimations provisoires (INSEE, 2017).

## 1.2 Présentation du PIB

Tout d'abord il semble important de rappeler comment est théoriquement construit le PIB. Il existe trois approches du PIB : par les dépenses, par les revenus moyens et enfin par la valeur ajoutée. Cette partie présentera ces différentes conceptualisations du PIB. Il faut néanmoins rappeler qu'il s'agit ici d'un rappel de définition du PIB et non de méthodes de calcul comme nous le verrons dans la suite du mémoire.

Il est possible de calculer le PIB par les dépenses au moyen de l'équation suivante :

$$PIB = C + G + I + (X - M)$$

Où  $C$  représente les dépenses de consommation,  $G$  les dépenses gouvernementales,  $I$  les investissements privés et publics,  $X$  les exportations et  $M$  les importations. Ainsi en additionnant toutes les dépenses de consommation des personnes d'une région, que ce soit en biens ou en services, les dépenses du gouvernement telles que les salaires des fonctionnaires ou dans les infrastructures publiques, les investissements privés et publics des entreprises pour offrir leurs biens et services puis la différence entre les exportations et les importations du pays, nous obtenons la mesure du PIB.

Le PIB peut ensuite se calculer par les revenus au moyen de l'équation suivante :

$$PIB = Sal + RM + EBE + Tin$$

Où  $Sal$  représente le total de la rémunération des travailleurs,  $RM$  le revenu mixte net des entreprises non incorporées,  $EBE$  les excédents bruts d'exploitation et  $Tin$  les taxes indirectes nettes sur la production et les importations. Ainsi en additionnant la masse salariale d'une région donnée, les autres types de revenus liés à la location, aux intérêts ou encore aux dividendes, les profits des entreprises et puis les taxes prélevées sur la production et les importations, nous obtenons la mesure du PIB.

Enfin le PIB peut se calculer par la valeur ajoutée au moyen de l'équation suivante :

$$PIB = Cout\ final - Prix\ des\ matieres\ premieres\ et\ des\ etapes\ intermediaires + Taxes$$

Ainsi en tenant compte du coût des produits finaux, le coût de la production des produits et des taxes que ces produits permettent de lever sur l'ensemble des secteurs et des industries d'une économie, nous obtenons la mesure du PIB. Cependant la liste des méthodes théoriques ci-dessus indiquent

aux moyens de quels versants de l'économie le PIB peut être calculé. S'agissant des méthodes il faut savoir que, comme mentionné précédemment, la récolte des données est longue. C'est pourquoi l'estimation provisoire du PIB est primordiale. Elle revêt une importance toute particulière pour la décision publique notamment.

Pour estimer le PIB de manière rapide il n'est pas possible de reposer les calculs sur les *hard data*. Il faut donc utiliser des méthodes particulières. Ces méthodes sont variées et n'offrent pas toutes le même niveau de précision. Elles peuvent se baser sur des enquêtes de conjoncture, des données économiques ou des séries financières. Les deux principaux types de méthodes d'estimation du PIB provisoire sont : les méthodes par sondage et les méthodes dites structurelles.

La première méthode, qui sera exposé plus en détail dans ce mémoire, repose donc sur des sondages effectués auprès de différents agents économiques. À la lumière des données des précédentes périodes, les nouvelles données récoltées permettront au moyen d'une régression linéaire d'estimer le PIB provisoire en très peu de temps. Naturellement la précision de cette estimation dépendra de la profondeur de données disponible. C'est ce que nous ferons dans ce mémoire à l'aide des données PMI que nous allons voir.

La deuxième méthode repose sur l'estimation des trois facteurs : travail, capital et productivité globale des facteurs de production (PGF). Cela peut s'estimer de deux manières : soit en considérant la PGF comme résiduelle, après avoir calculé la croissance des facteurs capital et travail ou alors en modélisant une fonction de production (JURUS, 2013).

### 1.3 Présentation du PMI

L'estimation du PIB d'un pays peut donc se faire sur la base de sondages tels que le Purchasing Manager Index (PMI). Le PMI est un indicateur permettant d'observer l'orientation dominante de la tendance économique dans les secteurs de la fabrication et des services. Il montre les conditions du marché ainsi que son expansion, tel que le perçoit les directeurs d'achats. Il permet aux décideurs d'entreprises ou aux analystes par exemple, d'évaluer les conditions commerciales actuelles. Cet indice a un avantage de taille : sa rapidité de diffusion et ce particulièrement dans un contexte d'incertitude important. Les enquêtes mensuelles permettant de construire les indices PMI, sont menées par plusieurs institutions telles que l'*Institute for Supply Management*, le *Singapore Institute of Purchasing and Materials Management* et *Markit Economics*. L'indice est donc construit sur la base d'enquêtes réalisées auprès des directeurs d'achats d'entreprises du secteur manufacturier et des services. Les entreprises sélectionnées pour mener cette enquête, sont choisies de façon à représenter le plus fidèlement possible la structure de l'économie. De plus, ces enquêtes sont réalisées de manière identique dans tous les pays sondés, permettant ainsi de pouvoir les comparer.

Les enquêtes menées ont pour objectif d'établir la variation observée de plusieurs caractéristiques mesurables par les directeurs d'achats par rapport au mois précédent. Ces caractéristiques mesurables sont : le niveau de production, l'emploi, les nouvelles commandes, les stocks, les délais de livraison et les prix. Les entreprises interrogées ont le choix entre trois réponses : niveau supérieur, niveau égal ou niveau inférieur à celui du mois précédent. Sera construit ensuite ce que l'on appelle le PMI

Composite. C'est une moyenne pondérée du PMI des secteurs manufacturiers et des services pour une économie donnée. Les pondérations de ces secteurs découlent directement de la part de ces secteurs dans le PIB.

Une fois les enquêtes réalisées, une pondération est appliquée aux réponses des entreprises en fonction de leur taille avant d'être transformées en indices de diffusion. Un PMI égal à 50 indique qu'il n'y a pas eu de changement ce mois comparativement au mois précédent, moins de 50 signifie qu'il y a davantage de réponses négatives et plus de 50 signifie qu'il y a davantage de réponses positives (MARKIT, 2022). Les résultats ne sont pas révisés après la première publication.

L'indice PMI présente de nombreux avantages. D'abord il est basé sur des enquêtes reposant sur des données réelles et non sur des anticipations ou des opinions, ce qui limite les possibilités d'erreurs d'estimation de la part des directeurs d'achats. De plus, les variations observées sont effectuées pour plusieurs variables ce qui permet de mieux définir la nature d'un ralentissement ou d'une expansion économique. Le PMI est diffusé à la fin de chaque mois ce qui en fait un indice particulièrement réactif. L'enquête est effectuée auprès d'un grand nombre d'entreprises qui y participent chaque mois, permettant ainsi de maintenir une cohérence et une comparabilité des données dans le temps. Enfin la méthodologie de l'enquête est identique dans les nombreux pays où elle est menée, ce qui permet une comparaison internationale comme expliqué précédemment.

Le PMI se heurte aussi à certaines limites. Par exemple il ne comprend pas tous les secteurs de l'économie ce qui peut être un problème lorsque l'on veut estimer le PIB si un des secteurs exclus évolue de manière opposée au reste de l'économie. Une autre limite réside dans la structure même de l'enquête. Le sondage permet de connaître la direction des variations que subissent certains secteurs de l'économie cependant l'amplitude de ces variations n'est pas estimable à l'échelle microéconomique. De fait des variations différentes de l'activité d'une même entreprise se verront pondérées de la même manière.

## 1.4 Objectif du mémoire

L'objectif de ce mémoire est de rechercher si l'estimation du PIB peut être effectuée avec précision lorsque l'on utilise des valeurs synthétisées du PMI. La méthode d'estimation du PIB sur la base du PMI<sup>1</sup> est significative et améliore considérablement la qualité des prévisions de l'économie (GODBOUT et JACOB, 2010). Cependant cette méthode exige une certaine profondeur de données du PMI pour être précise<sup>2</sup>. Cependant de nombreux pays ne recensent leur PMI que depuis peu de temps, ce qui ne leur permet pas d'estimer avec précision leur PIB à l'aide de cette méthode. L'idée est donc de synthétiser les valeurs non recensées du PMI grâce à la série disponible et aux séries des pays qui ont une plus grande profondeur de données.

---

1. La méthodologie permettant d'estimer le PIB sur la base du PMI est explicitée dans la partie méthodologie du mémoire.

2. La significativité des coefficients que l'on obtient lorsque l'on régresse le PIB avec le PMI augmente avec le nombre d'observations. Ce qui est directement lié aux degrés de liberté.

## 2 Revue de littérature

### 2.1 Synthétisation économétrique d'Abadie

La méthode de contrôle synthétique développée par Abadie et Gardeazabal est une approche qui a su faire ses preuves dans le monde de la recherche empirique (ABADIE et GARDEAZABAL, 2003). Elle est d'ailleurs considérée par certains comme l'innovation la plus importante dans la littérature de l'évaluation des politiques publiques depuis ces 15 dernières années (ATHEY et IMBENS, 2017). Elle permet de créer un contrefactuel pour une suite de données traitées grâce à un groupe de suites dit "donateur". L'idée est de construire une version synthétique d'une série de données en pondérant les séries du groupe donateur minimisant la distance entre notre version synthétique et la série de notre unité traitée. Cette méthode s'avère particulièrement efficace pour observer et quantifier l'effet d'un choc exogène subi seulement par l'unité traitée. En effet, en observant l'écart entre les données de notre version synthétique et de notre unité traitée après le choc exogène, il est possible de calculer l'effet du traitement sur notre unité traitée. De plus, contrairement aux techniques classiques d'analyse de régression la synthétisation permet de connaître la contribution de chaque unité comparative à notre contrefactuel. Dans ce mémoire l'usage de cette méthode va être un peu particulier. En effet, l'objectif ne va pas être d'observer l'effet d'un traitement sur une unité, mais simplement de prolonger une série présentant des données manquantes. Cette recherche repose sur l'intuition que les données de PMI que nous n'avons pas, s'inscrivent dans une certaine continuité et une logique similaire aux données du groupe donateur. Ce qui est en parfaite adéquation avec la philosophie qui se cache derrière la méthode de contrôle synthétique, selon laquelle lorsqu'une unité observée est un agrégat d'un certain nombre d'entités, une combinaison de ces mêmes entités permet de synthétiser les valeurs de l'unité observée (ABADIE, 2021).

### 2.2 Différence avec la méthode du *Diff and diff*

En économétrie, et en particulier dans l'évaluation des politiques publiques, la méthode du *Diff and diff* (à effets fixes) est largement utilisée. Elle peut s'apparenter à la méthode de contrôle synthétique c'est pourquoi il est important de pointer ce qui différencie ces approches et donc justifie l'emploi du contrôle synthétique. Le *Diff and diff* permet d'observer l'effet d'un traitement en comparant un groupe traité avec un groupe de contrôle sur une variable dépendante. Cette méthode est plus rigide et se cantonne à l'évaluation de l'effet d'un traitement. Elle peut donc se rapprocher de l'usage classique qui est fait du contrôle synthétique<sup>3</sup>. Cependant ce mémoire ne cherche pas à évaluer l'effet d'un traitement mais bien à créer une série synthétique du PMI qui permettra d'estimer le PIB.

---

3. Voir partie sur la revue de ses applications dans différents secteurs.



## 2.3 Revue des applications de la méthode de contrôle synthétique

La méthode de contrôle synthétique a été utilisée dans de nombreuses études. Cette partie présentera brièvement les domaines dans lesquelles elle a su faire ses preuves. Elle a permis notamment d'observer l'effet des lois sur le droit du port d'armes (DONOHUE et al., 2018), de la légalisation de la prostitution (CUNNINGHAM et al., 2018), des politiques d'immigration (BOHN et al., 2014), des connexions politiques des entreprises (ACEMOGLU et al., 2016), du crime organisé (PINOTTI, 2015). Elle a aussi été utilisée comme un outil d'analyse lors des débats sur les effets de l'immigration (BORJAS, 2017) ou encore sur les effets de la fixation du niveau de salaire minimum (ALLEGRETO et al., 2017, NEUMARK et WASHER, 2017). La méthode a su trouver des applications dans d'autres domaines que l'économie tels que les sciences sociales, ingénierie ou la biomédecine (HEERSINK et al., 2017, PIETERS et al., 2017). Au-delà du monde académique cette méthode a été adoptée dans de nombreuses organisations multilatérales, des *think tanks*, des agences gouvernementales ou des entreprises de conseil. Par exemple la méthode a été largement utilisée pour mesurer les effets des partenariats pour le programme d'enseignement de la fondation de Bill et Melinda Gates (GUTIERREZ et al., 2016). Cet aperçu non exhaustif des domaines dans lesquels cette méthode a été appliquée montre sa fiabilité. Bien qu'elle fût utilisée dans des champs divers et variés, l'approche par synthétisation économétrique ne semble jamais avoir été utilisée dans un objectif de synthétisation de PMI pour estimer le PIB.

### 3 Méthodologie

#### 3.1 Fonctionnement de la synthétisation

Comme vu précédemment, la synthétisation d'une série se fait à partir d'un groupe de séries donateur. Cette partie traitera de la formalisation mathématique de la méthode.

Lors de la synthétisation on considère un échantillon de  $J + 1$  unités avec  $j \in J$ .  $j = 1$  va être notre sujet d'analyse, les valeurs de  $j = 2$  à  $j = J + 1$  constitueront notre groupe de donateurs. Il faut que notre groupe de donateurs soit influencé par des processus structurels similaires. L'analyse se fait sur des données qui comprennent plusieurs observations au cours du temps d'un même individu statistique, que l'on appelle données de panel ou données longitudinales. La contribution de chaque unité de notre groupe donateur dans la construction de notre série synthétique est pondérée par une valeur  $w$ . Les pondérations de chacune des valeurs de notre groupe de donateurs sont ensuite regroupées dans un vecteur de taille  $(J \times 1)$  des pondérations tel que :  $W = (w_2, \dots, w_{J+1})$  sachant que  $w_2 + \dots + w_{J+1} = 1$ . Afin de débiter la synthétisation on construit la matrice  $X_1$  de taille  $(k \times 1)$  qui contient les valeurs des caractéristiques de notre unité traitée. Ensuite on construit la matrice  $X_0$  de taille  $(k \times J)$  qui contient les valeurs des caractéristiques de notre groupe donateur. Par conséquent la différence entre les caractéristiques de l'unité traitée avant et après l'intervention est égale à  $X_1 - X_0$ . Enfin, pour construire notre unité synthétique, on choisit le vecteur  $W^*$  tel que la différence entre  $X_1$  et  $X_0 W^*$  soit la plus faible possible. Ainsi le résultat du contrôle synthétique sera tel qu'il permettra d'approcher au mieux l'unité traitée.

Ce qui nous donne :

$$\sum_{m=1}^k v_m (X_{1m} - X_{0m} W)^2 \quad (1)$$

Où  $v_m$  est la pondération de l'importance relative de la variable  $m$  quand on mesure la divergence entre  $X_1$  et  $X_0 W$ .

Dans l'utilisation qui est généralement faite de cette méthode, les auteurs cherchent à évaluer l'effet d'un choc exogène sur des données d'un pays, une région ou une entreprise. Pour ce faire nous séparons nos périodes en deux échantillons. On considère les périodes telles que  $t = 1, \dots, T$ . L'évènement a lieu en  $T_0 + 1$  ce qui implique que les périodes allant de  $t = 1, \dots, T_0$  sont avant l'intervention et  $t = T_0 + 1, T_0 + 2, \dots, T$  sont après l'intervention.

Deux résultats potentiels :

- $Y_{it}^N$  pour le résultat qui sera observé pour l'unité  $i$  à la période  $t$  si l'unité  $i$  n'est pas exposé à l'intervention.
- $Y_{it}^I$  pour le résultat si  $i$  est exposé à l'intervention

L'effet de l'intervention est donc égal à la différence entre les deux résultats possibles à savoir :

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N \text{ pour les périodes } T_0 + 1, T_0 + 2, \dots, T, \hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

Nous rappelons que ce mémoire ne cherche pas à estimer l'effet d'un traitement, mais seulement à synthétiser notre série sur la base des données disponibles. Ainsi nous n'avons pas à observer de

différence entre  $Y_{it}^I$  et  $Y_{it}^N$  (que nous ne pourrons de toute manière pas faire étant donné que nous n'avons pas la série  $Y_{it}^N$ ). Nous prolongeons seulement  $Y_{it}^I$  et regardons ensuite si sur la base de cette série l'estimation du PIB est suffisamment précise. Le but de la méthode étant de construire une série synthétique qui donne une estimation raisonnable des données manquantes. Pour ce faire, l'idéal serait donc de construire le groupe donateur de manière à ce qu'il soit sensible à toutes les caractéristiques pertinentes du PMI.

Pour construire notre estimateur de contrôle synthétique numériquement il faut que l'on définisse la distance entre notre unité de contrôle synthétique et notre unité traitée. La fonction **synth()** qui est utilisé sur R pour créer cette série synthétique va permettre de résoudre l'équation suivante :

$$\|X_1 - X_0W\|_V = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)} \quad (2)$$

Cette équation correspond au LOSS (W) dans l'affichage du code R.

$W^*$  réduisant donc cette distance à son minimum et qui peut être trouvé grâce au programme Synth.  $V$  étant la matrice de taille  $(k \times k)$  symétrique. Cette matrice est introduite pour permettre différentes pondérations des variables dans  $X_0$  et  $X_1$  en prenant leur pouvoir prédictif sur le résultat. Si  $V$  optimal alors la moyenne des carrés des erreurs est minimisée pour l'estimateur de contrôle synthétique, ce qui revient à notre prévision. Le programme Synth permet d'avoir une bonne flexibilité sur le choix de  $V$ . Ce choix peut se faire manuellement ou automatiquement. Lorsque l'on fait le choix de définir ce  $V^*$  de manière automatique le programme résout ce terme :

$$\arg \min_{V \in \mathcal{V}} (Z_1 - Z_0W^*(V))' (Z_1 - Z_0W^*(V)) \quad (3)$$

Cette équation correspond au LOSS (V) dans l'affichage du code R.

$\mathcal{V}$  est l'ensemble des matrices positives et diagonales.

Pour conclure, le programme de synthétisation va donc créer notre unité synthétique en pondérant de manière optimale (c'est à dire en minimisant la racine de l'erreur quadratique moyenne) la contribution des unités de notre groupe donateur dans la création de notre unité synthétique, ainsi qu'en pondérant de manière optimale la contribution des variables prédictives. Les deux vecteurs créés sont donc  $W^*$  et  $V^*$ .

### 3.2 L'algorithme de synthétisation

Afin d'effectuer cette synthétisation nous avons utilisé le package **synth()** sur R. Cet algorithme permet d'abord de construire les sous bases  $X_0$ ,  $X_1$ ,  $Y_0$ ,  $Y_1$  nécessaire à la synthétisation<sup>4</sup>. Cette étape demande une grande rigueur lors du transfert de la base de données sur R. Ensuite la fonction **synth()** va nous permettre d'obtenir le vecteur  $W^*$  des pondérations optimales des pays du groupe donateur. En pondérant les PMI de chaque pays du groupe donateur nous parvenons à construire

---

4. A l'aide de la fonction **dataprep()**

notre série synthétique. Enfin, pour visualiser la pertinence de notre construction, nous traçons la série synthétique sur le même graphique que notre série réelle. On peut aussi tracer un graphique représentant l'écart à chaque date entre notre série synthétique et notre série réelle.

Cependant la pertinence de notre synthétisation pourra être constatée lors de l'estimation du PIB grâce à notre série synthétique. En comparant la précision de notre estimation en utilisant la série synthétique avec la précision de l'estimation que l'on peut faire avec les vraies valeurs du PMI, nous pourrions être en mesure d'évaluer la robustesse de l'usage de la méthode de synthétisation.

### 3.3 Estimer le PIB grâce au PMI

L'estimation du PIB sur la base du PMI se fait au moyen d'une régression linéaire. Cependant la précision de notre estimation dépendra du nombre d'observation de la série PMI qui augmentera les degrés de liberté de notre régression et donc la significativité de nos coefficients. Les coefficients et l'ordonnée à l'origine (*intercept*) que nous trouvons permettent ensuite de créer notre série synthétique.

La régression linéaire que nous menons est la suivante :

$$\text{deltaPIBtrim} \sim \text{PMI\_M1} + \text{PMI\_M2} + \text{PMI\_M3} \quad (4)$$

Où *deltaPIBtrim* est la différence entre le PIB au trimestre  $t$  et celui au trimestre  $t - 1$  que nous retrouvons dans la base de données de l'INSEE, *PMI\_M1* est le vecteur du PMI des premiers mois de chaque trimestre, *PMI\_M2* le vecteur du PMI des deuxièmes mois de chaque trimestre et *PMI\_M3* celui du PMI des troisièmes mois de chaque trimestre. Cette régression nous permettra de trouver les coefficients ( $\beta$ ) et l'*intercept* ( $\alpha$ ) de l'équation suivante :

$$\text{deltaPIBtrim} = \alpha + \beta_1 \text{PMI\_M1} + \beta_2 \text{PMI\_M2} + \beta_3 \text{PMI\_M3} \quad (5)$$

Lorsque nous avons donc nos coefficients et notre ordonnée à l'origine, il suffit de les multiplier avec nos vecteurs des PMI de chaque mois des trimestres afin de reconstruire notre série de PIB estimée. Il est primordial de vérifier la précision de notre estimation en calculant sa racine de l'erreur quadratique moyenne. La formule de la racine de l'erreur quadratique moyenne (*Roots Mean Square Error*) est la suivante :

$$\sqrt{(\text{deltaPIBreal} - \text{deltaPIBestime})^2 / \text{NbrObservations}} \quad (6)$$

Ainsi une estimation précise sera vérifiée par un RMSE faible, signifiant que l'écart absolu moyen entre notre estimation et la réalité est faible.

## 4 Données

### 4.1 La base IHS Markit

La base de données qui est utilisée dans ce mémoire est récupérée auprès de l'IHS Markit. Elle s'étend sur un ensemble de 20 pays à savoir : la France (1), l'Allemagne (2), l'Italie (3), l'Espagne (4), le Royaume Uni (5), les États-Unis (6), la Chine (7), l'Inde (8), le Brésil (9), le Japon (10), la Russie (11), la Pologne (12), le Mexique (13), le Canada (14), la Turquie (15), les Pays Bas (16), l'Autriche (17), l'Indonésie (18), la Nouvelle-Zélande (19) et l'Égypte (20). La base de données nous permet d'observer plusieurs indicateurs de PMI pour chacun de ses pays, à savoir : le composite<sup>5</sup>. Dans le secteur de l'industrie on retrouve les indicateurs suivants : la synthétique industrie, la production passée, l'emploi, les nouvelles commandes, les stocks, les délais de livraisons, les commandes export, le prix des output et le prix des input. Dans le secteur des services on retrouve les indicateurs suivants : l'activité passée, les nouvelles affaires, l'emploi, les retards (backlogs of work), l'activité future, le prix des output et le prix des input. Pour des raisons techniques nous enlèveront la Nouvelle Zélande de notre base de données<sup>6</sup>. Cette base de données présente l'avantage de regrouper des séries de PMI de plusieurs pays. Cela nous permettra de pouvoir facilement créer notre groupe donateur lorsque l'on synthétisera une série PMI.

### 4.2 La base de l'INSEE

Pour la partie estimation du PIB nous avons besoin de données du PIB absente de notre base IHS Markit. Étant donné que dans le cadre de ce mémoire nous allons estimer le PIB de la France sur la base de son PMI réel et du PMI synthétique que nous aurons construit, nous utiliserons la base de l'INSEE, disponible sur leur site internet, qui recense le PIB trimestriel de la France ainsi que le delta du PIB<sup>7</sup>.

### 4.3 La base de la FRED

Nous avons utilisé une base de données disponible sur le site de la FRED Economic data pour retrouver le PIB trimestriel des États-Unis afin de pouvoir l'estimer avec nos PMI. Nos observations de PIB vont de 1999 à 2021.

## 5 Résultats

### 5.1 Synthétisation de la France à partir du PMI composite

Afin de visualiser le fonctionnement de la synthétisation voyons d'abord une série simple de données de PMI, pour la France par exemple. On voit donc l'évolution du PMI composite de la

---

5. Aux États-Unis le  $PMI_{Composite} = 0,40 * PMI_{manufacturier} + 0,60 * PMI_{services}$

6. Pour la Nouvelle Zélande, le PMI n'est recensé que pour 251 périodes et non 252 comme le reste des pays du groupe donateur, entraînant des complications inutiles pour le fonctionnement de l'algorithme de synthétisation.

7.  $\Delta PIB_t = PIB_t - PIB_{t-1}$

France sur la période du 01/09/1999 au 01/12/2019.

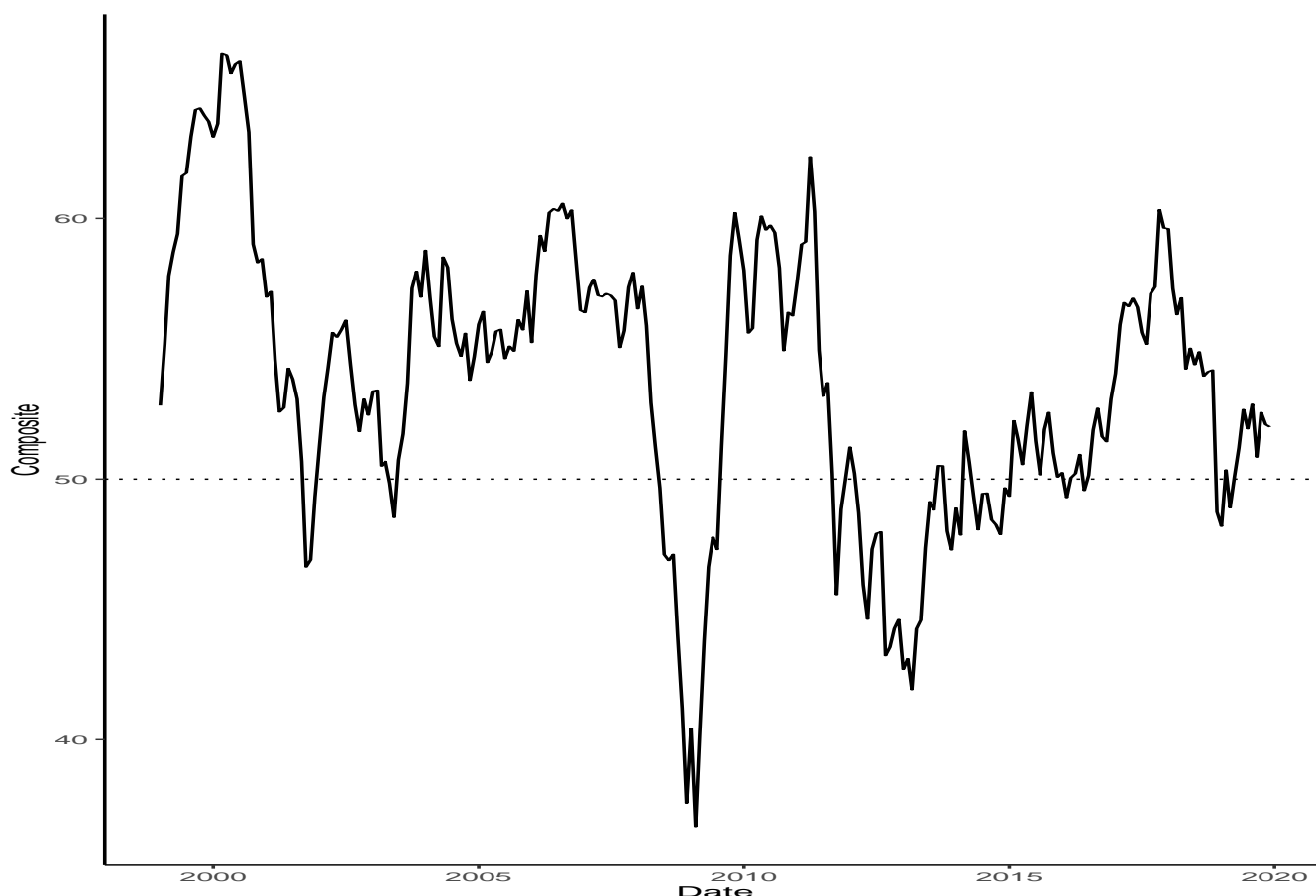


FIGURE 1 – PMI de la France de 1999-09-01 à 2019-12-01

Cette série de valeurs du PMI permet de visualiser la dynamique de l'économie Française durant cette période.

Le programme de synthétisation développé par Abadie disponible sur R sous le nom du package **synth()** permet donc de construire une suite synthétique de la France grâce à un groupe de pays donateurs. Cependant le groupe de pays donateurs doit contenir des séries de données de même taille que le pays que l'on cherche à synthétiser. En d'autres termes, l'algorithme ne permet pas de synthétiser automatiquement une série si le groupe donateur ou le pays que l'on cherche à synthétiser comporte des valeurs manquantes. Afin de créer notre série synthétique de la France nous devons donc choisir des pays dans notre groupe donateur qui ont le même nombre d'observations du PMI. En restreignant le groupe donateur aux pays qui ont la même profondeur de données on synthétise donc la France avec l'Allemagne, l'Italie, l'Espagne et l'Angleterre. Le vecteur de pondération montrant la contribution de chacun de ces pays à la construction de notre France synthétique est le suivant :

TABLE 1 – Contribution des pays donateurs

Pondérations	
Allemagne	0.173
Italie	0.088
Espagne	0.083
Angleterre	0.656

Maintenant que nous avons la contribution de ces pays à la construction de la France synthétique, nous l'appliquons aux valeurs du PMI de chacun de nos pays afin de tracer le graphique suivant :

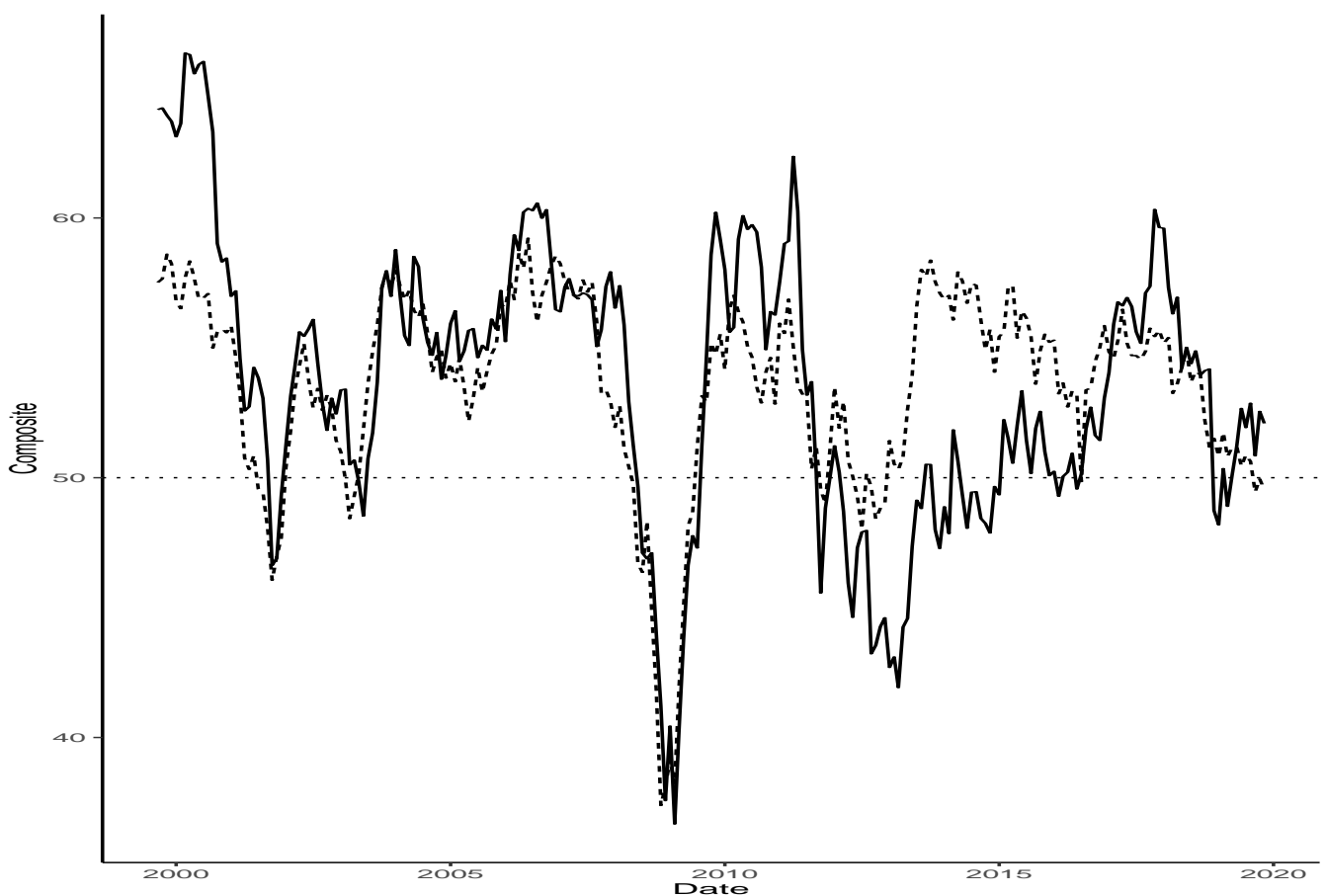


FIGURE 2 – PMI réel et synthétique de la France de 1999-09-01 à 2019-12-01 avec 4 pays donateurs

La figure 2 permet d'observer la série du PMI réel de la France et la série que l'on a construit à l'aide de notre groupe donateur de 4 pays. Nous rappelons que nous avons seulement choisis ces 4 pays car ils étaient les seuls à avoir une profondeur de données suffisante pour construire la synthèse jusqu'à septembre 1999. La racine de l'erreur quadratique moyenne de notre synthèse est de 3,95.

Maintenant pour améliorer la précision de synthèse (on entend par précision de la synthèse un écart faible entre les valeurs réelles et les valeurs synthétiques du PMI), nous pouvons

augmenter le nombre de pays dans notre groupe donateur. De fait, en l'absence de profondeur de données suffisante nous raccourcirons la période de synthétisation afin qu'elle débute en novembre 2009.

TABLE 2 – Contribution des pays donateurs (Figure 3)

	Pondérations
Allemagne	0.089
Italie	0.010
Espagne	0.825
Angleterre	0.010
US	0.022
Chine	0.008
Inde	0.008
Brésil	0.006
Japon	0.007
Russie	0.016

Le tableau permet d'observer la contribution des pays donateur pour la synthétisation du PMI de la France de 2009-11-01 à 2019-12-01. On peut voir que les pays que nous avons ajoutés ont une contribution relativement faible comparativement aux pays que nous avons dans notre groupe donateur, l'Espagne a une contribution nettement supérieure au reste des pays.

La figure 3 permet donc d'observer la série réelle du PMI de la France et la série synthétique construite à partir d'un groupe donateur de 10 pays sur une période allant de novembre 2009 à décembre 2019.

La racine de l'erreur quadratique moyenne de cette nouvelle synthétisation est de 2,48 ce qui permet de confirmer l'intuition selon laquelle un plus grand nombre de pays donateur permet d'améliorer la précision de notre synthétisation.

La figure 4 permet de visualiser l'évolution de l'écart entre les valeurs du PMI de la France réelle et les valeurs du PMI de la France synthétique que nous avons construit à l'aide d'un groupe donateur de 10 pays.

## 5.2 Synthétisation des États-Unis à partir du PMI composite

Désormais nous allons prolonger synthétiquement une série de données que l'on a. Pour ce faire il faut avoir une partie de la série que l'on veut prolonger (on ne peut pas la créer de toute pièce) et les séries de notre groupe donateur<sup>8</sup> sur toute la période que nous souhaitons synthétiser. Prenons l'exemple des États-Unis. Dans notre base de données, nous avons les valeurs du PMI des États-Unis seulement depuis novembre 2009.

Notre groupe donateurs comprend cinq pays. Le tableau 3 des pondérations montre une très grande contribution de l'Angleterre dans la construction de notre série synthétique.

8. France, Allemagne, Italie, Espagne, Angleterre



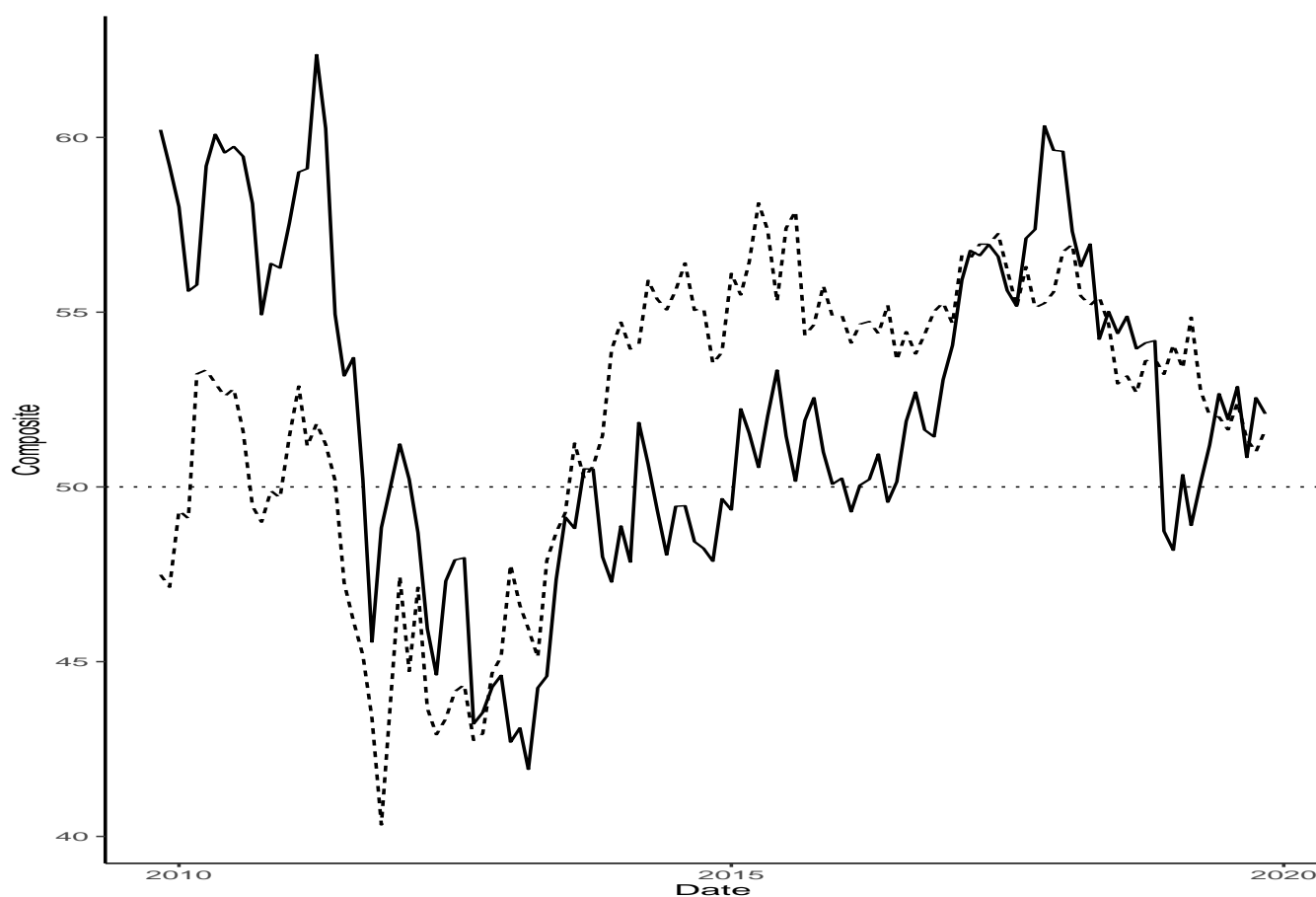


FIGURE 3 – PMI réel et synthétique de la France de 2009-11-01 à 2019-12-01 avec 10 pays donateurs

TABLE 3 – Contribution des pays donateurs

	Pondérations
France	0.00004
Allemagne	0.001
Italie	0.00003
Espagne	0.00004
Angleterre	0.999

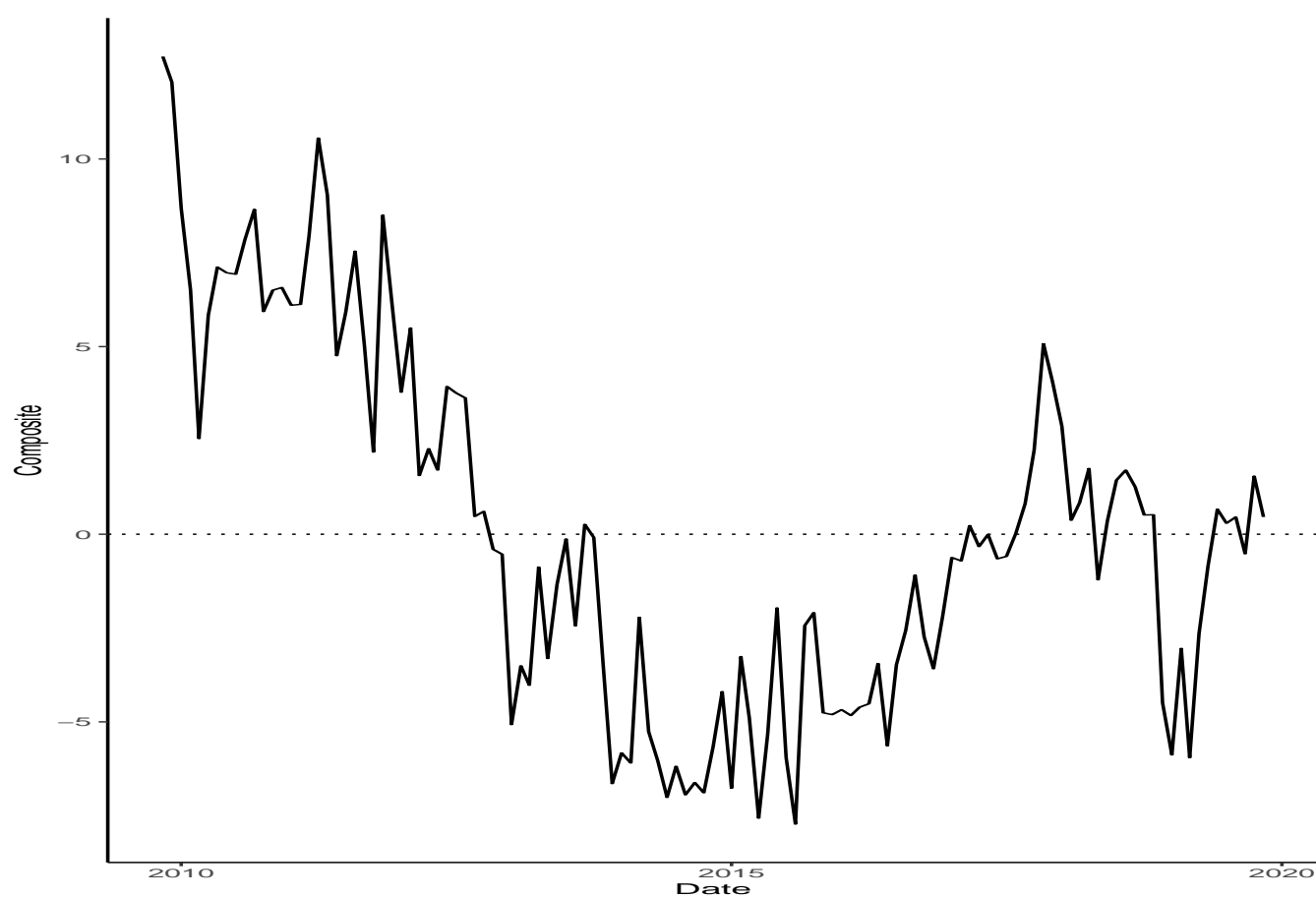


FIGURE 4 – Évolution de l'écart entre les PMI de la France réelle et de celle synthétique

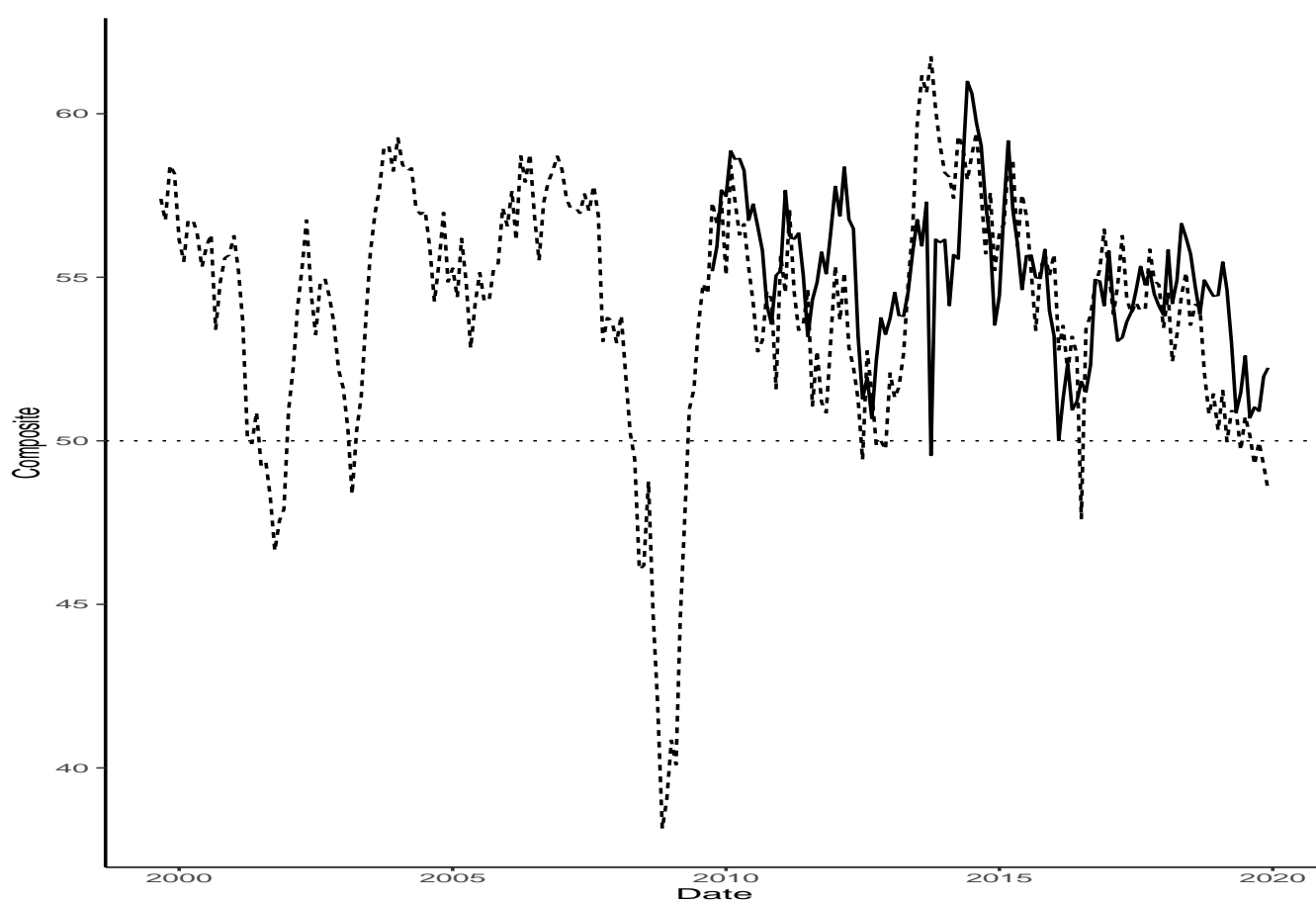


FIGURE 5 – PMI réel et synthétisé des US de 1999-09-01 à 2019-12-01

La figure 5 permet de visualiser ce à quoi la série des PMI des États-Unis aurait ressemblé si nous avions pu récolter les PMI avant la période de novembre 2009. Cette synthétisation a donc été construite en appliquant les pondérations que nous avons dans la table 3 aux pays donateurs sur toute la période allant de novembre 1999 à novembre 2019. La racine de l'erreur quadratique moyenne de cette synthétisation est de 2,54. Un niveau de synthétisation comparable à celui que nous avons constaté lors de la synthétisation du PMI de la France avec 10 pays donateurs.

### 5.3 Estimation du PIB à l'aide du PMI

#### 5.3.1 Estimation du PIB français avec le PMI réel

Maintenant que nous sommes en mesure de synthétiser une série de PMI pour n'importe quel pays à condition d'avoir un groupe de pays donateurs sur la période que l'on cherche à synthétiser, nous allons vérifier si cette série synthétique que nous avons créée peut permettre d'estimer le PIB avec précision.

Comme nous l'avons vu précédemment le PIB est donc estimé chaque trimestre à l'aide des trois PMI mensuels<sup>9</sup> de ce même trimestre. Au moyen d'une régression, décrite dans la partie méthodologie, nous sommes en mesure de trouver les coefficients de chaque PMI mensuels sur la valeur du PIB trimestriel. Pour faire ce test nous allons prendre les données PMI de la France que nous avons vu précédemment ainsi que les PIB trimestriels de la France disponibles sur le site de l'INSEE. Nous allons donc vérifier la précision de l'estimation du PIB avec la série de PMI réels de la France, puis nous estimerons ensuite le PIB de la France avec la série de PMI synthétique que nous avons construit.

Le tableau 4 est le résultat de notre régression du PIB trimestriel de la France avec le PMI du premier, deuxième et troisième mois de chaque trimestre. Il nous permet d'observer les coefficients que l'on utilise pour créer notre estimation du PIB avec les PMI réels. Ensuite nous construisons notre estimation avec l'équation de droite suivante :

$$\text{deltaPIBestim} = (-0,049) * \text{PMI\_M1} + (0,078) * \text{PMI\_M2} + (0,034) * \text{PMI\_M3} - 2,997$$

Ainsi nous pouvons tracer notre PIB estimé que nous retrouvons en pointillé sur la figure 6. Afin de mieux visualiser la proximité entre notre PIB estimé et le PIB réel nous tracerons aussi en trait plein le PIB réel.

Comme nous pouvons le constater l'estimation est relativement fiable. La figure 7 permet de visualiser la différence entre le PIB réel et le PIB estimé à l'aide des vraies valeurs du PMI. L'estimation a une racine carrée de l'erreur quadratique de 0,311<sup>10</sup>.

---

9. PMI\_M1, PMI\_M2, PMI\_M3

10.  $\sqrt{(\text{PIBreel} - \text{PIBestime})^2 / 84}$

TABLE 4 – Coefficients de la régression du PIB avec les PMI réels

	<i>Dependent variable :</i>
	deltaPibTrimInsee
PMI_M1	−0.049** (0.024)
PMI_M2	0.078** (0.033)
PMI_M3	0.034* (0.019)
Constant	−2.997*** (0.361)
Observations	84
R <sup>2</sup>	0.571
Adjusted R <sup>2</sup>	0.555
Residual Std. Error	0.318 (df = 80)
F Statistic	35.454*** (df = 3 ; 80)
<i>Note :</i>	*p<0.1 ; **p<0.05 ; ***p<0.01

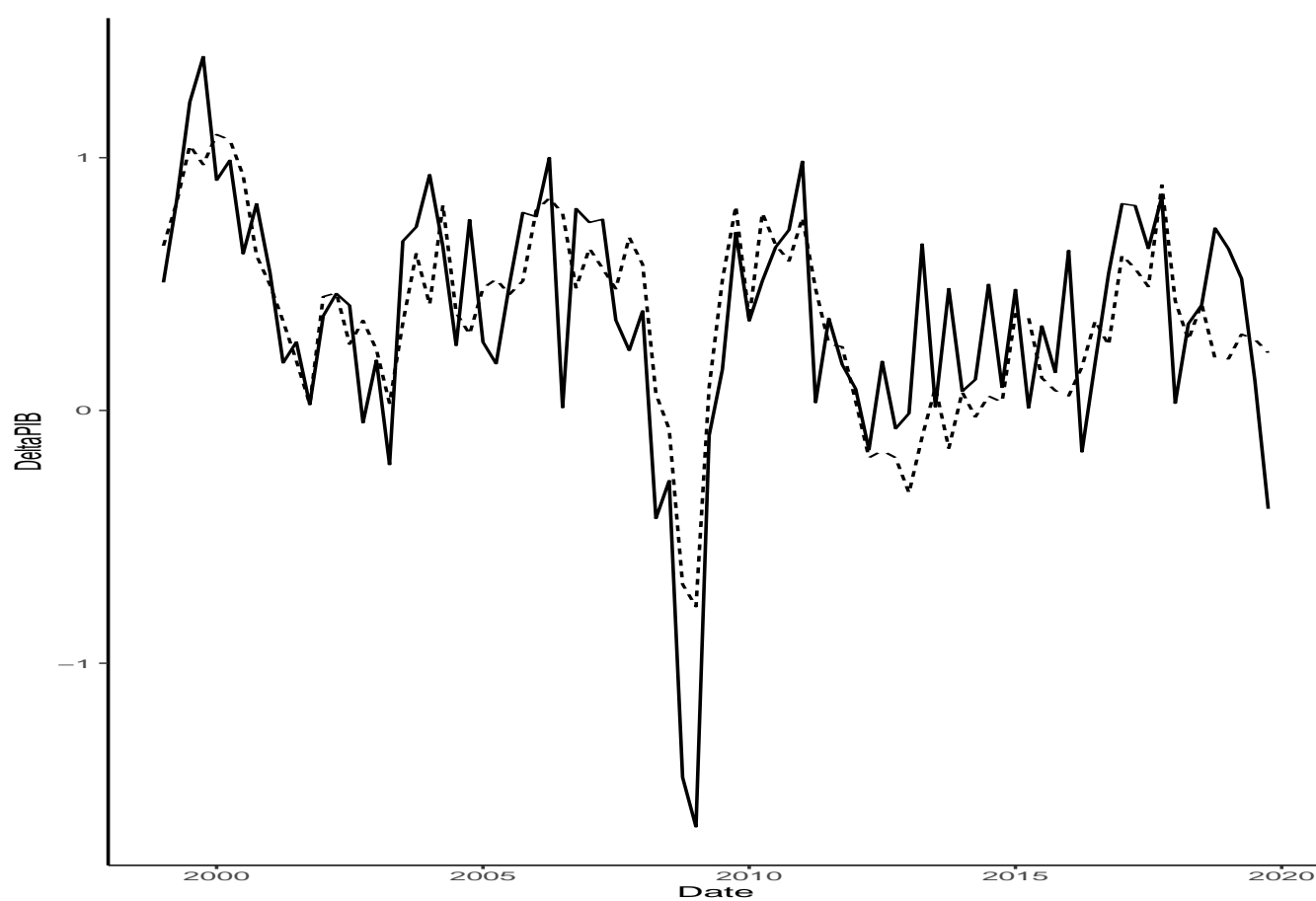


FIGURE 6 – PIB réel et estimé grâce au PMI réel de la France

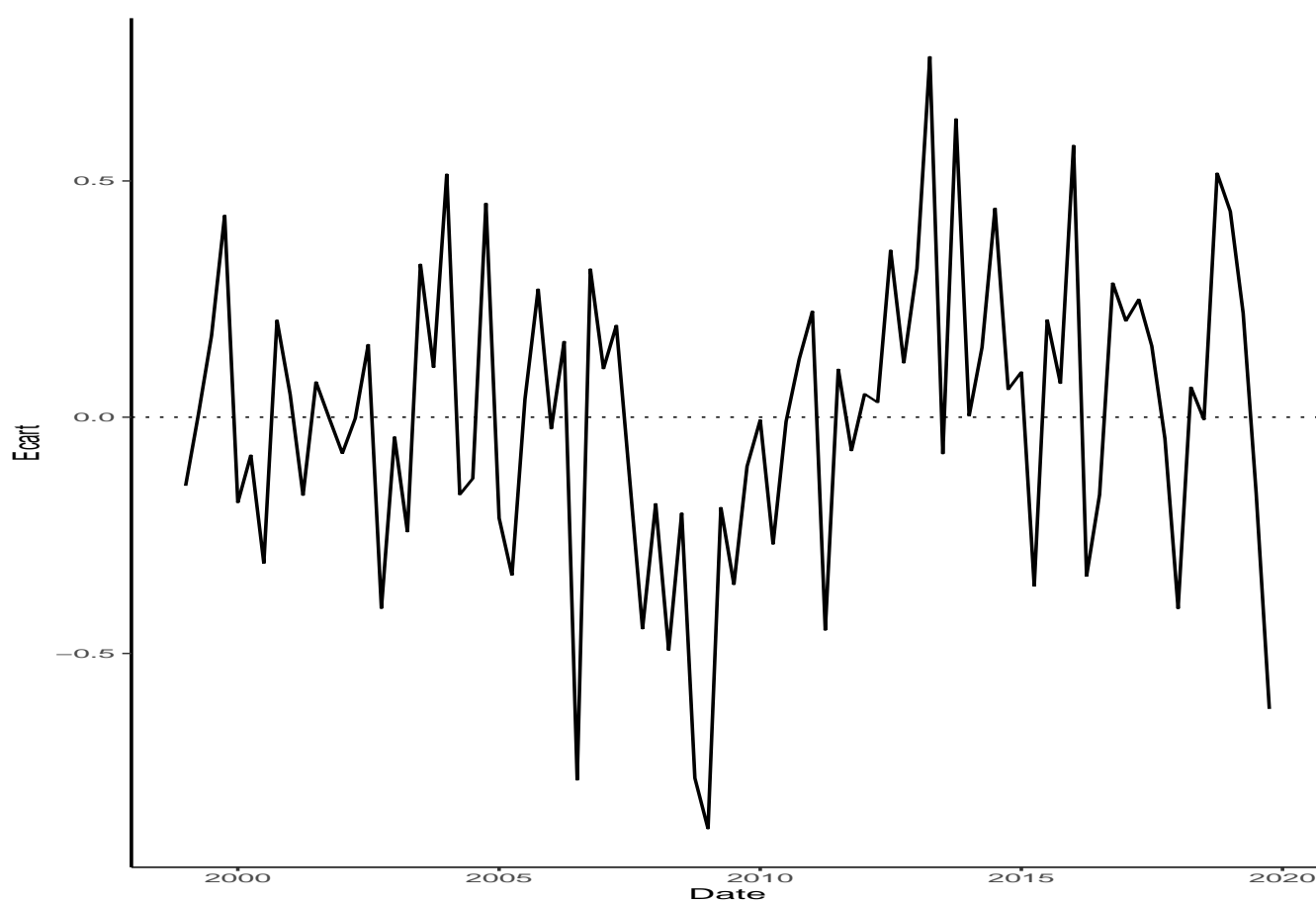


FIGURE 7 – Différence entre le PIB réel et le PIB estimé grâce au PMI réel

### 5.3.2 Estimation du PIB français avec le PMI synthétique

Désormais nous allons estimer le PIB avec la série synthétique que nous avons construit à l'aide du groupe de 4 pays donateurs <sup>11</sup>. Comme pour l'estimation avec le PMI réel nous allons régresser le PIB trimestriel avec les premiers, deuxièmes et troisièmes mois de chaque trimestre de notre série synthétique.

TABLE 5 –

<i>Dependent variable :</i>	
deltaPibTrimInsee	
PMI_M1synth	–0.007 (0.025)
PMI_M2synth	0.004 (0.040)
PMI_M3synth	0.097*** (0.029)
Constant	–4.700*** (0.512)
Observations	81
R <sup>2</sup>	0.587
Adjusted R <sup>2</sup>	0.571
Residual Std. Error	0.310 (df = 77)
F Statistic	36.474*** (df = 3 ; 77)

*Note :* \*p<0.1 ; \*\*p<0.05 ; \*\*\*p<0.01

Comme pour l'estimation faite sur la base des PMI réels, la régression nous permet d'obtenir les coefficients qui seront nécessaires à la construction de notre estimation. Le tableau 5 permet de les visualiser. Nous construisons donc notre nouvelle prévision avec l'équation de droite suivante :

$$\text{deltaPIBestim} = (-0,007) * \text{PMI\_M1synth} + (0,004) * \text{PMI\_M2synth} + (0,097) * \text{PMI\_M3synth} - 4,700$$

La figure 8 permet de visualiser notre prévision construite sur la base de notre série de PMI synthétique ainsi que le PIB réel. Elle semble elle aussi relativement fiable.

La figure 9 confirme cette fiabilité montrant très peu d'écart entre notre estimation et notre PIB réel. La racine de notre erreur quadratique moyenne est de 0,30. Autrement dit l'estimation que nous avons faite du PIB sur la base de notre série synthétique du PMI est plus précise que celle que nous avons pu faire sur la base de notre série de PMI réelle.

11. Voir Figure 2



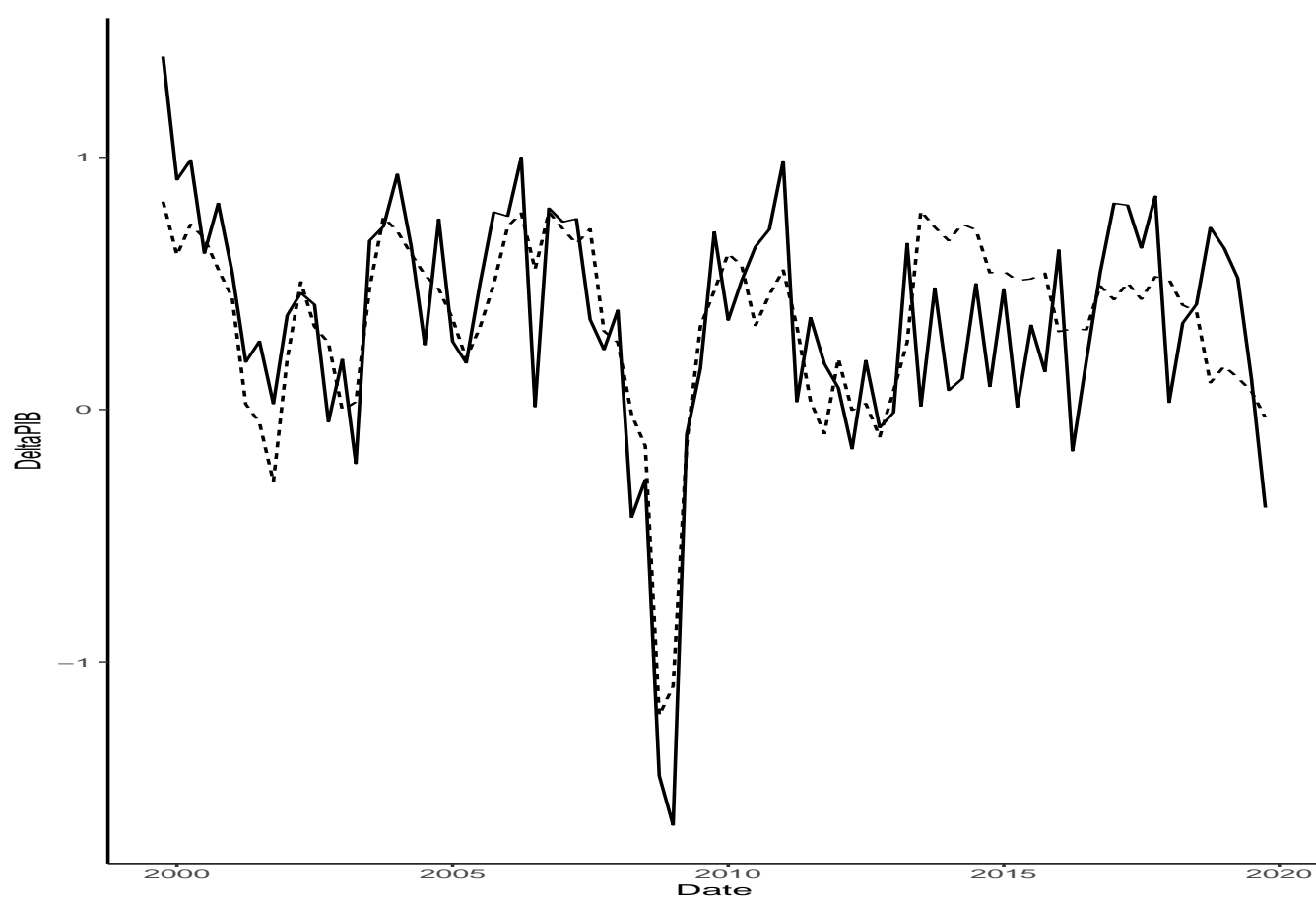


FIGURE 8 – PIB réel et estimé grâce au PMI synthétique de la France

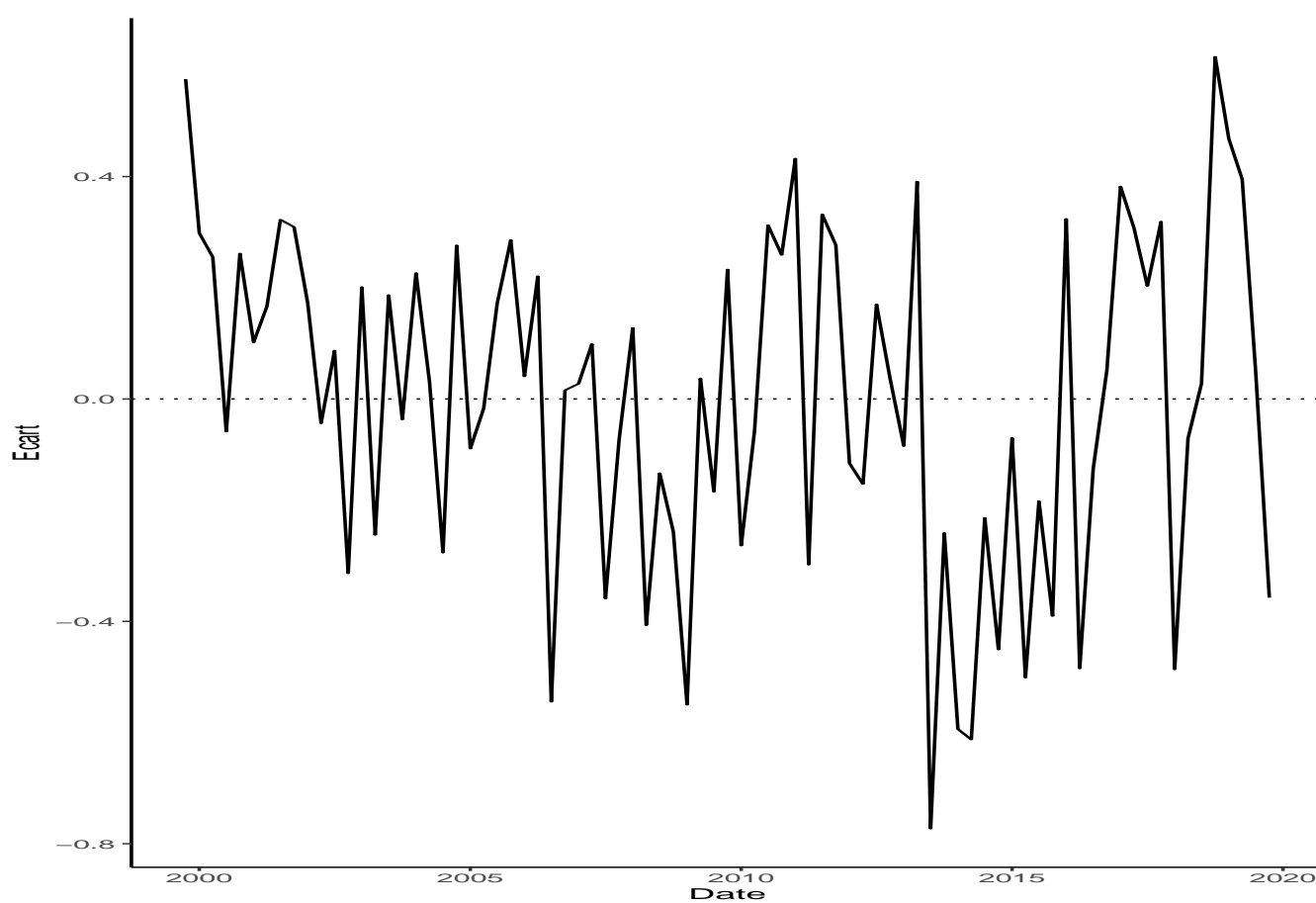


FIGURE 9 – Différence entre le PIB réel et le PIB estimé grâce au PMI synthétique

### 5.3.3 Estimation du PIB américain avec le PMI réel

Nous avons obtenu de très bons résultats en estimant le PIB Français avec notre série synthétique du PMI, voyons maintenant si cela peut aussi fonctionner lorsque l'on reprend une série synthétique construite avec des données de PMI manquantes. Comme nous l'avons vu dans la partie synthétisation des résultats, nous sommes parvenus à recréer une série synthétique du PMI des Etats-Unis juste à partir d'un groupe donateur. Reprenons donc cette série et voyons si elle permet d'estimer le PIB avec précision.

TABLE 6 – Coefficients de l'estimation du PIB américain avec le PMI réel

	<i>Dependent variable :</i>
	Delta_PIB
PMI_M1	−0.276 (0.246)
PMI_M2	0.428 (0.367)
PMI_M3	−0.206 (0.300)
Constant	6.546 (10.355)
Observations	41
R <sup>2</sup>	0.044
Adjusted R <sup>2</sup>	−0.033
Residual Std. Error	2.400 (df = 37)
F Statistic	0.572 (df = 3 ; 37)

*Note :* \*p<0.1 ; \*\*p<0.05 ; \*\*\*p<0.01

Nous reprenons la méthode d'estimation que nous avons évoquée dans la partie Estimation du PIB français et nous obtenons les coefficients visibles dans le tableau 6. Ensuite nous construisons notre estimation que l'on retrouve en pointillé dans la figure 10 aux côtés du PIB réel des États-Unis sur la période allant du 2009/10/01 au 2019/12/01.

L'estimation ne semble pas parvenir à capter toute l'amplitude des variations. La racine de l'erreur quadratique moyenne est de 2,280. Ce qui n'est pas très performant. Ceci est probablement dû à la courte période de PMI dont nous disposons pour effectuer notre estimation.

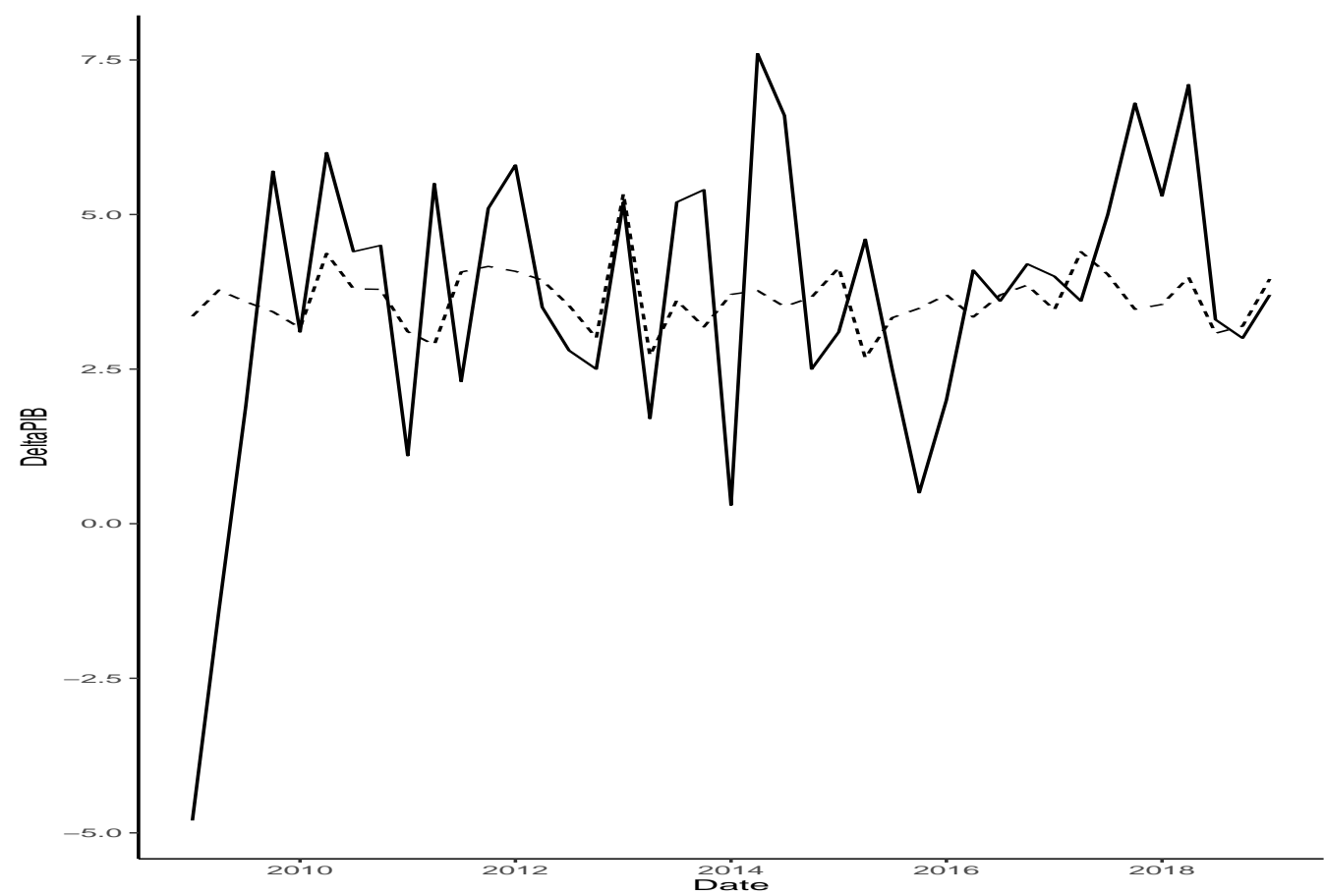


FIGURE 10 – PIB réel et estimé avec le PMI réel des États-Unis

### 5.3.4 Estimation du PIB américain avec le PMI synthétique

Voyons maintenant quelle est la précision de notre estimation lorsque l'on utilise la série synthétique du PMI. Comme précédemment nous effectuons notre régression qui permet d'obtenir le tableau 7. A partir de ce tableau nous construisons la figure 11 qui permet de visualiser notre estimation du PIB sur base de PMI synthétique en pointillé ainsi que nos vraies valeurs du PIB trimestriel. La racine de l'erreur quadratique moyenne de notre estimation est de 2,142. Cela n'est pas une estimation suffisamment précise, en revanche elle est plus précise que celle que nous avons effectuée avec nos PMI disponibles réels. Ce résultat confirme donc le potentiel de cette méthode.

TABLE 7 – Coefficients de l'estimation du PIB américain avec le PMI synthétique

	<i>Dependent variable :</i>
	Delta_PIB
PMI_M1synth	0.100 (0.164)
PMI_M2synth	0.341 (0.205)
PMI_M3synth	0.019 (0.172)
Constant	−20.774*** (3.648)
Observations	81
R <sup>2</sup>	0.387
Adjusted R <sup>2</sup>	0.363
Residual Std. Error	2.198 (df = 77)
F Statistic	16.199*** (df = 3 ; 77)

Note : \*p<0.1 ; \*\*p<0.05 ; \*\*\*p<0.01

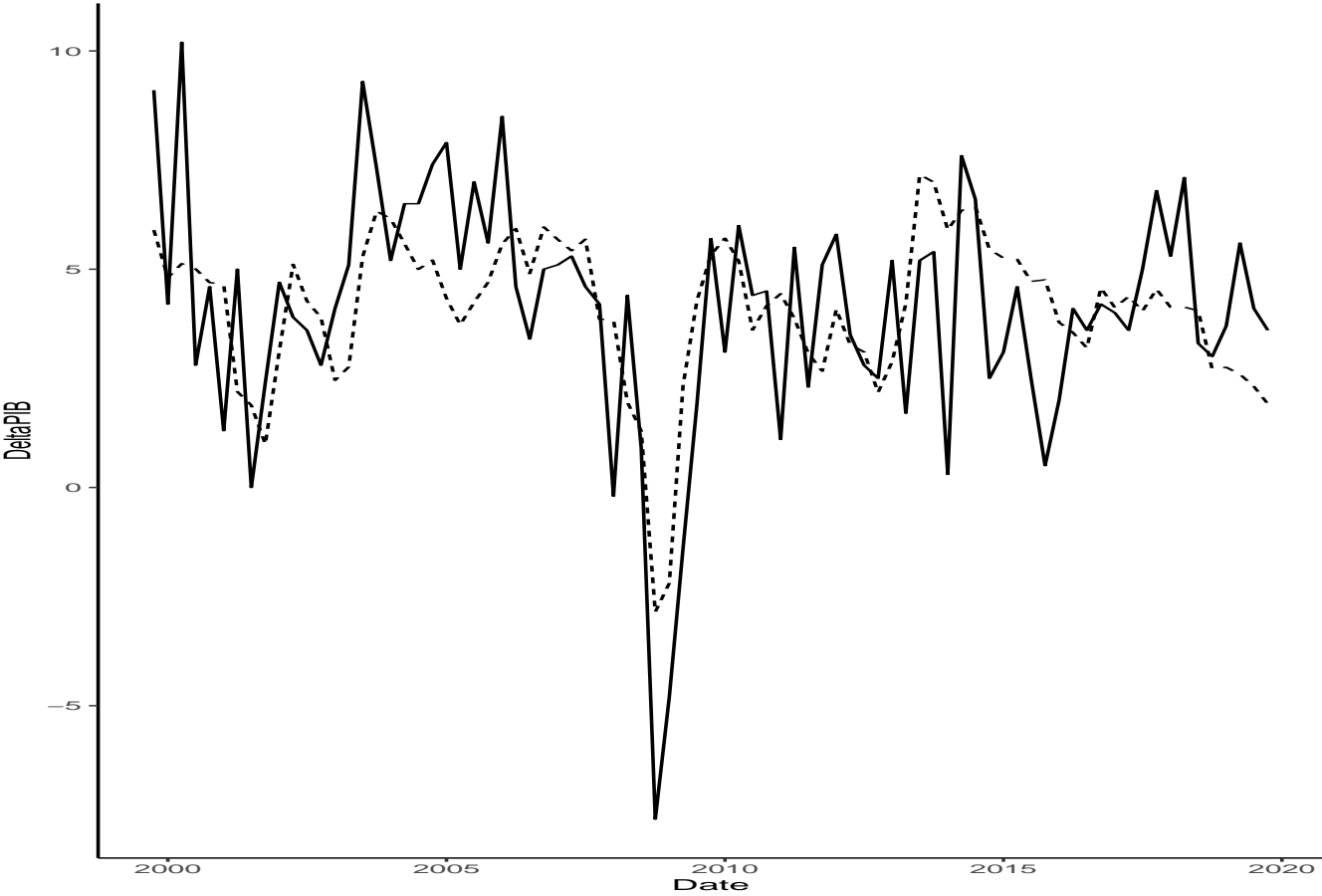


FIGURE 11 – PIB réel et estimé avec le PMI synthétique des États-Unis

## 6 Conclusion

L'intuition à l'origine de ce mémoire selon laquelle nous pouvions estimer le PIB sur la base d'un PMI synthétique a donc été confirmée. La précision de l'estimation du PIB français est importante. La précision de l'estimation du PIB américain est moins précise cependant elle est plus précise avec notre série de PMI synthétique qu'avec notre série de PMI réelle. Nous rappelons que ces résultats ont été trouvés dans ces deux seuls cas particulier. La méthode que nous avons appliquée n'est donc peut-être pas si précise dans tous les cas de figures, et peut potentiellement avoir fonctionné grâce à des biais que nous ne pouvons pas vérifier sans tests. Néanmoins, c'est un résultat encourageant qui montre qu'il serait peut-être possible d'étendre cette méthodologie à d'autres données avec d'autres groupes donateurs sur des périodes différentes pour vérifier sa robustesse. De plus dans le cadre de ce mémoire nous avons construit nos séries synthétiques sur la seule base du PMI composite, alors que comme nous l'avons explicité <sup>12</sup>, les bases qui regroupent les PMI de chaque pays comme la base IHS Markit par exemple donnent accès à plusieurs PMI par secteur. Il serait donc intéressant de rajouter des PMI de certains secteurs particuliers pour essayer d'améliorer encore notre synthétisation.

---

12. Voir partie Base de données

## Références

- ABADIE, A. (2021). Using Synthetic Controls : Feasibility, Data Requirements, and Methodological Aspects.
- ABADIE, A. & GARDEAZABAL, J. (2003). The Economic Costs of Conflict : A Case Study of the Basque Country.
- ACEMOGLU, JOHNSON, S., KERMANI, A., KWAK, J. & MITTON, T. (2016). The Value of Connections in Turbulent Times : Evidence from the United States.
- ALLEGRETO, S., ARINDRAJIG, D., ZIPPERER, B. & REICH, M. (2017). Credible Research Designs for Minimum Wage Studies : A Response to Neumark, Salas, and Wascher.
- ATHEY, S. & IMBENS, G. W. (2017). The State of Applied Econometrics : Causality and Policy Evaluation.
- BOHN, SARAH, LOFSTROM, M. & RAPHAEL, S. (2014). Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population ?
- BORJAS, G. J. (2017). The Wage Impact of the Marielitos : A Reappraisal.
- CUNNINGHAM, SCOTT & SHAH, M. (2018). Decriminalizing Indoor Prostitution : Implications for Sexual Violence and Public Health.
- DONOHUE, J. J., ABHAY, A. & WEBER, K. D. (2018). Right-to-Carry Laws and Violent Crime : A Comprehensive Assesment Using Panel Data and a State-Level Synthetic Control Analysis.
- FONTANEL, J. & GUILHAUDIS, J.-F. (2019). Les Effets « Pervers » de l'usage Du PIB Pour La Décision Politique et Les Relations Internationales. Comment En Sortir ?
- GODBOUT, C. & JACOB, J. (2010). Le pouvoir de prévision des indices PMI.
- GUTIERREZ, I. A., WEINBERGER, G. & ENGBERG, J. (2016). Improving Teaching Effectiveness : Impact on Student Outcomes : The Intensive Partnerships for Effective Teaching Through 2013–2014.
- HEERSINK, B., BRENTON, P. D. & JENKINS, J. A. (2017). Disasters and Elections : Estimating the Net Effect of Damage and Relief in Historical Perspective.
- INSEE (Éd.). (2017). La Croissance Économique.
- JANY-CATRICE, F. & MÉDA, D. (2011). Femmes et Richesse - Au Delà Du PIB.
- JURUS, A. (2013). Estimation Du PIB Potentiel et Impact Sur Les Finances Publiques.
- LARCH, M. (2014). Stability and Growth Pact of the European Union.
- MARKIT, I. (Éd.). (2022). PMI - L'indicateur Économique Le plus Suivi Au Monde.
- NEUMARK, D. & WASHER, W. (2017). Reply to Credible Research Designs for Minimum Wage Studies.
- PIETERS, H., CURZI, D., OLPER, A. & SWINNEN, J. (2017). Effect of Democratic Reforms on Child Mortality : A Synthetic Control Analysis.
- PINOTTI, P. (2015). The Economic Costs of Organised Crime : Evidence from Southern Italy.
- STIGLITZ, J. & SEN, A. (2009). Rapport de La Commission Sur La Mesure Des Performances Économiques et Du Progrès Social.