

Une année de crise COVID : impact sur la dynamique de l'activité des entreprises en France. Une évaluation sur données individuelles*

Benjamin Bureau[†], Anne Duquerroy[†], Julien Giorgi[‡],
Mathias Lé[†], Suzanne Scott[‡] et Frédéric Vinas[†]

Juillet 2021, WP #823

RÉSUMÉ

Nous étudions l'impact de la crise sanitaire sur l'activité de plus de 645 000 entreprises, à partir de données individuelles permettant d'estimer leur chiffre d'affaires à une fréquence mensuelle. Notre approche, fondée sur un modèle de micro-simulation, est innovante à triple titre. Premièrement, nous quantifions la perte d'activité par rapport à une situation contrefactuelle dans laquelle la crise n'aurait pas eu lieu. Deuxièmement, nous estimons ce choc au niveau individuel, permettant une analyse détaillée de l'hétérogénéité des chocs d'activité. Nous mettons en lumière la dispersion du choc à la fois entre secteurs et au sein des secteurs. Nous montrons que le secteur de l'entreprise explique jusqu'à 48% de la variance des chocs d'activité mensuels pondérés par l'emploi en 2020, soit une part beaucoup plus importante que lors d'une année normale. Enfin, nous identifions quatre profils de trajectoires, caractéristiques de l'évolution de l'activité en 2020. Le secteur est le principal déterminant d'appartenance à un profil donné. Conditionnellement au secteur, le profil de trajectoire est également corrélé à la capacité d'adaptation organisationnelle des entreprises.

Mots-clés : Covid-19 ; activité économique ; sociétés non-financières ; micro-simulation.

Classification JEL : D22 ; G38 ; H32

* Nous remercions Jean Boissinot, Nicolas Carnot, Gilbert Cette, Olivier Garnier, Alain Gerbier, Dominique Goux, Loriane Py, Sébastien Roux et Jean-Pierre Villette pour les nombreux échanges et conseils qui ont nourri cette analyse présentée ici. Nous remercions également Tanguy Aubert, Julien Demuynck, Olivier Hubert, Martial Ranvier ainsi que les participants aux réunions techniques du Comité de suivi et d'évaluation des mesures de soutien financier aux entreprises confrontées à l'épidémie de Covid-19, présidé par Benoit Cœuré.

[†] Banque de France, benjamin.bureau@banque-france.fr, anne.duquerroy@banque-france.fr, mathias.le@banque-france.fr, frederic.vinas@banque-france.fr

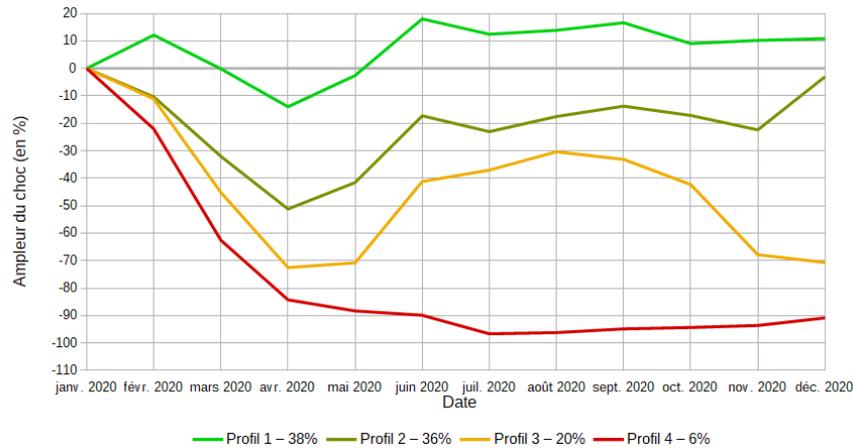
[‡] Insee, julien.giorgi@insee.fr, suzanne.scott@insee.fr

NON-TECHNICAL SUMMARY

Au début de l'année 2020, la pandémie de COVID-19 et les mesures de restriction ayant visé à l'endiguer ont provoqué un choc important sur l'activité des entreprises françaises. En s'appuyant sur un échantillon inédit de plus de 645 000 entreprises permettant d'estimer leur chiffre d'affaires à une fréquence mensuelle à partir des déclarations de TVA, nous proposons une analyse triplement innovante dans le débat concernant l'impact économique de la crise sanitaire. D'abord, nous évaluons l'impact de la crise en estimant des chocs d'activité qui mesurent l'évolution de l'activité économique mesurée en écart par rapport à une situation contrefactuelle simulée. Puis, le modèle de micro-simulation développé permet d'évaluer le choc d'activité lié à la crise au niveau individuel, rendant possible une analyse détaillée de l'hétérogénéité des situations des entreprises après un an de crise, y compris au sein d'un même secteur. Enfin, l'analyse exploite la dimension infra-annuelle des données pour caractériser différents profils de trajectoires d'activité au cours de l'année 2020.

Nos résultats montrent que le premier confinement a constitué un choc généralisé dont l'ampleur a varié selon les secteurs alors que le second confinement a été plus faible et plus localisé. Au cours de la crise, le secteur d'activité explique une part nettement plus importante de la dispersion des chocs individuels d'activité qu'en période normale. Elle explique ainsi jusqu'à 48 % de la variance des chocs observés lorsque ces derniers sont pondérés par les effectifs salariés. L'analyse dynamique des pertes d'activité individuelles révèle quatre profils de trajectoire, illustrant les différents niveaux de résilience des entreprises face à la crise. L'appartenance à ces différents profils s'explique principalement par le secteur d'activité – à un niveau de découpage très fin. A secteur d'activité similaire, elle s'explique par les capacités d'adaptations organisationnelles et productives des entreprises pendant la crise. L'existence d'une activité exportatrice avant le début de la crise est très défavorable à la trajectoire d'activité connue par les entreprises en 2020.

Chocs moyens par profil de trajectoire



Notes : La classification des trajectoires est réalisée par clustering de courbe qui permet d'identifier quatre profils-types représentatifs des trajectoires de chocs des entreprises pendant l'année 2020. Lecture : les entreprises du profil 1 ont connu un choc moyen de -14% en avril 2020. Sources : DGFIP-Insee. Calculs des auteurs.

Corporate activity in France amid the Covid-19 crisis. A granular data analysis.

ABSTRACT

Taking advantage of detailed firm-level data on VAT returns, we estimate the monthly impact of the Covid-19 crisis on the turnover of more than 645,000 French firms. Our approach, based on a micro-simulation model, is innovative in a triple way. Firstly, we quantify the activity loss with respect to a counterfactual situation in which the crisis would not have hit. Secondly, we estimate this shock at the firm level, enabling a thorough analysis of activity loss heterogeneity throughout the crisis. In particular, we shed light on the dispersion of the shock both within and between industries. We show that the industry the firm operates in explains up to 48% of the monthly activity shocks' variance weighted by employment, a much larger share than in a normal year. Finally, we leverage our monthly firm-level data on sales to show how corporate activity has evolved along four distinct trajectories throughout 2020. The main determinant of belonging to a given profile of activity is the firm industry – defined at a very granular level. Conditional on industry, the activity trajectory is also correlated with the ability to adapt some firms have demonstrated during the crisis in terms of organization and production.

Keywords: Covid-19 ; business dynamics ; micro-simulation ; non-financial corporations.

Working Papers reflect the opinions of the authors and do not necessarily express the views of the Banque de France or Insee. This document is available on publications.banque-france.fr/en

Introduction

La pandémie de COVID-19 a constitué un choc macro-économique majeur sur l'activité économique des entreprises françaises pouvant masquer une hétérogénéité individuelle ou sectorielle forte

Au début de l'année 2020, la pandémie de COVID-19 a provoqué un choc majeur sur l'activité économique mondiale et les mesures de restrictions sanitaires mises en œuvre par la plupart des gouvernements ont provoqué un ralentissement brutal de l'activité économique. En France, ces mesures se sont concrétisées par plusieurs périodes de confinement puis de déconfinement progressif, ainsi que par un ensemble de mesures évolutives comme les couvre-feux et les restrictions de déplacement. *In fine*, en France, le PIB a connu une chute de 7,9 % entre 2019 et 2020 [Amoureux et al., 2021]. La contrepartie en valeur monétaire de ce recul en volume du PIB est une diminution de 6,3 % du revenu national, soit 150 milliards d'euros (Md€). Rapporté à la taille de la population, cette diminution correspond à un recul du revenu monétaire de 2 300 euros par habitant [Carnot, 2021]. Au mois d'avril 2020, la chute de la valeur ajoutée a dépassé les 30 %, plaçant la France parmi les pays les plus touchés de la zone euro, au niveau des pertes subies par l'Italie ou l'Espagne [Heyer et Timbeau, 2020]. Sur l'ensemble de l'année 2020, la valeur ajoutée des entreprises françaises a connu une baisse de 8,1 % et de 8,3 % pour les seules sociétés non-financières (SNF).

Ces évaluations, s'appuyant sur l'examen des taux de variations annuels des agrégats macroéconomiques, permettent d'établir un diagnostic sur la situation économique globale. Ces approches omettent toutefois que les situations économiques des entreprises sont hétérogènes, et que l'impact de la crise au niveau individuel peut être très varié, y compris au sein d'un même secteur d'activité. La pandémie de COVID-19 a affecté l'activité des entreprises françaises en provoquant un choc exogène et non anticipé sur l'offre et la demande de biens et services [Baldwin et Di Mauro, 2020], mais ces chocs n'ont pas été de même ampleur dans tous les secteurs [Brinca et al., 2020].

La production des entreprises a été inégalement affectée du fait de l'inégal impact des mesures de restriction sanitaires et de la structure des chaînes d'approvisionnement

L'activité économique des entreprises a d'abord été affectée par les facteurs exogènes liés à la crise les ayant contraintes à une diminution ou à une cessation de leur production. L'impact de ces restrictions a inégalement touché les différents secteurs. Les périodes de confinement ont par exemple contraint à la fermeture un certain nombre d'activités jugées non-essentiels¹. Les différents secteurs n'ont ainsi pas connu les mêmes contraintes dans l'exercice de leur activité du fait de l'inégale réparti-

¹ Secteur culturel, restauration, établissements sportifs couverts, etc.

tion des mesures de restriction sanitaire [Baleytes et *al.*, 2020] [Dauvin et Sampognaro, 2021]². En France, les 20 % des secteurs les plus contraints par la crise sanitaire, c'est-à-dire faisant face aux contraintes de production les plus fortes, ont représenté près de 75 % de la baisse du PIB sur l'année [Baleytes et *al.*, 2020]. Parmi ces secteurs se trouvent principalement les secteurs industriels, la construction, les transports, l'hébergement-restauration et les autres activités de services³. Les mesures de distanciation sociale ont par ailleurs pu engendrer des chocs de productivité négatifs pour de nombreuses firmes, liés à la réorganisation des équipes et des processus de production [Blanchard et *al.*, 2020]. Par ailleurs, la chute de l'activité de certaines économies, en particulier chinoise, a conduit à un ralentissement de certains secteurs ou entreprises spécifiques⁴ du fait de leur dépendance à d'autres économies pour leur approvisionnement [Baldwin et Tomiura, 2020]. Ainsi, les degrés d'ouverture et de spécialisation des pays peuvent avoir un impact sur les fluctuations observées de l'activité économique au niveau national, et la structure des chaînes d'approvisionnement peut jouer un rôle sur l'hétérogénéité sectorielle ou individuelle du choc attribuable à la crise [Brancati et *al.*, 2020].

Les inégalités observées dans la chute du revenu disponible, la baisse de la demande étrangère et du tourisme et l'altération des comportements d'épargne et de consommation conduisent à une baisse de la consommation plus prononcée dans certains secteurs et entreprises

Au-delà des contraintes de fermeture et des difficultés d'approvisionnement, la consommation de biens et services a été affectée par les confinements, réduisant les possibilités de consommer [Boone et *al.*, 2020]. Ainsi, si le revenu disponible des ménages a progressé de 1 % en valeur en 2020 malgré une baisse des heures travaillées [Jauneau et Vidalenc, 2021], cette légère croissance masque de grandes disparités et la consommation totale a diminué [Baleytes et *al.*, 2021]. Entre mars 2020 et mars 2021, un ménage sur quatre déclare avoir connu une baisse de ses revenus [Clerc et *al.*, 2021] et certaines franges de la population comme les artisans, les commerçants ou encore des jeunes ont été particulièrement touchées [Pouget et *al.*, 2021]. L'inégale vulnérabilité des individus à une baisse de leurs revenus a donc pu entraîner une baisse de la demande plus marquée sur les postes de consommation présentant une élasticité-revenu forte chez les individus dont le revenu a été le plus affecté par la crise.

La diminution de la demande étrangère – touchant les entreprises exportatrices –, et la baisse du tourisme [Škare et *al.*, 2021] provoquant le ralentissement de l'activité de l'ensemble des secteurs tirés par celui-ci, comme l'hébergement-restauration participe également de l'inégal impact de la baisse de la

2 Au niveau international, le rôle direct de l'intensité des mesures de restriction sur l'activité est illustré par la forte corrélation entre l'indice de restriction de l'université de Oxford, synthétisant en temps réel le degré de restriction liés aux mesures sanitaires au niveau national [Hale et *al.* 2020], et le taux de (dé)croissance du PIB au premier trimestre de l'année 2020.

3 Les « Autres services » sont essentiellement composés des arts, spectacles et activités récréatives, des services de coiffure et de soins corporel et des réparations d'ordinateurs et autres biens personnels.

4 En particulier le secteur automobile

consommation, en particulier entre secteurs. La dégradation inédite des anticipations du fait de la crise [Insee, 2020a] peut également être un facteur clef conduisant à une augmentation de l'épargne de précaution des ménages et à un recentrage de leur consommation sur les biens de première nécessité [Bonnet et al, 2021].

Une compréhension fine de l'impact de la crise ne peut faire l'économie d'une approche individuelle

L'ensemble de ces facteurs d'offre et de demande identifiés par la littérature met en lumière des effets complexes et non-homogènes de la pandémie de COVID-19 en fonction du secteur d'activité ou des caractéristiques des entreprises. Des études américaines sur données d'enquête établissent que la crise liée au COVID-19 distingue des entreprises « gagnantes » et « perdantes », en construisant une mesure de la réallocation de l'activité nette des variations macroéconomiques de cette dernière [Barreiro et al., 2020] et en évaluant la pérennité de cette réallocation de l'activité au sein de l'économie américaine [Barrero et al., 2021]. Toutefois, ces méthodes ne permettent que de construire des mesures macro-économiques de réallocation, sans pouvoir caractériser plus finement l'hétérogénéité des situations. Le premier enjeu analytique de l'évaluation de la crise sanitaire est donc de pouvoir estimer les effets liés à cette dernière au niveau des entreprises afin de compléter les diagnostics macroéconomiques. Cette étude s'attache à répondre à cet objectif en s'appuyant sur des données individuelles d'entreprises afin de rendre compte de l'hétérogénéité de l'impact de la crise sanitaire.

L'impact de la crise doit être évalué par rapport à un scénario contrefactuel hors-crise

D'un point de vue méthodologique, les approches en taux de variations, annuels ou trimestriels, traditionnellement utilisées en comptabilité nationale ne permettent pas de quantifier pleinement l'impact de la crise sur les évolutions de l'activité, et ce même lorsque ces taux de variations sont calculés sur données individuelles [Guerini et al., 2020] [Hadjibeyli et al., 2021]. En effet, l'évaluation de l'impact de la crise sanitaire sur le niveau d'activité des entreprises ou des secteurs impose de ré-insérer les montants d'activité observés dans la dynamique de développement de chaque entreprise et secteur afin de pouvoir quantifier un manque à gagner correspondant à l'écart entre l'activité observée et le niveau d'activité qui aurait dû être observé au niveau de l'entreprise ou du secteur si la crise n'avait pas eu lieu. C'est pourquoi les choix utilisés dans cette étude s'appuient sur la construction d'une activité contrefactuelle individuelle, simulant les évolutions de l'activité des entreprises sur l'année 2020 dans un scénario où la pandémie ne serait pas survenue.

L'analyse dynamique des trajectoires d'activité des entreprises permet de compléter les diagnostics en coupe des confinements ou de l'ensemble de l'année 2020

Ces résultats doivent permettre de réaliser des analyses en coupe sur l'hétérogénéité des effets de la crise sur les entreprises. Ces approches, qu'elles soient effectuées au mois le mois ou à l'issue de l'année 2020 ne permettent cependant pas de suivre les entreprises dans le temps et d'étudier si les entreprises subissant les plus fortes pertes un mois donné sont également les entreprises subissant les pertes d'activité les plus importantes le mois suivant. Une caractérisation des trajectoires individuelles des entreprises pendant la crise vient donc offrir un éclairage supplémentaire à l'analyse de l'impact de la crise sanitaire sur leur activité en informant sur l'hétérogénéité de la résilience des entreprises françaises, c'est-à-dire leur inégale capacité à faire face à un choc d'une telle ampleur et à se rétablir sur le sentier de croissance après un choc initial. La méthode exploite la diversité des chocs d'activité et des trajectoires observés pour dresser une typologie des profils et des situations après un an de crise. La caractérisation des déterminants des chocs d'activité permet de donner un éclairage sur les facteurs explicatifs de la situation des entreprises. En particulier, le rôle joué par le secteur, souvent utilisé comme unique variable de caractérisation, doit être discuté et quantifié. Conjointement au secteur d'activité, soumis à l'évolution des contraintes légales et administratives au fil de la crise, les stratégies d'adaptation des entreprises, que ce soit en termes d'évolution de leur offre de biens et services, de la réorganisation de leur chaîne de production, ou de la réalisation d'investissements spécifiques pour pallier les difficultés rencontrées peuvent être corrélées aux trajectoires observées. L'étude de ces corrélations doit venir documenter les difficultés d'activité spécifiquement liées à la crise.

Le reste de l'article se compose d'abord d'une présentation méthodologique de l'évaluation du niveau d'activité qu'aurait connu chaque entreprise en l'absence de crise et de la manière dont ces évaluations sont mobilisées pour quantifier les pertes d'activités attribuables à la crise et leur inégale répartition dans le tissu productif français (I). Une seconde partie développe les principaux résultats du modèle d'évaluation développé, en insistant sur l'hétérogénéité des trajectoires et situations individuelles après un an de crise ainsi que sur le rôle que joue le secteur dans cette hétérogénéité parmi un ensemble d'autres caractéristiques (II). Puis, la méthode et les résultats sont discutés afin de clarifier leur interprétation (III). Enfin, une conclusion suggère des pistes de développement et de prolongement de l'analyse (IV).

I. Données et construction de l'échantillon

1. Mobilisation des données de TVA

La mesure de l'activité retenue est celle du *chiffre d'affaires*. Ce choix est motivé par le fait que le chiffre d'affaires est une mesure brute de l'activité économique dont l'estimation n'est pas affectée par des hypothèses de reconstruction comme c'est le cas avec la valeur ajoutée. Il permet ainsi d'approcher l'impact de la crise sur l'activité, indépendamment des ajustement effectués ensuite dans la gestion financière et opérationnelle des entreprises. Les données mobilisées sont issues des déclarations mensuelles faites par les entreprises dans le cadre du paiement de la taxe sur la valeur ajoutée (TVA). Les entreprises détaillent dans celles-ci leurs ventes et achats chaque mois, trimestre ou année à la Direction Générale des Finances Publiques (DGFIP). Ces déclarations permettent de reconstituer le chiffre d'affaires de chaque entreprise, en sommant l'ensemble de ses opérations, imposables ou non, sur le territoire français ou à l'étranger (formules et dictionnaire des variables en [Annexe A](#)). Seules les entreprises effectuant une déclaration mensuelle sont retenues dans l'analyse.

2. Correction des données brutes

Les séries de chiffre d'affaires construites à partir des données de déclarations nécessitent certaines corrections afin de pouvoir obtenir des chroniques robustes. Des reports de déclaration d'un mois sur l'autre, se traduisant par une déclaration nulle un mois donné suivie d'une déclaration égale à l'activité de deux mois le mois suivant, ont été détectés et corrigés en répartissant l'activité du second mois entre le mois de déclaration nulle et le mois de rattrapage ([Annexe B.1](#)). Des valeurs aberrantes, en termes de niveau ou d'évolutions mensuelle et annuelle, ont également été corrigées en les ramenant dans la tendance de la série ([Annexe B.2](#)). Enfin, les entreprises déclarant leur chiffre d'affaires de manière irrégulière, et pour lesquelles des simulations robustes de leurs évolutions d'activité ne pouvaient être effectuées, ont été retirées de l'échantillon de travail ([Annexe B.3](#)). Cette restriction concerne surtout des micro-entreprises avec un faible chiffre d'affaires annuel et n'affecte donc que marginalement le taux de couverture de l'étude en termes de chiffre d'affaires⁵.

3. Cohérence avec les données de bilan et restriction du champ d'étude

Les déclarations de TVA ont été enrichies d'informations issues de Fare - Fichier Agrégé de Résultats d'Esane (Élaboration des statistiques annuelles d'entreprise) – 2018, date du dernier millésime

⁵ Cette restriction fait baisser le taux de couverture annuel de 0.2 points de pourcentage

disponible. L'échantillon est ainsi restreint aux entreprises présentes dans Fare 2018, déclarant leur TVA au régime mensuel depuis janvier 2018, ainsi qu'aux entreprises dont les données ne sont pas imputées dans Fare 2018. Cet appariement permet de disposer d'informations sur la date de création, la taille et l'existence ou non d'une activité exportatrice des entreprises. Il se justifie également par la perspective de la construction d'un tableau de flux individuel afin d'étudier l'impact de la crise sanitaire sur la trésorerie des entreprises et les conséquences qui en découlent en termes d'endettement. Ce travail sur la trésorerie des entreprises est effectué dans une seconde étude complémentaire à la présente [Bureau et al., 2021a] [Bureau et al., 2021b]. Enfin, le recours aux données de bilan permet de vérifier la cohérence des chiffres d'affaires reconstruits à partir des données de TVA.

Pour assurer la cohérence des sources, les entreprises dont le chiffre d'affaires issu de Fare diffère de plus de 35 % du chiffre d'affaires annuel reconstitué à partir des déclarations TVA en 2018 sont également exclues des données. Cette cohérence au seuil de 35 % a été vérifiée à la fois au niveau des unités légales et des groupes profilés. Lorsque la cohérence n'était pas vérifiée au niveau unité légale, mais l'était au niveau du groupe profilé, ce dernier était conservé dans l'échantillon en agrégeant les chiffres d'affaires des différentes unités légales le composant⁶. Cette condition de cohérence entre les sources statistiques conduit à éliminer de l'échantillon certaines grosses entreprises françaises pour lesquelles les déséquilibres entre les données de bilan et celles de déclarations TVA sont souvent prononcés.

Enfin, les entreprises appartenant aux secteurs financier, de l'enseignement et de l'administration publique, ainsi que les travailleurs indépendants et les personnes physiques ont été supprimées de l'échantillon.

4. Description de l'échantillon : taux de couverture et distribution par secteur d'activité et taille d'entreprise

L'échantillon final est constitué de plus de 740 000 unités légales. Ces unités légales correspondent à un peu plus de 645 000 unités d'observation : 578 000 unités légales analysées en tant que telles, et 68 000 groupes profilés composés d'au moins deux unités légales. Cet échantillon représente 85 % de la valeur ajoutée (VA) des sociétés non financières (SNF), hors travailleurs indépendants, des secteurs conservés dans l'étude. Sur l'ensemble des SNF, l'échantillon couvre 71 % de la VA et en particulier 81 % de la VA des entreprises de taille intermédiaire et des grandes entreprises (ETI-GE), 72 % de la VA des petites et moyennes entreprises (PME) et 38 % de la VA des très petites entreprises (TPE)⁷. Le faible part de la valeur ajoutée des TPE couverte par l'échantillon est lié au fait que les pe-

6 En comptabilité d'entreprises, les chiffres d'affaires des différentes unités légales ne se somment pas exactement (contrairement à la valeur ajoutée). La comparaison du CA profilé issu des comptes profilés de FARE avec le *proxy* de CA obtenu par somme des CA des unités légales calculés à partir des déclarations de TVA permet de conserver des entreprises dont la cohérence n'est pas vérifiée au niveau UL, mais dont une approximation au niveau profilé est cohérente avec les données de bilan. Cela permet notamment d'augmenter la taille de l'échantillon et le taux de couverture.

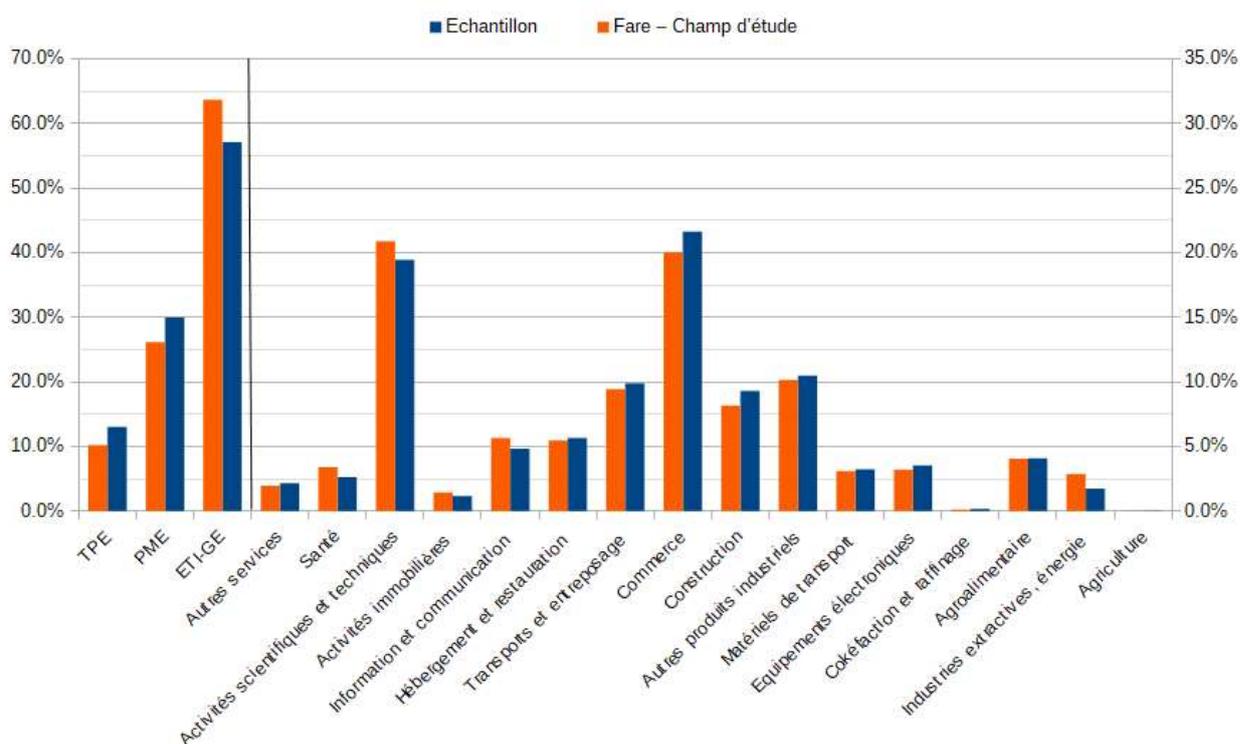
7 Les taux de couverture de l'échantillon sont obtenus en comparant les montants de valeur ajoutée de l'échantillon au montant de la valeur ajoutée de la population totale d'entreprises présentes dans FARE 2018 sur le même champ.

tites entreprises souscrivent majoritairement aux régimes de déclaration trimestriel et annuel de la TVA.

La répartition des effectifs salariés par secteur au sein de cet échantillon est proche de la ventilation sectorielle de l'ensemble des entreprises françaises du champ d'étude. Par rapport à la base Fare restreinte au champ de travail de l'étude, le secteur du commerce présente une légère surreprésentation, et les secteurs de l'énergie et des activités scientifiques et techniques une sous-représentation. Par taille d'entreprise, la structure des effectifs est également proche de la structure générale, mais les ETI-GE sont sous-représentées dans l'échantillon au profit des PME et TPE (**Figure 1**-Chiffres détaillés en **Annexe C**).

Les procédures de correction et de vérification de la cohérence des déclarations conduisant à l'échantillon final permettent donc de fournir une image peu déformée de la population des SNF françaises et de leur montant d'activité mensuelle.

Figure 1 – Ventilation des effectifs salariés par secteur d'activité et taille d'entreprises



Note de lecture : Les données issues de FARE comportent l'ensemble des entreprises françaises. Elles sont ici restreintes au champ d'étude de l'échantillon et se voient appliquer les mêmes restrictions sectorielles.

Source : Déclarations TVA (DGFIP) ; Données FARE 2018 (Insee) ; Calcul des auteurs.

II. Méthode

Le premier point auquel s'attache cette étude est d'estimer l'impact de la crise sur l'activité des entreprises françaises. La théorie économique permet en effet d'identifier un ensemble de facteurs liés à la crise ayant pu avoir un impact sur l'offre et de la demande de biens et services (cf. supra). Il reste que tous ces effets ne sont pas nécessairement présents ou de la même ampleur pour toutes les entreprises. Ainsi, l'évaluation de l'impact de la crise doit s'appuyer sur l'activité observée *in fine* au niveau individuel. Une estimation naïve de l'effet de la crise à travers le taux de croissance entre 2019 et 2020 permet déjà de fournir un ordre de grandeur aux pertes d'activité subies. Toutefois, que ce soit au niveau sectoriel ou de l'entreprise, les taux de croissance annuels ne constituent pas une estimation pleinement satisfaisante de l'effet de la crise. En effet, les entreprises, comme les secteurs auxquelles elles appartiennent, connaissent des trajectoires de croissance ou de décroissance auxquelles s'ajoutent d'éventuelles fluctuations saisonnières de l'activité. Un secteur dont l'activité diminue de manière structurelle année après année devrait donc connaître un taux de croissance négatif entre 2019 ou 2020, indépendamment de l'apparition de la crise sanitaire. L'utilisation d'un simple taux de croissance conduirait donc à surestimer l'impact négatif de la crise sur l'activité du secteur. Inversement, un secteur ou une entreprise en phase d'expansion avant la crise pour lequel un taux de croissance légèrement négatif est observé en 2020 correspond à une grande sous-estimation du réel effet de la crise : on lui attribue une légère baisse d'activité alors que, hors crise, le taux de variation aurait pu être nettement positif. Ce scénario empire si l'activité attendue s'inscrit dans un scénario de très forte croissance : il est alors possible d'observer un taux de variation positif de l'activité, malgré de fortes pertes par rapport à ce qui aurait dû être observé sans crise. Ce phénomène peut concerner en particulier de jeunes entreprises en forte croissance.

Pour estimer l'effet de la crise, il faut donc construire un scénario « contrefactuel » pour l'année 2020, qui soit cohérent à la fois au niveau agrégé et au niveau individuel. Autrement dit, l'objectif est d'élaborer un scénario de ce qu'aurait été l'évolution de l'activité en l'absence de crise, de sorte que les montants contrefactuels totaux d'activité se ventilent mois par mois entre l'ensemble des entreprises composant le tissu productif français de manière cohérente avec la trajectoire de croissance ou de décroissance, ainsi qu'avec la saisonnalité de chacune d'entre elles.

Pour satisfaire cet objectif, l'estimation des contrefactuels est effectuée en deux étapes. La première consiste à estimer un contrefactuel du montant d'activité attendu hors-crise à un niveau d'agrégation correspondant au croisement de la taille et du secteur d'activité. La seconde ventile ces montants contrefactuels entre les entreprises composant chaque groupe *secteur x taille*.

1. Estimation des dynamiques hors-crise à un niveau méso-économique

La première étape estime un chiffre d'affaires contrefactuel total par secteur et catégorie de taille. Pour cela, 16 secteurs de la nomenclature agrégée A17⁸ et 3 tailles d'entreprise (TPE, PME et ETI-GE) sont mobilisés, pour un total de 44 séries différentes étudiées⁹. Le choix du croisement du secteur et de la taille dans l'analyse des dynamiques hors crise a été motivé par plusieurs raisons : (i) il permet de conserver une bonne finesse d'analyse, même au niveau le plus agrégé des simulations ; (ii) dans certains secteurs, la saisonnalité des TPE n'est pas identique à celles des entreprises de taille plus importante ; et (iii) effectuer séparément les prédictions pour les montants d'activités générés par les petites entreprises permet d'obtenir des prédictions plus robustes pour ces dernières.

Le chiffre d'affaires total de ces groupes *secteur x taille* est d'abord reconstitué mensuellement entre janvier 2015 et janvier 2020, soit avant le début de la crise. Les observations sur cette période sont utilisées pour modéliser la dynamique hors-crise de chacune des 44 séries *secteur x taille s* étudiées. La modélisation consiste, pour chacune d'elles, en une stationnarisation puis une sélection du meilleur couple (p_s, q_s) de paramètres d'ordres auto-régressif et de moyenne mobile à incorporer dans un modèle SARIMA. Le couple sélectionné est celui minimisant le critère AIC parmi 64 combinaisons possibles de paramètres¹⁰ allant de $(p_s=1, q_s=1)$ à $(p_s=8, q_s=8)$ ¹¹.

In fine, cette procédure permet d'obtenir une modélisation robuste de chacune des séries *secteur x taille* transformées et stationnarisées correspondant à l'équation (1). En notant $Y_{t,s}$ le chiffre d'affaires du groupe *secteur x taille s* à la date t , B l'opérateur de retard et $X_{t,s} = (1 - B^{12}) \log(Y_{t,s})$ ¹² les équations correspondant à chaque série s'écrivent :

$$X_{t,s} = (\varphi_{1,s} X_{t-1,s} + \dots + \varphi_{p_s,s} X_{t-p_s,s}) + \varepsilon_{t,s} - (\psi_{1,s} \varepsilon_{t-1,s} + \dots + \psi_{q_s,s} \varepsilon_{t-q_s,s}), \forall s \quad (1)$$

Avec $(\varepsilon_{t,s})_{t=1, \dots, T}$ désignant un bruit blanc gaussien de variance σ^2

8 Seul le secteur financier est exclu de l'analyse, cf partie I.3 et les restrictions imposées à l'échantillon.

9 Le croisement de ces deux dimensions constitue 48 groupes *secteur x taille*. Certains de ces groupes comptent un faible nombre d'entreprises et sont fusionnés avec d'autres groupes du même secteur. Dans l'agriculture et le secteur de la santé, toutes les entreprises de plus de 10 salariés sont regroupées dans le même groupe *secteur x taille*. Dans le secteur de la cokéfaction et du raffinage, toutes les entreprises sont analysées ensemble.

10 Le critère AIC vaut $AIC = 2k - 2 \log(L)$ avec L la vraisemblance du modèle estimé, et k le nombre de paramètres libres du modèle. L'AIC repose sur un compromis entre la qualité de l'ajustement et la complexité du modèle, en pénalisant les modèles ayant un grand nombre de paramètres, ce qui limite les effets de sur-ajustement [Akaike, 1998].

11 La stationnarisation des séries transformées est testée statistiquement par des tests de Dickey-Fuller et de Dickey-Fuller augmentés [Dickey et Fuller, 1979] (Annexe D.1). À l'issue de la sélection du couple de paramètres, des tests de robustesse de la modélisation sélectionnée sont effectués. L'absence d'auto-corrélation sérielle des résidus, leur normalité et leur blancheur sont testées [Box et Pierce, 1970] [Ljung et Box, 1978] (Annexes D.2 à D.4). La significativité des coefficients associés aux paramètres des termes autorégressifs et de moyenne mobile est également testée par un z-test (Annexe D.5). Lorsque plus d'un de ces critères n'est pas vérifié pour une série, le couple de paramètres donnant la seconde valeur la plus faible pour l'AIC est sélectionné et la procédure est répétée.

12 La différence par rapport au même mois de l'année précédente est une approche classique dans la stationnarisation des séries temporelles. Dans une étape liminaire de l'analyse, la décomposition mensuelle des 44 séries *secteur x taille* a par ailleurs permis d'identifier une tendance saisonnière, justifiant ainsi le recours à des retards d'ordre 12 pour la stationnarisation des séries.

Ces équations sont ensuite mobilisées pour calculer la prévision linéaire optimale mensuelle d'horizon h de chaque série *secteur x taille*. Dans le cadre de l'étude, $h \in [1, 11]$ car la prévision est effectuée entre février et décembre 2020 :

$$X_{T+h,s}^{\hat{}} = EL[X_{T+h,s} | X_{1,s}, \dots, X_{T,s}], \forall h \in [1, 11] \quad (2)$$

Ces prévisions prennent la période janvier 2015-janvier 2020 comme période d'apprentissage. La transformation de la série des prévisions obtenues avec l'équation (2) permet d'obtenir la série $(Y_{T+h,s}^{\hat{}})_{h \in [1, 11]}$ correspondant à l'estimation du chiffre d'affaires contrefactuel au cours de chaque mois de 2020 pour chaque groupe *secteur x taille*.

2. Calcul des montants d'activités contrefactuels et des chocs individuels

La seconde étape consiste à répartir le chiffre d'affaires contrefactuel total estimé précédemment entre toutes les entreprises d'un groupe *secteur x taille*. Cette ventilation se fait successivement mois par mois, en commençant par février 2020 pour finir par décembre 2020. La part de marché mensuelle de chaque entreprise dans son groupe d'appartenance incorpore la saisonnalité propre de son activité et sa dynamique récente de développement. Formellement, la part individuelle attribuée à chaque entreprise i du groupe *secteur x taille* s au premier mois t (ici février 2020) vaut :

$$S_{i,s,t}^{\hat{}} = \frac{1}{2} \left(S_{i,s,t-12} + \frac{1}{3} \sum_{j=1}^3 S_{i,s,t-j} \right) \cdot \left(1 + \frac{1}{2} \frac{\left(\sum_{j=1}^3 Y_{i,t-j} - \sum_{j=1}^3 Y_{i,t-12-j} \right)}{\frac{1}{2} \left(\sum_{j=1}^3 Y_{i,t-j} + \sum_{j=1}^3 Y_{i,t-12-j} \right)} \right) \quad (3)$$

Avec $S_{i,s,t}$ la part de marché de l'entreprise i au sein du groupe *secteur x taille* s à la date t . La part de marché attribuée à chaque entreprise au mois de février correspond à la moyenne de sa part de marché au cours des 3 mois précédents et de sa part de marché en février 2019, à laquelle est ajouté un coefficient de pondération individuel visant à incorporer la tendance de croissance ou de décroissance de l'entreprise sur l'année écoulée.

L'idée derrière cet indicateur est que la part de marché des entreprises peuvent connaître une saisonnalité différente de celle de leur groupe *secteur x taille*, justifiant le recours à la part de marché un an auparavant comme contribution à la part attendue pour modéliser cette saisonnalité¹³. De plus, la seconde hypothèse était que les parts de marché mensuelles pouvaient connaître des fluctuations importantes. La prise en compte de la moyenne glissante sur les 3 mois précédant permet par exemple de lisser un potentiel résultat exceptionnel un mois donné et de donner une image plus robuste de la position de l'entreprise au sein de son groupe *secteur x taille*. Au-delà de la prise en compte de ces deux

13 Par exemple, les hôtels des stations de ski connaissent un pic d'activité pendant l'hiver, mais appartiennent au secteur de l'hébergement-restauration, dont le pic saisonnier intervient pendant l'été. Ce phénomène peut concerner à la fois des saisonnalités propres à des sous-secteurs mais peut également concerner des entreprises dont la saisonnalité individuelle est particulière. Dans les deux cas, il était important de prendre en compte cette dimension dans la modélisation.

dimensions (saisonnalité et moyenne glissante sur les mois précédants), la réflexion a été menée à partir de cas limites de développement et de saisonnalité d'entreprises afin de choisir un indicateur permettant de modéliser le mieux possible les différentes trajectoires possibles pour chaque entreprise. Par ailleurs, les différents indicateurs possibles ont été départagés en effectuant une simulation en pseudo temps réel sur l'année 2019 afin de sélectionner le meilleur indicateur sur la base de deux critères de décision : l'indicateur ayant le taux d'erreur le plus faible sur le signe (positif ou négatif) des taux de croissance mensuels simulés d'une part, et l'indicateur conduisant aux taux de variation entre les CA simulé et observé le plus faible d'autre part¹⁴.

Le coefficient de (dé)croissance s'appuie sur la structure des indicateurs de Haltiwanger et Davis et est compris par construction entre 0 et 2. En dessous de 1, il simule une tendance annuelle de décroissance. Au-dessus de 1, une tendance annuelle de croissance. La forme de cet indicateur de (dé)croissance a été dicté par la volonté de disposer d'un indicateur borné, en particulier pour les entreprises en croissance forte, et centré en 1 ((dé)croissance nulle).

La part et le montant d'activité contrefactuels du mois de mars (t+1) sont calculés en s'appuyant sur la même équation (2) mais où la part de marché du mois précédent (février) n'est pas celle observée, mais celle estimée à l'étape précédente, et ainsi de suite jusqu'en décembre (t+10) :

$$\left\{ \begin{array}{l} S_{i,s,(t+1)}^{\hat{}} = \frac{1}{2} (S_{i,s,(t+1)-12} + \frac{1}{3} (S_{i,s,t}^{\hat{}} + \sum_{j=2}^3 S_{i,s,(t+1)-j}^{\hat{}})) * (1 + \frac{1}{2} \frac{((Y_{i,(t+1)-1}^{\hat{}} + \sum_{j=2}^3 Y_{i,(t+1)-j}^{\hat{}}) - \sum_{j=1}^3 Y_{i,(t+1)-12-j}^{\hat{}})}{(\sum_{j=2}^3 Y_{i,(t+1)-j}^{\hat{}} + \sum_{j=1}^3 Y_{i,(t+1)-12-j}^{\hat{}})}) \\ \dots \\ S_{i,s,(t+10)}^{\hat{}} = \frac{1}{2} (S_{i,s,(t+10)-12} + \frac{1}{3} \sum_{j=2}^3 S_{i,s,(t+10)-j}^{\hat{}}) * (1 + \frac{1}{2} \frac{(\sum_{j=1}^3 Y_{i,(t+10)-j}^{\hat{}} - \sum_{j=1}^3 Y_{i,(t+10)-12-j}^{\hat{}})}{(\sum_{j=1}^3 Y_{i,(t+10)-j}^{\hat{}} + \sum_{j=1}^3 Y_{i,(t+10)-12-j}^{\hat{}})}) \end{array} \right. \quad (4)$$

Les parts de marchés mensuelles ainsi calculées sont ensuite ajustées de sorte qu'elles se somment à 1 au sein de chaque groupe *secteur x taille* :

$$\widetilde{S}_{i,s,t} = \frac{S_{i,s,t}^{\hat{}}}{\sum_{i=1}^n S_{i,s,t}^{\hat{}}} \quad (5)$$

Le chiffre d'affaires contrefactuel individuel s'obtient en multipliant la part de marché individuelle ainsi estimée par l'activité totale contrefactuelle du groupe *secteur x taille* auquel appartient l'entreprise à la date correspondante :

$$CA_{i,s,t}^{\hat{}} = \widetilde{S}_{i,s,t} \hat{Y}_{t,s} \quad (6)$$

Le choc d'activité mensuel estimé est la différence, exprimée en pourcents, entre le chiffre d'affaires observé et ce chiffre d'affaires contrefactuel.

14 En particulier, différentes pondérations entre les parts de marchés des trois mois précédants et du même mois de l'année précédente ont été comparées

$$Choc_{i,s,t} = \left(\frac{\hat{CA}_{i,s,t} - CA_{i,s,t}}{CA_{i,s,t}} \right) \cdot 100 \quad (7)$$

Une seconde estimation des chocs, dite à la Haltiwanger et Davis est également calculée [Davis et Haltiwanger, 1992]. Ces indicateurs de chocs sont bornés entre -200 (cessation complète d'activité alors qu'une activité positive était attendue) et 200 (activité observée positive et montant d'activité contrefactuel nul).

$$Choc_{i,s,t}^{HD} = \left(\frac{\hat{CA}_{i,s,t} - CA_{i,s,t}}{\frac{1}{2}(\hat{CA}_{i,s,t} + CA_{i,s,t})} \right) \cdot 100 \quad (8)$$

Sauf mention contraire, les chocs de l'équation (7) sont mobilisés. Les chocs à la Haltiwanger-Davis sont mobilisés dans les tests de robustesse.

L'obtention de ces séries de chiffres d'affaires observés et contrefactuels permet de varier les approches. En effet, en sommant – mois par mois ou sur l'année – les chiffres d'affaires contrefactuels des entreprises de tout l'échantillon ou d'un secteur donné et en comparant ces sommes à l'agrégation des chiffres d'affaires observés sur le même périmètre, il est possible de construire des chocs d'activité agrégés. Ces chocs traitent l'activité des secteurs ou de l'économie dans leur ensemble et renseignent l'écart du montant d'activité par rapport à une année normale, sans prendre en compte la diversité des situations individuelles. Au contraire, les analyses de la distribution des chocs d'activité individuels tels que calculés en (7) permettent d'affiner les analyses en identifiant entreprises gagnantes et perdantes, y compris au sein d'un même secteur.

3. Réplication de l'exercice en 2019, année sans crise

L'analyse de ces entreprises « gagnantes » et « perdantes » doit cependant être nuancée à un niveau individuel. En effet, la méthode utilisée pour attribuer les chiffre d'affaires contrefactuels individuels repose sur une projection déterministe. De ce fait, elle ne permet pas de prendre en compte les aléas irréductibles, empêchant ainsi de formellement considérer que chaque situation contrefactuelle individuelle serait exactement ce qui se serait passé en l'absence de crise. Plutôt, les contrefactuels individuels et leurs méthodes participent d'un scénario de ce qui *aurait pu* être observé sans crise, sur la base de l'information disponible au moment du déclenchement de cette dernière. En ce sens, l'analyse de la prévalence des entreprises gagnantes ou perdantes (et de l'ampleur de ces gains ou pertes) dans la simulation doit plutôt être effectuée en termes de leur distribution et du caractère exceptionnel de cette distribution du fait de la crise. Dans cette optique, l'exercice de simulation sur l'année 2020 a été répliqué une année sans crise : en 2019. Cela permet, en creux, de mettre en évidence le caractère exceptionnel du choc de 2020. La comparaison des résultats obtenus en 2019 et 2020 et les enseignements

pouvant en être tirés sur le fonctionnement et la robustesse du modèle sont développés longuement dans la discussion.

La dernière étape de l'analyse consiste à replacer la dimension temporelle des séries de chocs au centre des traitements en caractérisant les similitudes et différences saillantes parmi les trajectoires de l'activité des entreprises en 2020.

4. Classification des entreprises selon leur trajectoire de chocs

a) Répartir les entreprises selon leur dynamique de pertes d'activité mensuelles : construction d'une distance dans l'espace des trajectoires d'activité des entreprises par clustering de courbes

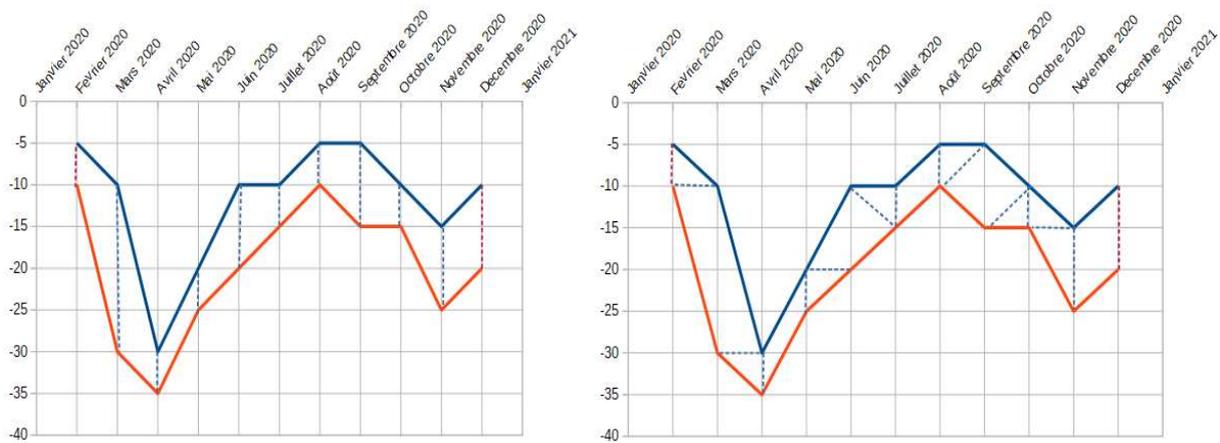
Afin de rendre comparables les trajectoires des entreprises, indépendamment de leur montant d'activité observé ou attendu, ces dernières ont été analysées à travers le prisme de la série de leurs chocs d'activité mensuels entre février et décembre 2020. Cette approche permet de réunir les entreprises ayant connu en 2020 des séries de chocs d'activité similaires. Cela permet en particulier d'identifier les entreprises les plus touchées lors des différents confinements, et celles ayant été capables de compenser tout ou partie de leurs pertes d'activité du premier confinement pendant la période de reprise estivale. Surtout, raisonner sur les séries des chocs rend *a priori* comparables les trajectoires de toutes les entreprises et rend possible l'identification de groupes homogènes dont l'identification n'est pas guidée par des différences de montants d'activité réalisés. Au contraire, étudier les trajectoires à travers les séries des montants d'activité réalisés ou de l'estimation nominale des pertes d'activité (en euros par exemple) aurait conduit à réunir des entreprises de tailles comparables dans chaque groupe.

L'identification des profils de chocs des entreprises sur l'ensemble de l'année 2020 a été réalisée par un clustering de courbes à partir des données des trajectoires de chocs mensuels individuels $(Choc_{i,s,t})_{i \in [1,N], t \in [fév. 2020, déc. 2020]}$ avec N le nombre d'entreprises de l'échantillon. Cette méthode consiste à partitionner une population de séries temporelles en un nombre donné de classes, de sorte à regrouper des séries les plus semblables possibles au sein de chaque classe. La mesure utilisée pour comparer les séries afin de réaliser cette partition est la distance DTW (*Dynamic Time Warping*). Cette distance compare les séries deux à deux en calculant dans un premier temps la matrice de coût local LCM pour chaque paire. Les éléments de cette matrice sont définis comme la distance entre les couples de points de chacune des deux séries selon l'équation (9) :

$$\forall x, y \text{ séries de chocs}, \forall i, j \in [1, T], LCM(i, j) = |x_i - y_j| \quad (9)$$

Cette matrice carrée de dimension $T \times T$ avec T le nombre de dates étudiées, est ensuite parcourue de $LCM(1, 1)$ à $LCM(T, T)$ en choisissant à chaque étape l'élément $LCM(i, k) = \min_{j \in [i-t, i+t]} (LCM(i, j))$ où t est la largeur de la fenêtre de déformation. Le chemin ainsi défini minimise le coût de passage de la série x à la série y . Du fait de la fenêtre de déformation, la valeur d'un mois donné n'est pas nécessairement comparée au même mois de la seconde série. Afin de prendre en compte la dimension temporelle des données, c'est-à-dire la succession d'observations selon un ordre déterminé, cette déformation n'a cependant lieu qu'au sein d'une fenêtre de largeur fixe. Conformément à la littérature [Aghabozorgi et al, 2015] la taille t de fenêtre a ici été fixée à 10 % de la taille de la série, soit un mois. La **Figure 2** offre un exemple d'alignement de deux séries selon cette méthode.

Figure 2 - Alignement de séries selon deux méthodes : point à point à gauche et avec déformation à droite



On note Φ l'ensemble des éléments de la matrice LCM constituant le chemin optimal, m_Φ le poids associé aux éléments de Φ et M_Φ le coefficient de normalisation associé. Ces deux coefficients assurent la comparabilité des distances malgré les distorsions différentes appliquées selon les chemins choisis. La distance entre les séries x et y est alors définie selon l'équation (10) :

$$DTW(x, y) = \left(\sum_{(i,j) \in \Phi} \frac{m_\Phi LCM(i, j)^2}{M_\Phi} \right)^{1/2} \quad (10)$$

L'algorithme de classification répartit itérativement les séries au sein de différentes classes de sorte à minimiser cette distance au sein de chaque classe et à la maximiser entre les classes. À chaque étape, les séries sont comparées au centre de chaque classe (ici la médiane), et assignées à la classe dont le centre est le plus proche. Ce processus est répété jusqu'à obtenir une partition stable des séries. Le résultat de cette classification dépend à la fois du nombre de classes choisi et des centres tirés aléatoirement et ayant initialisé la partition. Le choix d'une partition en 4 classes a été déterminé afin

d'optimiser la qualité de la partition tout en conservant un nombre de classes important (**Annexe H**). La classification a de plus été répétée afin de s'assurer de la stabilité de la partition finale. La confusion, *i.e.* la part d'entreprises changeant de classe entre différentes partitions prenant des centres de classes initiaux différents est proche de zéro dans l'ensemble des répétitions. Les classes présentées constituent donc une partition fiable de la population étudiée.

b) Expliquer la ventilation des entreprises entre les différents profils-types identifiés : mise en place d'un modèle de classification

L'identification des profils de trajectoire et la répartition des entreprises s'appuie exclusivement sur les chocs d'activité mensuels estimés pour chaque entreprise. Une fois ces groupes d'entreprises aux trajectoires d'activité similaires pendant la crise identifiés, on étudie l'éventuelle homogénéité des caractéristiques des entreprises composant chaque groupe. Autrement dit, le fait pour une entreprise d'être assignée à un profil de trajectoire plutôt qu'aux autres ne dépend par construction que de la structure de ses pertes d'activité, mais ces dernières peuvent être corrélées avec des caractéristiques individuelles des entreprises, comme son secteur d'activité ou sa taille. Ainsi, pour tenter d'éclairer la répartition des entreprises entre les différents types de trajectoires, les corrélations entre le profil assigné aux entreprises et leurs caractéristiques sont étudiées.

Cette analyse s'appuie sur l'enquête Insee de 2020 relative à l'*Impact de la crise sanitaire sur l'organisation de l'activité des entreprises* [Duc et Souquet, 2020]. Outre les informations disponibles dans Fare (secteur d'activité, taille, date de création de l'entreprise) et dans les déclarations de TVA (existence d'une activité exportatrice), cette enquête fournit des informations sur le comportement des entreprises durant la crise, en particulier sur leur stratégie d'adaptation de l'organisation de leur activité : la proportion de salariés en télétravail, la réorganisation de la logistique commerciale des entreprises pendant les confinements (développement de systèmes de vente en ligne, de vente directe ou de nouveaux systèmes de livraison), l'adaptation de l'offre à travers le développement de nouveaux produits, activités ou services, la réalisation d'investissements spécifiques notamment dans les nouvelles technologies, ainsi que la réorganisation de l'activité via une modification des fournisseurs et partenaires commerciaux ou encore via la mise en commun de ressources avec d'autres entreprises. La plupart de ces variables sont catégorielles. Parmi les variables continues, seule la proportion des effectifs salariés en télétravail est conservée. La date de création de l'entreprise est recodée pour en faire une variable catégorielle. Les modalités de cette variable sont définies en fonction des quartiles de la distribution de la date de création renseignée dans l'enquête et distinguent quatre catégories d'effectifs comparables (créées avant 1997, entre 1998 et 2006, entre 2007 et 2012, après 2013).

L'appariement avec les données d'enquête restreint l'échantillon à 13 500 entreprises. Cette restriction altère la structure de l'échantillon, notamment en termes de tailles d'entreprises puisque les petites entreprises ne faisaient pas partie de l'échantillon de l'enquête. De plus, rien ne garantissait *a*

priori que les données d'enquêtes comportent des entreprises des différents profils de trajectoires identifiés dans les mêmes proportions. Afin de conserver un échantillon dont la proportion d'entreprises affectée à chaque profil de trajectoire est similaire à la structure de la répartition de l'échantillon total de l'échantillon¹⁵, et une distribution identique en termes de taille, secteur d'activité, date de création et existence d'une activité exportatrice, chaque entreprise de l'échantillon *post*-appariement se voit attribuer un poids. Ces poids sont déterminés à travers une méthode de calage sur marge [Deville, J.-C. et Särndal, 1992] [Rebecq, 2016]. Après calage, la similitude de la distribution jointe de ces variables dans l'échantillon apparié et dans l'échantillon original est également vérifiée.

Ces données sont ensuite mobilisées dans un modèle de classification logit multinomial non-ordonné afin d'étudier les corrélations statistiques entre le profil de trajectoire assigné par le clustering de courbe, et un ensemble de caractéristiques des entreprises : la taille de l'entreprise, sa date de création, l'existence d'une activité exportatrice, son secteur d'activité ainsi que les variables organisationnelles issues de l'enquête *Impact de la crise sanitaire sur l'organisation de l'activité des entreprises* (Détails du modèle en [Annexe K](#)).

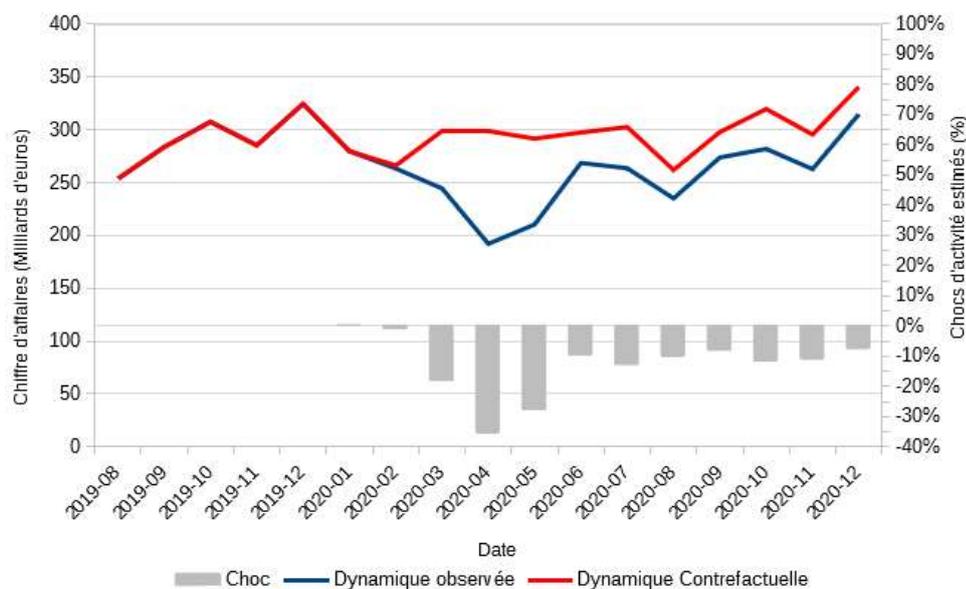
III. Résultats

1. Un choc d'activité brutal d'ampleur variable selon le secteur

L'activité économique, au sens de chiffre d'affaires, a été très ralentie pendant le premier confinement. Entre mars et mai, son niveau se situe 27 % en dessous du niveau d'activité estimé en l'absence de crise sanitaire. Sur le seul mois d'avril, cet écart est de -35 %. L'activité économique a ensuite connu un rebond marqué, son niveau restant cependant 10 % en dessous du niveau attendu hors crise entre juin et octobre. La perte d'activité du printemps n'a donc pas été compensée par une activité plus élevée au cours de l'été ou au début de l'automne. La perte d'activité au quatrième trimestre, qui inclut le second confinement, est estimée à environ 10 %. D'une part, le second confinement a été plus court et moins contraignant que le premier. D'autre part, les entreprises ont pu adapter leur stratégie et leur organisation, ce qui était moins le cas au début de la crise. Sur le mois de novembre, concerné en intégralité par le second confinement, les pertes d'activité sont estimées à -11 %, chiffre bien inférieur aux -35 % d'avril ([Figure 3](#)).

15 Cet élément est important dans la mesure où le modèle compare *in fine* les probabilités d'appartenir à chaque profil de trajectoire.

Figure 3 – Evolution du choc d'activité économique agrégé au cours de l'année 2020



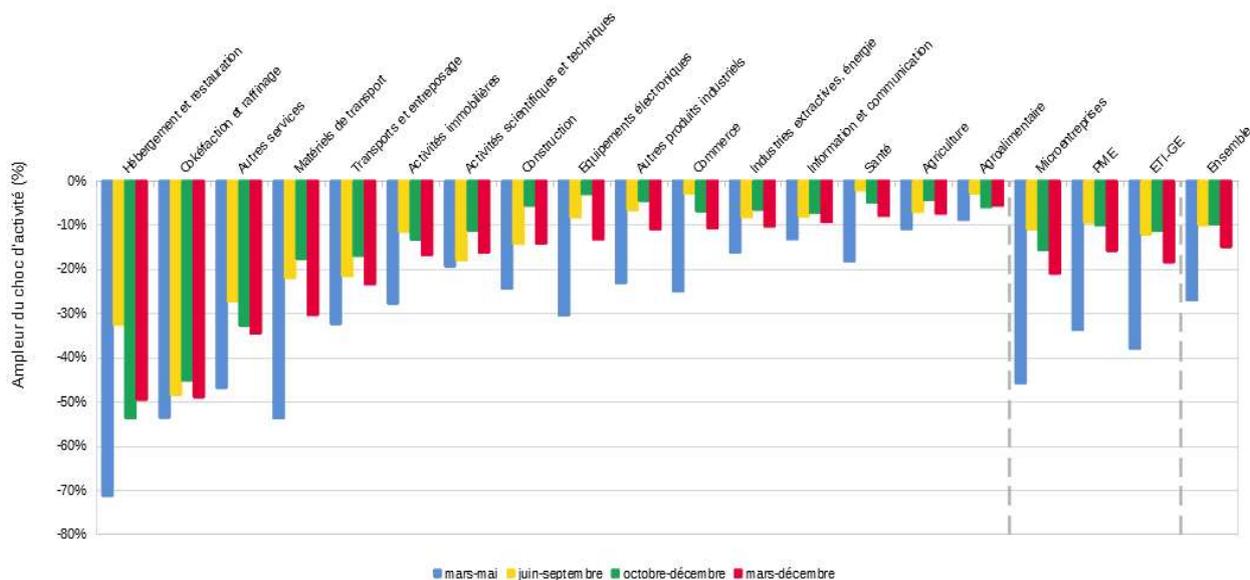
Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

Le premier confinement a constitué un choc pour tous les secteurs, mais son ampleur a été variable. L'hébergement-restauration et la fabrication de matériels de transport sont les deux secteurs dont l'activité économique a été la plus déprimée avec une perte d'activité estimée respectivement à -71 % et -54 % entre mars et mai. Le secteur de l'information et de la communication, l'agriculture et pêche et la fabrication alimentaire ont bien mieux résisté (respectivement -13 %, -11 % et -9 %). Au contraire, lors du second confinement, seuls certains secteurs ont vu leurs pertes d'activité s'amplifier à nouveau de manière significative après la reprise modérée de la période estivale. Se retrouvent dans cette situation l'hôtellerie-restauration (-54 %) et les « autres services »¹⁶ (-33 %). Pour l'essentiel des autres secteurs, l'activité ne baisse que de façon plus limitée (**Figure 4**).

Dans un certain nombre de secteurs industriels, comme l'électronique et les autres produits industriels, l'activité économique a rebondi entre les deux confinements et s'est presque rétablie au niveau attendu lors du second confinement (-3 % et -5 % respectivement pendant le second confinement).

16 Les « autres services » sont essentiellement composés des arts, spectacles et activités récréatives, des services de coiffure et de soins corporel et des réparations d'ordinateurs et autres biens personnels.

Figure 4 – Le choc d’activité économique cumulé de mars à décembre 2020 : ventilation sectorielle



Source : Déclarations TVA (DGFiP) ; Calculs des auteurs.

2. Une hétérogénéité individuelle dépassant la seule appartenance sectorielle et éclairant les difficultés des petites entreprises

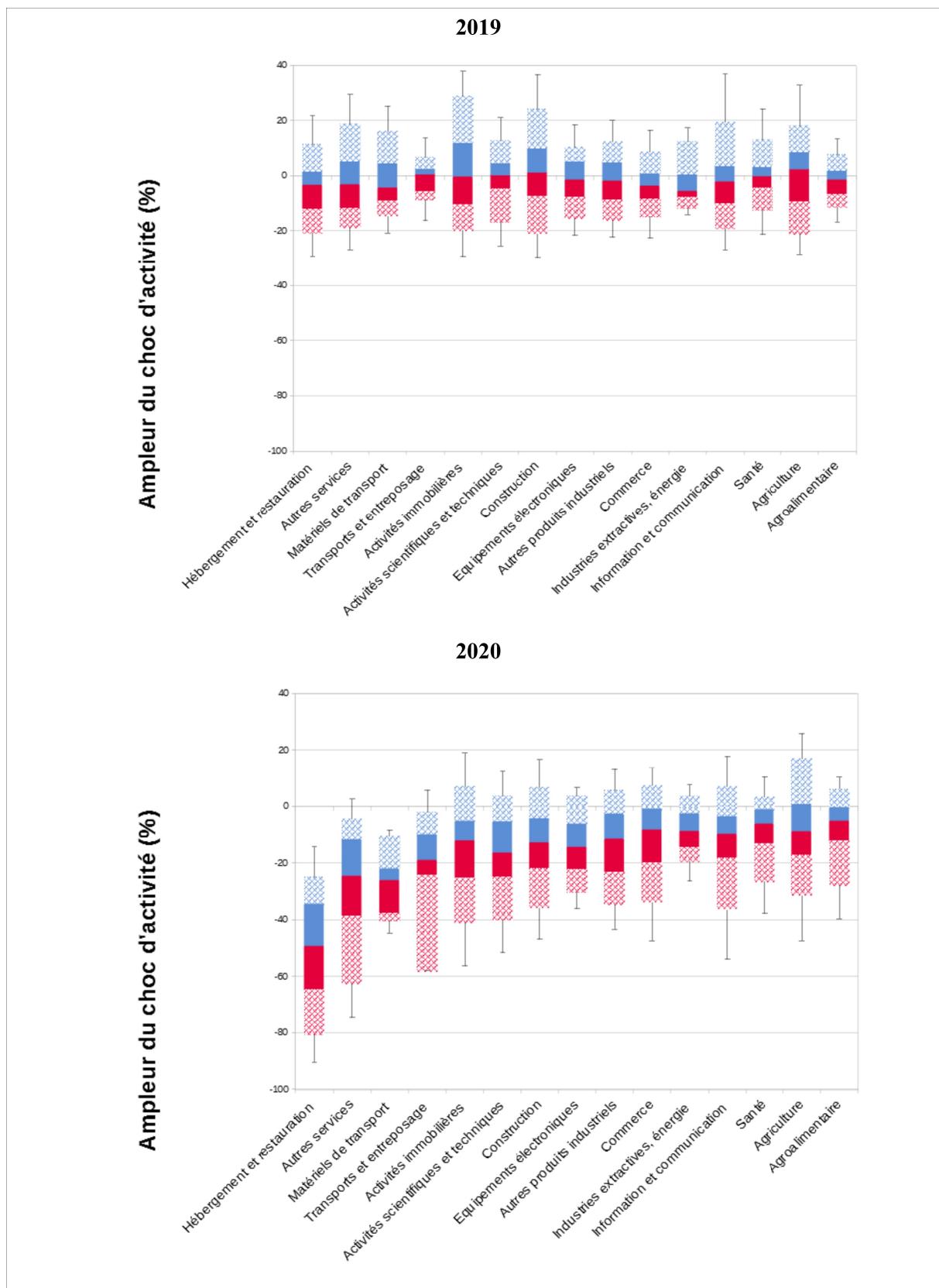
La dispersion des situations individuelles des entreprises depuis le début de la crise permet de compléter l’analyse sectorielle. Pour apprécier la dispersion des trajectoires individuelles de chocs d’activité tout au long de la crise, la dispersion des trajectoires de chocs qui serait constatée lors d’une année sans crise est prise comme référence. En effet, quel que soit le contexte, des différences s’observent toujours entre la trajectoire effective et la trajectoire estimée par un modèle à partir des observations des années précédentes. La question est de savoir si ces différences sont particulières en 2020 (*cf* méthode, partie II.3).

La **Figure 5** présente, pour chaque secteur, les principaux quantiles, pondérés par les effectifs salariés, de la distribution des chocs d’activité estimés pour l’année 2019, année « normale », et pour 2020¹⁷. En 2019, les situations individuelles sont dispersées, certaines entreprises enregistrent une activité plus élevée qu’anticipée, d’autres plus faible, mais on recense à peu près autant de chocs positifs que de chocs négatifs. En outre, la dispersion intra-sectorielle varie peu entre les différents secteurs. En 2020, la situation diffère triplement. D’abord, le recul général de l’activité se traduit par une proportion importante de chocs négatifs et une proportion faible de chocs positifs. Ensuite, la dispersion intra-sectorielle des chocs individuels est plus importante en 2020, et ce pour tous les secteurs. Enfin,

17 La dispersion des chocs individuels d’activité pondérés par les effectifs reflète la dispersion des chocs pour les salariés appartenant à ces entreprises. La dispersion des chocs individuels (non pondérés par les effectifs) reflète la dispersion des chocs pour les entreprises : les TPE étant largement majoritaires dans l’économie et dans l’échantillon, cette dispersion non pondérée donne surtout une image de la variété des chocs pour les micro-entreprises (80 % de l’échantillon).

les chocs médians, tous négatifs, diffèrent entre secteurs en 2020 alors qu'ils sont comparables (et proche de zéro) en 2019.

Figure 5 – Dispersion des chocs d'activité par secteur en 2019 et 2020



Note de lecture : Chaque ligne représente la ventilation des chocs individuels d'activité au sein d'un secteur via une boîte à moustaches. Les différents segments des boîtes distinguent les quantiles à 10 %, 25 %, 50 % (médiane), 75 % et 90 %. Les extrémités (moustaches) représentent la valeur des quantiles à 5 % et 95 %. En raison des faibles effectifs le composant, le secteur Cokéfaction-raffinage n'est pas représenté sur ce graphique.

Lecture : En 2020, dans le secteur dans le secteur "Hébergement-Restauration", 5 % des salariés travaillent dans une entreprise ayant connu un choc d'activité au moins inférieur ou égal à -90,4 %, 10 % dans une entreprise avec un choc inférieur ou égal à -80 %, 25 % avec un choc inférieur ou égal à -65 %; 50 % avec un choc inférieur ou égal à -50 %, 75 % inférieur ou égal à -35 %, 90 % inférieur ou égal à -25 %, 95 % inférieur ou égal à -14 %.

Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

La forte hétérogénéité des chocs observés au sein de chacun des grands secteurs en 2020 s'explique par la présence simultanée de chocs négatifs importants, parfois de cessation d'activité, et d'entreprises parvenant à l'inverse à réaliser leur niveau d'activité attendu malgré la crise. L'inégale répartition des mesures de restriction sanitaire au sein d'un même grand secteur peut expliquer cette situation. En particulier, les décisions de fermeture administrative, très ciblées, n'ont touché aucun grand secteur dans son ensemble, mais des sous-secteurs définis très finement. Cette hypothèse est en partie confirmée par la comparaison des chocs médians, pondérés par les effectifs, subis par les entreprises en fonction de leur secteur défini à un niveau plus fin (classes NAF, 637 classes sur le champ de cette étude). Y compris dans les grands secteurs les plus touchés par la crise, les chocs médians annuels varient beaucoup entre sous-secteurs.

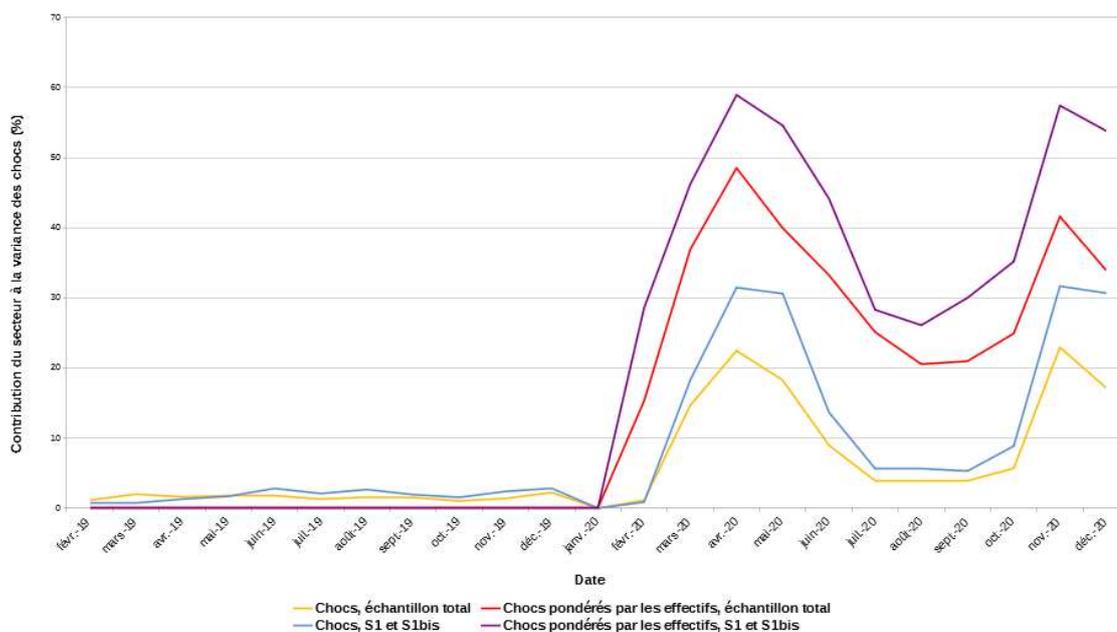
Dans l'hébergement-restauration, les établissements de restauration rapide se sont mieux maintenus (-34 %) que les débits de boissons (-55 %) ou les services de traiteurs (-70 %), tous deux concernés par une fermeture administrative en mars. Dans les « autres services », la gestion des salles de spectacle (-80 %) connaît la perte annuelle la plus importante alors que les services funéraires (-4 %) font partie des sous-secteurs s'étant le mieux maintenu. De la même façon, les grands secteurs les moins touchés, comme le commerce et la fabrication alimentaire, distinguent également des sous-secteurs fortement touchés et d'autres connaissant des gains, bien que modérés, d'activité. Pour autant, le secteur d'activité, même au niveau le plus fin de la nomenclature, peut-il être considéré comme unique facteur de l'hétérogénéité des chocs d'activité individuels ?

La décomposition de la variance des chocs mensuels entre une part attribuable au secteur et une part résiduelle¹⁸ montre bien que le secteur d'activité explique une part de la variance des chocs d'activité largement plus importante en 2020 (43 % avec pondération par les effectifs) qu'une année sans crise (0,02 % en 2019) et que le rôle du secteur dans l'hétérogénéité des chocs observés est plus important durant les mois de confinement, ayant inégalement affecté les secteurs d'activité. De plus, l'application de la même décomposition aux sous-secteurs les plus concernés par des mesures de restric-

¹⁸ La méthode de décomposition de la variance est assez usuelle et présentée notamment dans [Gibbons et al., 2014] et [Helpman, 2017]. La formule ainsi que sa démonstration complète sont présentées en [Annexe G](#).

tion sanitaires et de fermetures administratives (secteurs S1 et S1 bis¹⁹) attribue aux secteurs une part de la variance systématiquement plus élevée. En avril 2020, la contribution sectorielle à la variance des chocs est de 23 % et de 48 % avec pondération par les effectifs. Sur les secteurs S1 et S1 bis, cette part est de 31 % et de 59 % avec pondération (**Figure 6**). A titre de comparaison, la reproduction du même exercice en avril 2019 conduit à une estimation de la contribution sectorielle à la variance des chocs de 1,6 % et 0,03 % avec pondération des effectifs.

Figure 6 – Contribution de la variance intersectorielle à la variance des chocs d'activité



Note de lecture : Chaque ligne correspond à la part de variance intersectorielle dans la variance total des chocs en chaque mois. Les valeurs sur 2020 correspondent à la simulation centrale à ce dossier, celles sur 2019 correspondent aux résultats obtenus en appliquant le même exercice à l'année 2019.

Lecture : A son pic, en avril 2020 la variance intersectorielle représente 23 % de la variance totale des chocs sur l'échantillon. Si l'on pondère par l'emploi, elle représente 48 % de la variance totale. Au sein des secteurs S1 et S1bis, la part de variance intersectorielles s'élève à 31 % sur l'échantillon non pondéré et à 59 % sur l'échantillon pondéré. En répliquant l'analyse sur l'année 2019, la contribution de la variance intersectorielle est presque nulle : elle ne dépasse jamais 2,5 % sur l'année.

Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

Si la dimension sectorielle explique une part importante de la variance des chocs, la diversité des situations après un an de crise ne se limite donc pas à la nature de l'activité exercée. En particulier, dans la plupart des secteurs, les PME et TPE ont connu des pertes d'activité plus élevées que les ETI-GE lors des deux confinements, et un choc souvent plus important sur l'ensemble de l'année. Cela illustre les difficultés spécifiques rencontrées par les petites entreprises pendant la crise, au-delà du

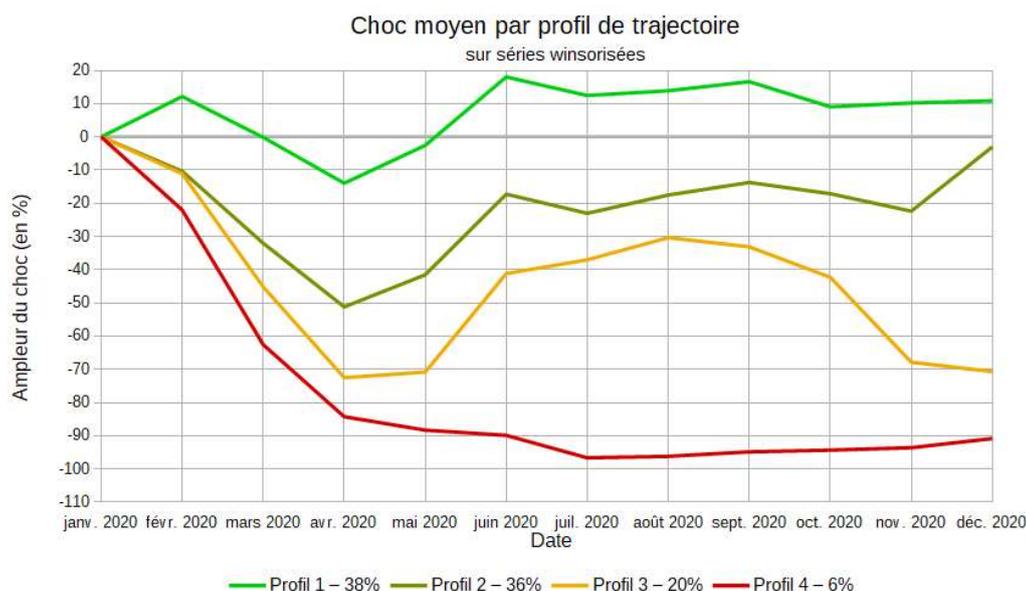
19 Les listes des secteurs S1 et S1bis sont définies par les modifications successives du décret 30 mars 2020 relatif au fonds de solidarité. L'évolution de ces listes a été reconstruite, mois par mois, sur l'ensemble de l'année 2020. La liste S1 concerne des secteurs particulièrement touchés par la crise et les fermetures administratives, en particulier dans les domaines de la restauration, du tourisme, de l'événementiel, de la culture et du sport. La liste S1bis concerne des secteurs liés, par exemple, à la distribution de films ou l'édition de livres.

seul secteur d'activité. Au-delà de la taille, les autres facteurs individuels expliquant la diversité des situations individuelles peuvent donc être explorés. De plus, les analyses qui précèdent décrivent au mois le mois les chocs subis par les entreprises et leur variabilité intra-sectorielle (analyse en coupes répétées), sans regarder si ce sont les mêmes entreprises qui d'un mois sur l'autre subissent les chocs les plus importants ou les plus faibles. La suite de l'analyse complète cette approche en s'intéressant aux trajectoires individuelles des entreprises (analyse en dynamique) et à leurs déterminants.

3. Quatre profils-types de trajectoire de chocs pendant la crise se dégagent

Dans l'objectif de mieux comprendre les facteurs, individuels ou collectifs, affectant la trajectoire de choc de chaque entreprise pendant la crise, les chocs sont approchés de façon dynamique. La classification des trajectoires par clustering de courbe permet d'identifier quatre profils-types représentatifs des trajectoires de chocs des entreprises pendant l'année 2020 (Figure 7, ou Annexe I pour la distribution des chocs mensuels au sein de chaque classe).

Figure 7 – Chocs moyens par profil de trajectoire



Note de lecture : Les entreprises du profil 1 ont connu un choc moyen de -14% en avril 2020.

Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

- **Les entreprises « non affectées »** (36 % des entreprises et 42 % des salariés) :

En moyenne, le 1^{er} confinement a eu un impact négatif limité sur ces entreprises, avec un choc moyen²⁰ de -14 % en avril. A partir de mois de juin, leur activité a rebondi et s'est établi en moyenne 10 % au-dessus de l'activité attendue. Le choc médian au sein de ce groupe est resté proche de zéro sur l'ensemble de l'année, à l'exception du premier confinement. Dans l'ensemble la distribution des

20 Toutes les moyennes évoquées dans ce paragraphe sont calculées sur des séries winsorisées à droite : tous les chocs supérieurs au 95^{ème} centile de choc sont ramenés à ce quantile.

chocs au sein de ce groupe est comparable à celle des chocs observés lors d’une année « normale » ([Annexe I](#)).

- **Les entreprises « résilientes »** (38 % des entreprises et 44 % des salariés) :

La perte d’activité moyenne de ces entreprises est plus substantielle au printemps, avec un choc moyen de –51 % en avril. À partir du mois de juin, les pertes sont moindres et le choc moyen reste stable autour de –20 % sur la fin d’année. Au cours de cette période, environ 75 % des entreprises de ce groupe a subi un choc négatif.

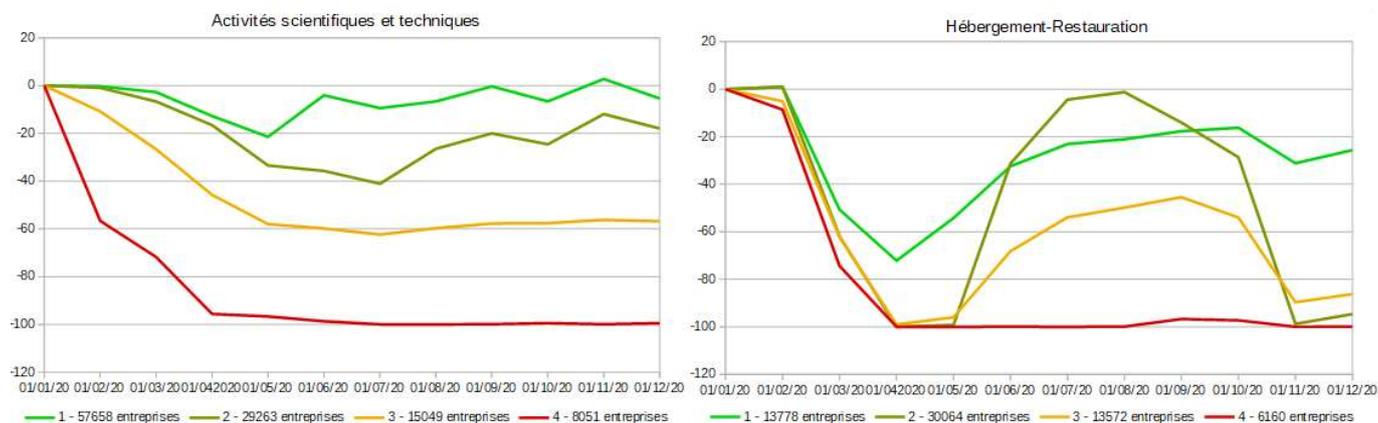
- **Les entreprises « confinées »** (20 % des entreprises et 12 % des salariés) :

La trajectoire moyenne de ce groupe se caractérise par des chocs importants lors des confinements (–72 % en avril et –70 % en novembre et décembre), avec une reprise limitée de l’activité pendant l’été. Le profil médian de ce groupe est très proche de son profil moyen.

- **Les entreprises « déprimées »** (6 % des entreprises et 2 % des salariés) :

L’activité de ces entreprises s’est effondrée lors du premier confinement (–84 % en moyenne en avril), sans reprise à l’été. Les chocs mensuels médians de ce groupe sont proches de –100 % d’avril à décembre 2020, et un tiers des entreprises de ce groupe déclare un chiffre d’affaires nul sur l’ensemble de cette période.

Figure 8 – Chocs médians par profil de trajectoire au sein des secteurs



Note de lecture : Les classes correspondent au résultat de la classification de trajectoire au sein de chaque secteur.

Lecture : Au sein du secteur « Activités scientifiques et techniques », les entreprises les moins affectées subissent un choc médian maximal de –22 % en mai 2020. Au sein du secteur « Hébergement-Restauration », le choc médian des entreprises les moins affectées atteint –74 % en avril 2020.

Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

La reproduction d’une partition au sein des secteurs d’activité ([Annexe J](#)) identifie des profils similaires à ceux observés sur l’ensemble de l’échantillon : on y retrouve une gradation allant des entreprises moins affectées aux entreprises plus affectées. Une différence notable entre ces partitions concerne les entreprises « confinées » qui n’apparaissent que dans certains secteurs.

La **Figure 8** trace par exemple les trajectoires des profils médians de l'hébergement-restauration, qui présente trois types d'entreprises « confinées » *i.e.* connaissant une nouvelle chute marquée de leur activité lors des deux confinements, et aucun profil « non-affecté ». Au contraire, le secteur des activités scientifiques et techniques ne présente pas de profil « confiné ». Cet écart confirme l'importance de la dimension sectorielle dans la trajectoire d'activité des entreprises, en particulier pendant le second confinement.

4. Le type de trajectoire est très corrélé au secteur d'activité et à l'adaptation organisationnelle

Les informations sur le secteur, la taille et l'âge des entreprises complétées par celles disponibles dans l'enquête sur l'*Impact de la crise sanitaire sur l'organisation et l'activité des entreprises* de l'Insee permettent de caractériser les entreprises appartenant à chacune des trajectoires-types.

Les coefficients issus du modèle de classification sont statistiquement significatifs²¹ (Tableau de régression disponible en **Annexe K**). Le secteur d'activité est le facteur prépondérant de la répartition des entreprises entre trajectoires-types. Il représente près de 85 % de la part de la ventilation des entreprises expliquée par le modèle²². Cette part importante est attribuable à la très forte dépendance sectorielle des profils les plus touchés, composés quasi intégralement d'entreprises de secteurs fermés administrativement pendant les confinements. Conditionnellement aux autres variables, les secteurs ayant la plus forte probabilité d'appartenir au profil le moins affecté sont ceux de la fabrication de produits électroniques grand public, les sous-secteurs de l'industrie alimentaire, les activités vétérinaires et le secteur médical. Dans le profil « résilient » se retrouvent majoritairement les secteurs de fabrication d'articles de joaillerie, d'ordinateurs et d'équipements périphériques et d'équipements automobiles. Les secteurs ayant la plus forte probabilité d'appartenir au profil « confiné » sont ceux des transports ferroviaires et des bibliothèques et musées. Enfin, dans le profil le plus affecté, les sous-secteurs de la culture de l'hébergement-restauration et du tourisme et des transports de voyageurs sont les plus surreprésentés.

Conditionnellement au secteur, l'effet de chacune des variables explicatives sur la probabilité de se retrouver dans les différentes classes est donc plutôt faible en valeur comparativement aux différences observées pour les secteurs²³. Autrement dit, la différence de probabilité d'appartenir au profil 4 est beaucoup plus forte entre deux secteurs ayant été concernés par des mesures de restriction différentes qu'entre deux modalités d'une autre variable incluse dans le modèle. Toutefois, en exprimant

21 Les observations SONT pondérées par les coefficients de pondération issus du calage sur marge lors de la régression. Cette pondération peut avoir un impact positif sur la significativité des effets affichés.

22 Estimation par mobilisation des R^2 ajustés de Cox-Snell [Cox et Snell, 1989]

23 Rappelons que ce modèle est une étude des corrélations et qu'il ne peut pas être interprété dans un sens de causalité.

les effets de chacune de ces variables en pourcentage d'évolution de la probabilité d'être affecté à chaque profil²⁴, plusieurs éléments se dégagent (**Figure 9**) :

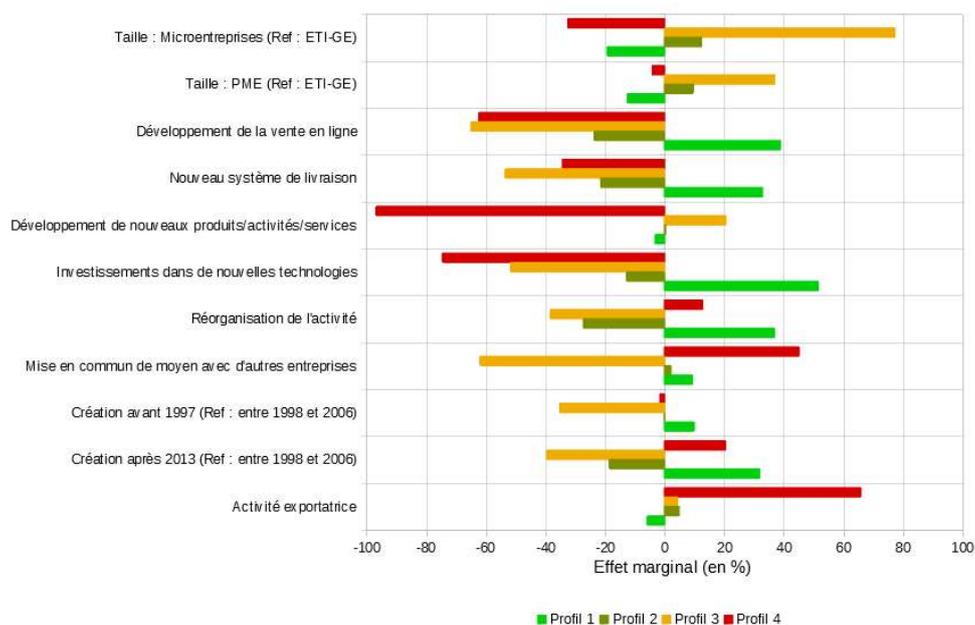
- Les PME et TPE, plus touchées en moyenne pendant les confinements ont, toutes choses égales par ailleurs, plus de chance d'appartenir au profil « confiné », tout particulièrement les TPE.
- Le fait d'avoir une activité exportatrice avant le début de la crise est associé à une probabilité plus élevée d'appartenir au profil « déprimé ». La dépendance aux débouchés étrangers et la chute de la demande extérieure peuvent expliquer une partie de cet effet.
- Le développement de nouveaux produits et systèmes de ventes suite à la crise est associé à une probabilité plus élevée d'appartenir au profil « non-affecté », et plus faible d'appartenir aux profils les plus touchés. La capacité à s'adapter aux restrictions sanitaires, en particulier celles concernant la manière d'organiser la distribution de ses produits a donc été importante. Cet effet va de pair avec un effet très positif de l'investissement dans des technologies nouvelles sur la probabilité d'appartenir au profil le moins touché, et très négatif sur la probabilité d'appartenir aux deux profils les plus touchés.
- La réorganisation de l'activité et la mise en commun de ressources avec d'autres entreprises est liée à une plus forte probabilité d'appartenir à la fois au profil « non-affecté », et au profil « déprimé ». Ce résultat peut s'expliquer par le fait que les entreprises ayant rapidement adapté leur activité ont pu mieux maintenir leur niveau de chiffre d'affaires. Cet effet peut être attribuable au contournement des pénuries des chaînes d'approvisionnement ayant pu toucher les entreprises en réorganisant ses circuits d'approvisionnement. Au contraire, la mise en commun de ressources ou la réorganisation du mode de production a pu être contrainte *a posteriori* pour les entreprises les plus touchées, expliquant un effet marginal positif dans le profil « déprimé » par un mécanisme de causalité inverse.

Ces résultats de régression permettent d'offrir une exploration plus détaillée des variables corrélées à l'hétérogénéité des situations observées. En particulier, si le secteur est bien le facteur explicatif principal de l'explication de cette hétérogénéité, le rôle joué par d'autres caractéristiques des entreprises ainsi que par leur comportement en termes de ré-organisation de leur activité pendant la crise permet d'affiner la compréhension de la dispersion observée. Ces variables offrent également une piste supplémentaire à l'explication de la diversité des situations observées à un niveau intra-sectoriel.

24 Ces effets s'appuient sur le calcul des probabilités prédites d'appartenir à chacun des profils pour toutes les modalités de l'ensemble des variables explicatives catégorielles. En comparant les probabilités prédites pour chaque profil de trajectoire pour différentes modalités d'une même variable catégorielle, il est possible de calculer l'effet relatif du passage d'une modalité à une autre sur la probabilité d'appartenir à chaque profil. Formellement, l'effet d'une variable binaire j sur la probabilité d'appartenir au profil c vaut :

$$Effet_{j,c} = \frac{\mathbb{P}(Profil_c | X_j=1, X_{-j}) - \mathbb{P}(Profil_c | X_j=0, X_{-j})}{\mathbb{P}(Profil_c | X_j=0, X_{-j})}, \forall c \in [1,4]$$

Figure 9 – Effets marginaux des variables du modèle de classification



Note de lecture : Le fait d'avoir développé la vente en ligne depuis le début de la crise est associé à une plus forte probabilité des entreprises d'appartenir au profil 1 de trajectoire. Par rapport aux entreprises n'ayant pas développé la vente en ligne, l'ampleur de cette différence est de +38 %. Autrement dit, les entreprises ayant développé la vente en ligne ont 1,38 fois plus de chance d'appartenir au profil 1 que les entreprises n'ayant pas développé de vente en ligne. Formellement :

$$\text{Effet}_{\text{vente en ligne}, \text{Profil}_1} = \frac{\mathbb{P}(\text{Profil}_1 | X_{\text{vente en ligne}} = 1, X_{-\text{vente en ligne}}) - \mathbb{P}(\text{Profil}_1 | X_{\text{vente en ligne}} = 0, X_{-\text{vente en ligne}})}{\mathbb{P}(\text{Profil}_1 | X_{\text{vente en ligne}} = 0, X_{-\text{vente en ligne}})} = 1,38$$

Source : Déclarations TVA (DGFIP) ; Enquête Impact de la crise sanitaire sur l'organisation et l'activité des entreprises (Insee) ; Calculs des auteurs.

IV. Discussion méthodologique

1. Cohérence des résultats : comparaison des enseignements du modèle de simulation avec les résultats issus des comptes nationaux, des études de conjoncture et des données d'enquête

La première dimension relative à la cohérence est qu'au sein du modèle, les prévisions individuelles sont cohérentes, par construction, avec les prévisions effectuées au niveau sectoriel. Ce souci de construction des contrefactuels est central dans la comparaison des résultats agrégés, *i.e.* au niveau de la taille ou du secteur, avec les résultats portant sur les distributions sur les chocs d'activité individuels.

Par ailleurs, les résultats de l'étude présentent une bonne cohérence avec les diagnostics macro-économiques issus des comptes nationaux dont les résultats ont été publiés au mois de mai 2021 [Amoureux et *al.*, 2021]. Dans ces derniers, le montant macro-économique du chiffre d'affaires n'est pas renseigné, mais il est possible d'effectuer une comparaison avec les fluctuations estimées entre 2019 et 2020 des montants de valeur ajoutée (VA). À périmètre équivalent à celui de notre champ d'étude, les comptes nationaux estiment une valeur ajoutée de 905 milliards pour les sociétés non-financières en 2019 et de 829 milliards en 2020, soit un taux de croissance de $-8,4\%$. Dans le cadre de l'étude, la valeur ajoutée [Bureau et *al.*, 2021a] est estimée à 869 milliards en 2019 et 770 milliards en 2020, soit un taux de croissance de $-11,4\%$ ²⁵.

Les chocs estimés par nos modèles, au niveau de l'économie totale, sont aussi en ligne avec les pertes d'activité mensuelles estimées dans les *Notes de conjoncture* publiées par l'Insee en termes de profil de choc, et avec l'ampleur du choc au pic de la crise (-35% en avril 2020 dans nos travaux, contre -30% au même mois dans la note de juillet 2021). Les deux estimations présentent de légers écarts dans l'ampleur estimée suite au premier confinement : les pertes se stabilisent autour de -5% dans la *Note de conjoncture*, et autour de -10% dans nos travaux. Dans l'ensemble ces écarts restent faibles et peuvent s'expliquer par des différences méthodologiques comme le travail à des niveaux d'agrégation différents, ou encore le calcul d'une perte par rapport au niveau d'activité du quatrième trimestre 2019, et non d'un contrefactuel, dans la *Note de conjoncture*.

Enfin, les données d'enquête confirment également la robustesse des résultats mis en avant par les modèles. Par exemple, l'Enquête Mensuelle de Conjoncture (EMC) effectuée par la Banque de France vise à évaluer les pertes attribuables à la crise par une méthodologie d'enquête plutôt que par un exercice de simulation et peut constituer un point de comparaison pertinent. Au niveau sectoriel, les pertes d'activité estimées à chaque mois sont proches entre l'enquête et le modèle développé (**Annexe F.2 – Figure F.4**). Le coefficient de corrélation sur l'ensemble des points mensuels des tous les secteurs se situe autour de 0.8, ce qui indique une bonne qualité des résultats du modèle.

2. Discussion de l'évaluation du rôle joué par la crise sur l'activité des entreprises

Si la cohérence avec des sources externes, que celles-ci soient issues de données d'enquête ou des comptes nationaux, rassure sur la qualité de la simulation effectuée, ces comparaisons ne permettent pas de juger de la qualité de l'estimation de l'activité contrefactuelle, puisque celle-ci n'est jamais observée. Une des manières de juger de la qualité de la méthode peut être de faire tourner le modèle déployé sur l'année 2020 sur une année sans crise, par exemple l'année 2019, et de comparer les résultats obtenus dans les simulations, avec les montants de chiffre d'affaires observés. En effet, le

²⁵ La vérification de la cohérence de la simulation avec la valeur ajoutée des comptes nationaux est détaillée en **Annexe F.3**.

modèle sur l'année 2019 permet d'obtenir le montant de chiffre d'affaires réalisé sans crise entre février et décembre 2019 en prenant comme période apprentissage la période allant de janvier 2015 à janvier 2019. Ce montant obtenu est directement comparable au chiffre d'affaires effectivement observé sur l'année 2019. Les 44 équations au niveau *secteur x taille* appliquées à l'année 2019 montrent que la stratégie empirique représente une bonne estimation de l'activité hors-crise. Sur l'ensemble des mois des 44 séries *secteur x taille*, la moyenne de la valeur absolue de l'écart entre le montant simulé et le montant observé est d'environ 2 %²⁶. Dans la majorité des points mensuels, le montant observé est compris dans l'intervalle de confiance au seuil de 95 % du montant prédit (**Annexe F.1 – Figure F.1**).

L'un des apports principaux de l'analyse consiste à descendre jusqu'au niveau individuel de l'entreprise et d'essayer de fournir une estimation des chocs d'activité en fonction du niveau d'activité que les entreprises auraient dû attendre si elles avaient poursuivi leur développement d'avant crise plutôt que de subir cette dernière.

Au niveau individuel, la difficulté de prévision d'un contrefactuel robuste connaît plusieurs causes. D'une part, les montants de chiffre d'affaires réalisés par les entreprises et déclarés dans les déclarations mensuelles de TVA sont beaucoup plus volatiles que les montants agrégés qu'il est possible de calculer au niveau sectoriel par exemple. Cette volatilité touche en particulier les TPE, majoritaires dans l'échantillon et dont l'activité comme la déclaration de cette dernière connaissent des fluctuations importantes d'un mois sur l'autre, sans nécessairement observer de saisonnalité. D'autre part, la formule adoptée pour attribuer une part de marché contrefactuelle s'appuie sur la dynamique observée sur l'année précédant l'exercice de prévision. Or, pour certaines entreprises connaissant potentiellement un rythme de développement non-linéaire, l'exercice de prévision s'avère périlleux puisqu'il n'incorpore la tendance de croissance que sous la manière d'une extrapolation de la tendance observée sur l'année écoulée. La conséquence de l'ensemble de ces éléments est que même en l'absence de crise, la mise en œuvre du modèle de calcul des chocs individuels conduit à l'estimation de chocs non-nécessairement nuls et fluctuant autour de 0.

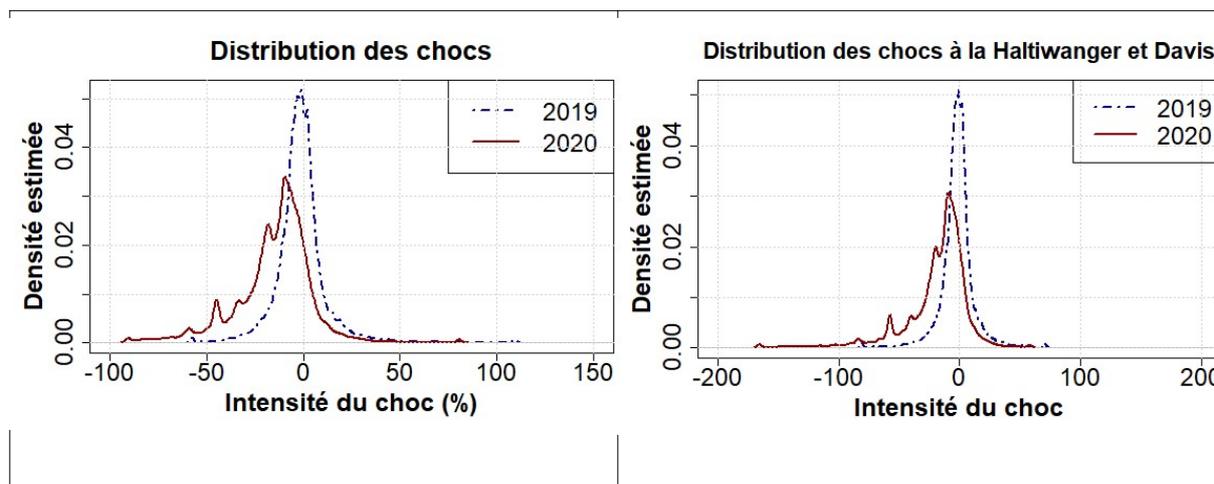
Ainsi, si les chocs individuels restent informatifs sur un écart potentiel de l'activité de chaque entreprise par rapport à un montant cible qui aurait pu être attendu, le réel impact de la crise sur les situations individuelles des entreprises se mesure en comparant les distributions des chocs d'activité individuels en 2020, à cette même distribution des chocs une année sans crise (**Figure 10**).

Dans cette optique, la distribution des chocs individuels obtenus par le modèle en 2019 peut servir de référence. En 2019, la distribution des chocs d'activité est de variance plus faible, tout en étant plus concentrée autour de 0. Par ailleurs, la densité estimée par noyau de la distribution des chocs individuels issus du modèle est symétrique autour de 0 en 2019. Au contraire en 2020 la distribution des

26 Une moyenne nulle des écarts indiquerait un modèle prédisant parfaitement l'ensemble des points 11 points des 44 séries

chocs annuels n'est plus du tout symétrique mais connaît une forte translation vers la gauche, due à une plus forte proportion de chocs négatifs.

Figure 10 – Distributions des chocs d'activités individuels en 2019 et 2020



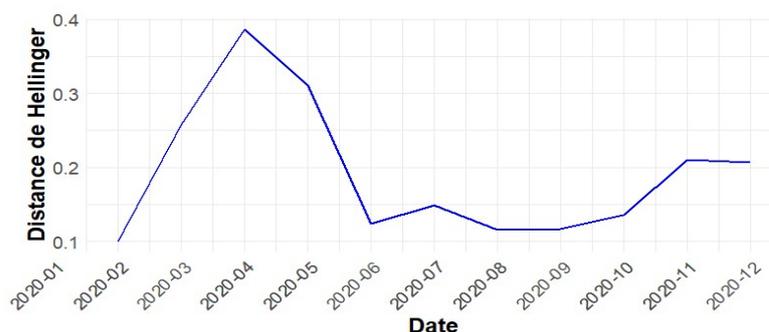
Note de lecture : la densité est estimée par la méthode non-paramétrique de l'estimation par noyau

Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

Ces éléments vont dans le sens d'une bonne robustesse des résultats du modèle, puisqu'en année normale, alors qu'aucune entreprise ne connaît de « choc » exogène dû à la crise, les déviations de l'activité des entreprises par rapport à ce qui pourrait raisonnablement être attendu en appliquant le modèle sont de faible amplitude et proche de zéro.

Une manière de se conforter dans cette opinion peut consister à reproduire cet exercice de comparaison des densités des chocs, non plus à l'échelle annuelle, mais au niveau mensuel. En effet, comme les résultats en ont fait état, les mesures de restriction et donc les difficultés auxquelles ont fait face les entreprises pour se maintenir sur leur trajectoire normale d'activité n'ont pas été constantes au cours de l'année, mais ont dépendu du calendrier des mesures de restriction. Ainsi, il serait raisonnable d'attendre une distorsion plus prononcée lors des mois les plus touchés par la crise, par exemple lors des premiers (mars, avril, mai) et second confinements (novembre, décembre). Les graphiques en [Annexe F.1.b.i](#) et [F.1.b.ii](#) vont dans ce sens. En effet, au mois de février, soit le premier mois modélisé dans l'étude et alors que la crise n'en était qu'à ses premiers soubresauts en 2020, les différences entre les densités estimées en 2020 et en 2019 sont très faibles. Puis, l'intensité de la déformation des chocs individuels semble suivre le calendrier des mesures de restriction en paraissant la plus déformée au mois d'avril, avant de se réduire progressivement pendant l'été et jusqu'au mois d'octobre, moment où le couvre-feu, puis le second confinement furent instaurés ([Annexe F.b](#)).

Figure 11 – Distance de Hellinger entre les distributions mensuelles des chocs individuels entre 2019 et 2020



Note de lecture : La distance de Hellinger entre la densité des chocs estimés pour un même mois en 2019 et 2020 est minimale au mois de février et maximale au mois d’avril. Plus de détails sur cette distance en [Annexe L](#).

Source : Déclarations TVA (DGFIP) ; Calculs des auteurs.

Une manière d’estimer l’intensité de cette déformation d’un mois sur l’autre peut être de calculer la distance de Hellinger entre les distributions des chocs d’activités des mêmes mois en 2019 et 2020²⁷ (**Figure 11**). Cette méthode confirme l’intuition graphique en mettant en lumière une dissimilarité très faible entre 2019 et 2020 pour le premier mois de la crise (février), et connaissant des fortes hausses durant les confinements avec un pic atteint en avril. Autrement dit, plus la conjoncture est affectée, plus le modèle simule une distribution des chocs différente des chocs estimés pendant le même mois d’une année sans crise.

Au contraire, la distance entre les distributions des chocs simulés pour les différents mois de l’année 2019 est faible, quel que soit le mois considéré. En prenant le mois de février 2019 comme distribution de référence, la distance de Hellinger est comprise entre 0,04 en mars et 0,14 en décembre 2019²⁸.

Une dernière manière d’évaluer l’apport du modèle est d’estimer l’amélioration des prévisions liée au calcul du contrefactuel par rapport à une estimation naïve de la crise exprimée en taux de variation. Cette comparaison est difficile à établir sur l’année 2020. Sur l’année 2019 cependant, il est possible de comparer les écarts entre l’observé et les prévisions du modèle en 2019 aux écarts entre l’observé et une reproduction du chiffre d’affaires du même mois de l’année précédente, correspondant à l’approche naïve de l’évaluation de la crise. Au niveau agrégé, le modèle mis en place dans cet article améliore la prévision pour plus de 80 % des mois de l’ensemble des 44 séries *secteur x taille*. Au niveau individuel, l’amélioration est plus limitée puisqu’elle ne concerne que 55 % des mois de l’en-

27 La distance de Hellinger est une mesure de la similarité entre deux distributions statistiques comprise entre 0 et 1. Voir notamment [Nikulin, 2001]

28 Le mois de juillet, un peu particulier du fait des congés du comptable qu’il est difficile de corriger complètement, présente une distance légèrement supérieure avec une valeur de 0,19.

semble des entreprises de l'échantillon. Cependant, pour les mois durant lesquels l'approche naïve est un peu meilleure, l'écart est très faible avec les prévisions du modèle. Les écarts au montant de chiffre d'affaires réalisé sont également plus faibles avec le modèle de prévision, de 11 000 euros par an et par entreprise. Autrement dit, le recours au modèle plutôt qu'à une estimation naïve améliore la prévision effectuée de 11 000 euros par entreprise et par an.

Conclusion

Cette étude s'attache à décrire l'impact de la crise sanitaire sur l'activité des entreprises françaises au cours de l'année 2020. Elle montre notamment que le premier confinement a constitué un choc généralisé dont l'ampleur a varié selon les secteurs, alors que le deuxième confinement a été plus faible et plus localisé à certains secteurs. Si les pertes d'activité sont des éléments fondamentaux pour caractériser les chocs subis par les entreprises, l'évaluation de l'impact de la crise sanitaire sur la situation financière des entreprises ne peut toutefois se résumer à la seule évaluation des chocs d'activité. La situation financière d'une entreprise dépend également des ajustements opérés dans la gestion opérationnelle et financière de la firme (baisse des consommations intermédiaires, de la distribution de dividendes, des dépenses d'investissement, etc.), ainsi que des mesures publiques de soutien, et ce d'autant plus que les montants versés au titre des mesures de soutien ont été importants. Ainsi, malgré les fortes pertes d'activité des entreprises, la perte de revenu national a été supportée en majorité par les administrations publiques [Pointeaux et *al.*, 2021], à hauteur de 70 %, plutôt que par les entreprises et les ménages, du fait du mécanisme des « stabilisateurs automatiques » [Blinder et Solow, 1974] par lequel en période de récession, les prélèvements diminuent et les montants versés de certaines prestations sociales augmentent. En outre, les dispositifs de soutien mis en place pour venir en aide aux entreprises ont permis une atténuation du recul du revenu disponible des entreprises [Dauvin et *al.*, 2020].

Dans ce contexte, un second document de travail [Bureau et *al.*, 2021a] complète la présente analyse en développant un modèle de micro-simulation pour mesurer l'impact de la crise sanitaire sur la situation financière des SNF en 2020. Cette modélisation complémentaire, qui incorpore à la fois l'adaptation des entreprises à leur fluctuation d'activité et le rôle joué par les mesures de soutien, permet d'apprécier la dispersion de l'impact de la crise sanitaire ainsi que l'impact des mesures de soutien sur la situation des plus de 645 000 entreprises étudiées dans ce papier.

Bibliographie

Aghabozorgi S, Shirghorshidi AS, Wah TY , Time-Series Clustering - A Decade Review, *Information Systems*, 53 (2015), 16-38.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199-213). Springer, New York, NY.

Amoureux V., Héam J.-C., Laurent T., **« Les comptes de la Nation en 2020 »**, *Insee Première* n° 1860, mai 2020

Baldwin, R., & Di Mauro, B. W. (2020). Economics in the time of COVID-19: A new eBook. *VOX CEPR Policy Portal*.

Baldwin, R., & Tomiura, E. (2020). Thinking ahead about the trade impact of COVID-19. *Economics in the Time of COVID-19*, 59.

Baleyte J., Bourgeois A., Favetto B., Heam J.-C., Lequien M. , Ralle P. (2021) **« L'économie française en 2020 : une année de bouleversements »**, *Insee Analyses* n° 64, mai 2021

Barrero, J. M., Bloom, N., & Davis, S. J. (2020). *Covid-19 is also a reallocation shock* (No. w27137). National Bureau of Economic Research.

Barrero, J. M., Bloom, N., Davis, S. J., & Meyer, B. H. (2021, May). COVID-19 Is a Persistent Reallocation Shock. In *AEA Papers and Proceedings* (Vol. 111, pp. 287-91).

Bentz, Y., & Merunka, D. (2000). Neural networks and the multinomial logit for brand choice modeling: a hybrid approach. *Journal of Forecasting*, 19(3), 177-200.

Blinder A., Solow R., " Analytical Foundations of Fiscal Policy ", in *The Economics of Public Finance*, Brookings Institution, 1974

Brancati, E., & Brancati, R. (2020). Heterogeneous Shocks in the COVID-19 Pandemic: Panel Evidence from Italian Firms. *Available at SSRN 3597650*.

Bonnet O., Olivia T., Roudil-Valentin T., « En 2020, la chute de la consommation a alimenté l'épargne, faisant progresser notamment les hauts patrimoines », mars 2021

Boone, L., Haugh, D., Pain, N., & Salins, V. (2020). Tackling the fallout from COVID-19. *Economics in the Time of COVID-19*, 37.

Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332), 1509-1526.

Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing* (pp. 227-236). Springer, Berlin, Heidelberg.

Brinca, P., Duarte, J. B., & Faria-e-Castro, M. (2020). Measuring sectoral supply and demand shocks during COVID-19. *Frb st. louis working paper*, (2020-011).

Blanchard, O., Philippon, T. & J. Pisani-Ferry, A New Policy Toolkit Is Needed as Countries Exit COVID-19 Lockdowns, PIIE Policy Brief 20-8, June 2020.

Bureau B., Duquerroy A., Giorgi J., Lé M., Scott S. & Vinas F., (2021a) « L'impact de la crise sanitaire sur la situation financière des entreprises en 2020 : une analyse sur données individuelles », Document de travail, N° 2021-003, *Insee*, juillet 2021

Bureau B., Duquerroy A., Giorgi J., Lé M., Scott S. & Vinas F., (2021b) « L'impact de la crise sanitaire sur la situation financière des entreprises en 2020 : une analyse sur données individuelles », Document de travail, Banque de France

Carnot N., « Comment s'est réparti le coût macroéconomique de la crise sanitaire ? », *Note de blog*, Insee, avril 2021

Clerc, M., Legleye, S., & Nougaret, A. (2021). « Au premier trimestre 2021, 22% des ménages déclarent une baisse de revenus par rapport à mars 2020 », *Insee Focus* n° 238, mai 2021

Cox, D. R., & E. J. Snell. (1989). *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.

Dauvin, M., Ducoudré, B., Heyer, E., Madec, P., Plane, M., & Sampognaro, R. (2020). Évaluation au 11 décembre 2020 de l'impact économique de la pandémie de COVID-19 en France et perspectives pour 2021. *OFCE Policy brief*, 81(11).

Dauvin M., Sampognaro R., « Dans les coulisses du confinement : modélisation de chocs simultanés d'offre et de demande », Sciences Po OFCE *Working Paper* n° 05/2021, février 2021.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227.

Davis, S. J., & Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, 107(3), 819-863.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418), 376-382.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.

Duc C., Souquet C., (2020) « L'impact de la crise sanitaire sur l'organisation et l'activité des sociétés », *Insee Première* n°1830, décembre 2020

Insee, (2020a), « En avril 2020, chute inédite de la confiance des ménages », *Informations Rapides*, n° 2020-106, 28 avril 2021

Insee, (2020b) « Que disent les données de transaction par carte bancaire sur l'évolution récente de la consommation des ménages », *Point de conjoncture*, 7 mai 2020.

Gibbons, S., Overman, H. G., & Pelkonen, P. (2014). Area disparities in Britain: Understanding the contribution of people vs. place through variance decompositions. *Oxford Bulletin of Economics and Statistics*, 76(5), 745-763.

Giorgi J. & Scott S., « Pandémie de COVID-19 et pertes d'activité : le cas des entreprises françaises en 2020 », in *Les entreprises en France*, coll. « Insee Références », novembre 2021

Guerini et al.(2020), «Dynamique des défaillances des entreprises en France et crise de la Covid-19», Policy Brief, n°73, OFCE, juin.

Hadjibeyli B. et al.(2021), «L'effet du COVID et du soutien public sur les entreprises», Direction générale du Trésor

Hale T., Webster S., Petherick A., Phillips T., Kira, B., Oxford COVID-19 Government Response Tracker, Blavatnik School of Government, 2020.

Helpman, E., Itskhoki, O., Muendler, M. A., & Redding, S. J. (2017). Trade and inequality: From theory to estimation. *The Review of Economic Studies*, 84(1), 357-405.

Heyer, É., & Timbeau, X. (2020). Évaluation de la pandémie de Covid-19 sur l'économie mondiale. *Revue de l'OFCE*, 166(2), 59-110.

Jauneau, Y., & Vidalenc, J. (2021). Une photographie du marché du travail en 2020. L'emploi résiste, le halo autour du chômage augmente, Insee Première, n° 1844, Mars 2021

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303

Nikulin, M. S. (2001). Hellinger distance. *Encyclopedia of mathematics*, 78.

Pointeaux A. *et al.*, « Les comptes des administrations publiques en 2020 - Un déficit public historique dans le contexte de la crise sanitaire », *Insee Première* n° 1859, mai 2021.

Pouget J., Simon O., Leclerc H., Poulon P., « Un an après... » , *Note de conjoncture*, mars 2021

Rebecq, A. (2016, October). Icarus: un package r pour le calage sur marges et ses variantes. In *9e colloque francophone sur les sondages*, Gatineau (Canada).

Ripley, B., Venables, W., & Ripley, M. B. (2016). Package 'nnet'. *R package version*, 7(3-12), 700.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.

Sardá-Espinosa, A. *Time-series clustering in r using the dtwclust package. The R Journal* (2019). doi: 10.32614. RJ-2019-023.

Škare, M., Soriano, D. R., & Porada-Rochoń, M. (2021). Impact of COVID-19 on the travel and tourism industry. *Technological Forecasting and Social Change*, 163, 120469.

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Annexes

A. Construction des séries de chiffre d'affaires

La formule d'estimation du chiffre d'affaires à partir des données de déclaration de la Taxe sur la Valeur Ajoutée est la suivante :

$$CA_{i,t} = CAF_{i,t} + CAE_{i,t}$$

$$\Leftrightarrow CA_{i,t} = (BI_{i,t} - AA_{i,t} - AOI_{i,t} + a \cdot b \cdot AONI_{i,t}) + (UE_{i,t} + HUE_{i,t} + a \cdot (1 - b) \cdot AONI_{i,t})$$

Avec a et b fixés par défaut à 1.

Table A.1 : Dictionnaire des variables issues des déclarations de TVA

7*Base imposable hors taxe France (BI)	Opérations réalisées : En France à taux normal 20 % + en France métropolitaine à taux réduit 5,5 % + en France métropolitaine à taux réduit 10 % + dans les DOM à taux normal 8,5 % + dans les DOM à taux réduit 2,1 % + Anciens taux + Opérations imposables à un taux particulier
6*Achats autoliquidés (AA)	Achats de prestations de services intracommunautaires + Importations + Acquisitions intracommunautaires + Livraisons d'électricité, de gaz naturel, de chaleur ou de froid imposables en France + Achats de biens ou de prestations de services réalisés auprès d'un assujetti non établi en France
1*Autres opérations imposables (AOI)	Autres opérations imposables
1*Autres opérations non-imposables (AONI)	Autres opérations non-imposables
2*Exportations vers l'Union Européenne (UE)	Livraisons intracommunautaires à destination d'une personne assujettie – Ventes BtoB +Livraisons d'électricité, de gaz naturel, de chaleur ou de froid non imposables en France
1*Exportations hors de l'Union Européenne (HUE)	Exportation hors UE

B. Correction des données brutes

1. Reports de déclarations

Les données de TVA présentent parfois des reports de déclaration d'un mois sur l'autre. Dans ces cas-là, les déclarations de TVA comportent une déclaration nulle un mois donné suivie d'une déclaration correspondant à l'activité de deux mois le mois suivant. Ces reports peuvent correspondre par exemple à une absence du comptable de l'entreprise un mois donné. Ce phénomène, dit de « congé du comptable » intervient particulièrement les mois d'été. Un lissage visant à répartir l'activité déclarée du mois de « rattrapage » entre le mois de déclaration nulle et le mois de rattrapage a été mis en place. Pour cela, les reports de déclarations sont d'abord identifiés comme étant les mois pour lesquels une déclaration nulle est située entre deux déclarations strictement positives, et pour lesquels la déclaration du mois suivant la déclaration nulle est au moins égale à 2 fois 70 % du montant de chiffre d'affaires déclaré lors du mois précédant la déclaration nulle :

$$D_{i,t} = 1_{(CA_{i,t}=0) \cap (CA_{i,t-1} \neq 0) \cap (CA_{i,t+1} \neq 0) \cap (CA_{i,t+1} > [2 \cdot (CA_{i,t-1} - (30\% \cdot CA_{i,t-1}))])}, \forall i, \forall t$$

La variable $D_{i,t}$ permet d'identifier les dates t à corriger pour chaque entreprise i . Les corrections des chiffres d'affaires mensuels de l'ensemble des entreprises est donc corrigé avec la formule suivante :

$$\begin{cases} CA'_{i,t-1} = CA_{i,t-1} \\ CA'_{i,t} = 1_{[D_{i,t}=0]} CA_{i,t} + 1_{[D_{i,t}=1]} \frac{CA_{i,t+1} + CA_{i,t-1}}{2} \\ CA'_{i,t+1} = 1_{[D_{i,t}=0]} CA_{i,t+1} + 1_{[D_{i,t}=1]} (CA_{i,t+1} - CA'_{i,t}) \end{cases}$$

2. Valeurs aberrantes

Le second problème lié au caractère déclaratif des données de TVA concerne les valeurs aberrantes pouvant se glisser dans les séries du chiffre d'affaires à la suite d'une erreur de report. Au-dessous d'un certain niveau de CA, il est particulièrement complexe d'identifier avec précision les valeurs aberrantes étant liées à une erreur de report de celles liées à la réalisation d'un mois exceptionnel. Afin d'assurer une correction adaptée, l'identification des valeurs aberrantes a été effectuée en considérant comme telles les valeurs de chiffre d'affaires mensuelles supérieures à un million d'euros et au moins 3 fois supérieures au montant déclaré le mois précédant et 3 fois supérieur à la déclaration du même mois de l'année précédente :

$$D_{i,t} = 1_{(CA_{i,t} > 10^6) \cap (CA_{i,t} > 3 \cdot CA_{i,t-1}) \cap (CA_{i,t} > 3 \cdot CA_{i,t-12})}$$

Ces valeurs sont ensuite corrigées en les rapportant dans la tendance de la série. Pour le mois d'avril, la formule de lissage a été adaptée. En effet, le mois d'avril 2020 ayant été le mois le plus fortement affecté par la crise en étant le premier mois complètement couvert par une période de confinement, il n'était pas satisfaisant de corriger une valeur aberrante ce mois-ci avec la même formule que pour les mois précédents.

La formule de lissage générale est la suivante :

$$\forall i, \forall t \neq \text{avril}_{2020}, CA'_{i,t} = 1_{[D_{i,t}=0]} CA_{i,t} + 1_{[D_{i,t}=1]} \frac{CA_{i,t-1} - CA_{i,t+1}}{2}$$

Pour le mois d'avril, le lissage a été calculé en appliquant au mois d'avril 2019, un taux de croissance égal à la moitié du taux de croissance observé entre mars 2019 et mars 2020²⁹ :

$$\forall i, CA'_{i,\text{avril}2020} = 1_{[D_{i,\text{avril}2020}=0]} CA_{i,\text{avril}2020} + 1_{[D_{i,\text{avril}2020}=1]} CA_{i,\text{avril}2019} * \frac{\left[1 + \frac{CA_{i,\text{mars}2020} - CA_{i,\text{mars}2019}}{CA_{i,\text{mars}2019}} \right]}{2}$$

3. Déclarations irrégulières

Enfin, certaines entreprises déclarent leur TVA de manière très irrégulière, malgré leur souscription au régime de déclaration mensuel. En particulier, certaines déclarent une activité nulle sur une majorité de mois de l'année ce qui rend difficile la modélisation de leurs trajectoires individuelles dans la suite des travaux. Pour assurer une estimation robuste de la situation individuelle des entreprises, les entreprises ayant déclaré une activité nulle au mois 9 mois sur la période février 2019 – février 2020, dont au moins deux déclarations nulles durant les 3 mois précédant février 2020, ont été supprimées de l'échantillon.

C. Composition de l'échantillon

Tableau C.1 - Ventilation des effectifs par secteur d'activité

	Échantillon	Fare – Champ d'étude
Agriculture	0.0%	0.0%
Énergie, industries extractives	1.7%	2.9%
Agroalimentaire	4.1%	4.1%
Cokéfaction et raffinage	0.2%	0.1%
Équipements électroniques	3.5%	3.2%
Matériels de transport	3.2%	3.1%
Autres produits industriels	10.5%	10.2%
Construction	9.3%	8.2%

29 L'hypothèse sous-jacente est que le mois d'avril a été plus touché que le mois de mars 2020 par la crise dans la mesure où seules deux semaines du mois de mars ont été concernées par des mesures de confinement contre l'ensemble du mois d'avril. Ainsi, un taux de croissance de 0,8 entre mars 2019 et mars 2020 conduit à lisser une éventuelle valeur aberrante du mois d'avril avec un taux de croissance de 0,4 par rapport à avril 2019.

Commerce	21.6%	20.0%
Transports et entreposage	9.9%	9.4%
Hébergement et restauration	5.6%	5.5%
Information et communication	4.8%	5.7%
Activités immobilières	1.2%	1.4%
Activités scientifiques et techniques	19.5%	20.9%
Santé	2.6%	3.4%
Autres services	2.2%	2.0%

Tableau C.2 - Ventilation des effectifs par taille d'entreprises

	Échantillon	Fare – Champ d'étude
ETI-GE	57.1%	63.6%
PME	29.9%	26.2%
TPE	13.0%	10.2%

D. Modèles de séries temporelles

1. Stationnarité des séries transformées

Dans l'analyse de séries temporelles et les modèles de prévision, la stationnarité est une propriété centrale. La stationnarité signifie que les propriétés statistiques d'une série ne varient pas dans le temps. Plus formellement, elle permet de vérifier que la distribution jointe des valeurs de la série pour les indices de temps $1 \leq t_1 < \dots < t_k$ est identique à la distribution jointe des valeurs de la série pour les indices de temps $1 \leq t_{1+h} < \dots < t_{k+h}, \forall h > 1$. Cette propriété permet de vérifier que la série étudiée est identiquement distribuée, mais aussi que la distribution jointe de toute suite de termes adjacents dans cette série est identique pour toutes les périodes de temps. La vérification de cette propriété permet de mettre en place des modèles de prévision statistiquement robustes.

Les séries de chiffre d'affaires $Y_{t,s}$ sont donc transformées afin de pouvoir les stationnariser. La transformation utilisée est la suivante :

$$X_{t,s} = (1 - B^{12}) \log(Y_{t,s})$$

Les tests de Dickey-Fuller augmentés visant à vérifier la stationnarité considèrent, pour chaque série

$X_{t,s}$ étudiée, un modèle auto-régressif d'ordre p avec une tendance temporelle :

$$X_{t,s} = \beta_1 + \beta_2 t + \varphi_1 X_{t-1,s} + \dots + \varphi_p X_{t-p,s} + \varepsilon_{t,s}$$

En prenant la différence première de ces équations, ces dernières se réécrivent :

$$\Delta X_{t,s} = \beta_1 + \beta_2 t + \gamma X_{t-1,s} + \delta_1 \Delta X_{t-1,s} + \dots + \delta_p \Delta X_{t-p+1,s} + z_t$$

Le test consiste à tester $H_0: \gamma=0$ contre $H_0: \gamma < 0$. L'hypothèse nulle est que la série n'est pas stationnaire. Le test a été effectué avec différentes valeurs de p et en incluant et supprimant les termes de dérive β_1 et de tendance $\beta_2 t$.

Le Tableau C.1 indique les p-valeurs pour 3 tests de Dickey-Fuller par série *secteur x taille* transformée en posant $p_s=0$: sans dérive ni tendance, avec dérive et sans tendance et avec dérive et tendance. Les résultats indiquent que les séries transformées sont stationnaires.

Dans la suite de l'étude, l'hypothèse que les séries des montants de chiffre d'affaires prises en logarithme et désaisonnalisées par un terme de retard d'ordre 12 sont stationnaires. C'est à partir de ces données que la modélisation et la prévision sont effectuées.

Tableau D.1 – P-valeur des tests de Dickey-Fuller par série secteur x taille

Groupe secteur/taille	Type de test de Dickey-Fuller		
	sans dérive ni tendance	Avec dérive et sans tendance	Avec dérive et tendance
AZ - Agri/Peche/TPE	0.01***	0.01***	0.01***
AZ - Agri/Peche/Plus de 10 salariés	0.01***	0.01***	0.01***
C1 - Fabrication alimentaire/ETI-GE	0.01***	0.01***	0.01***
C1 - Fabrication alimentaire/TPE	0.06*	0.01***	0.01***
C1 - Fabrication alimentaire/PME	0.01***	0.01***	0.01***
C2 - Cokefaction/Raffinage	0.02**	0.22	0.47**
C3 - Fabrication électronique/ETI-GE	0.01***	0.01***	0.01***
C3 - Fabrication électronique/TPE	0.01***	0.01***	0.01***
C3 - Fabrication électronique/PME	0.01***	0.01***	0.01***
C4 - Fabrication matériel de transport/ETI-GE	0.01***	0.01***	0.01***
C4 - Fabrication matériel de transport/TPE	0.02**	0.01***	0.01***
C4 - Fabrication matériel de transport/PME	0.01***	0.01***	0.01***
C5 - Autres produits industriels/ETI-GE	0.01***	0.01***	0.01***

C5 - Autres produits in- dus/TPE	0.01***	0.01***	0.01***
C5 - Autres produits in- dus/PME	0.01***	0.01***	0.01***
DE - Extraction + Energie/ETI-GE	0.02**	0.01***	0.05*
DE - Extraction + Energie/TPE	0.01***	0.01***	0.01***
DE - Extraction + Energie/PME	0.01***	0.01***	0.01***
FZ - Construction/ETI-GE	0.01***	0.01***	0.01***
FZ - Construction/TPE	0.01***	0.01***	0.01***
FZ - Construction/PME	0.01***	0.03**	0.07*
GZ - Commerce/ETI-GE	0.01***	0.01***	0.01***
GZ - Commerce/TPE	0.07*	0.01***	0.01***
GZ - Commerce/PME	0.02**	0.01***	0.01***
HZ - Transport/ETI-GE	0.01***	0.01***	0.01***
HZ - Transport/TPE	0.01***	0.01***	0.01
HZ - Transport/PME	0.01***	0.01***	0.01***
IZ - Restaurants / Hotels/ ETI-GE	0.01***	0.01***	0.01***
IZ - Restaurants / Hotels/ TPE	0.08*	0.08*	0.03**
IZ - Restaurants / Hotels/ PME	0.24	0.04**	0.04**
JZ - TIC/ETI-GE	0.01***	0.01***	0.01***
JZ - TIC/TPE	0.1	0.01***	0.01***
JZ - TIC/PME	0.01***	0.01***	0.01***
LZ - Immobilier/ETI-GE	0.01***	0.01***	0.01***
LZ - Immobilier/TPE	0.02**	0.01***	0.01***
LZ - Immobilier/PME	0.01***	0.01***	0.02**
MN - Scientifique / Tech- nique/ETI-GE	0.01***	0.01***	0.01***
MN - Scientifique / Tech- nique/TPE	0.06*	0.01***	0.01***
MN - Scientifique / Tech- nique/PME	0.01***	0.01***	0.01***
OQ - Santé/TPE	0.01***	0.01***	0.01***
OQ - Santé/Plus de 10 sa- lariés	0.01***	0.01***	0.01***

RU - Autres Services/ETI-GE	0.01***	0.01***	0.01***
RU - Autres Services/TPE	0.04**	0.01***	0.01***
RU - Autres Services/PME	0.01***	0.01***	0.01***

Note de lecture : Significativité des coefficients : 1 % : '***' ; 5 % : '**' ; 10 % : '*'

2. Normalité des résidus

La normalité des résidus a été testée en évaluant la qualité de l'ajustement de la distribution des résidus des équations des séries *secteur x taille* avec la distribution de la loi normale centrée réduite. Le

R^2 de la régression linéaire des quantiles théoriques de la loi normale centrée réduite sur les quantiles empiriques des résidus des équations permet d'estimer la qualité de cet ajustement. Pour l'ensemble des équations, la qualité de l'ajustement est très satisfaisante, avec des valeurs souvent supérieures à 0,95. La seule exception concerne les TPE du secteur de la santé pour lesquelles le R^2 vaut seulement 0,24.

Tableau D.2 : Qualité de l'ajustement de la distribution des résidus à la distribution de la loi normale

Groupe secteur/taille	R^2 Quantiles théoriques – Quantiles observés
AZ - Agri/Peche/TPE	0.989
AZ - Agri/Peche/Plus de 10 salariés	0.98
C1 - Fabrication alimentaire/ETI-GE	0.965
C1 - Fabrication alimentaire/TPE	0.982
C1 - Fabrication alimentaire/PME	0.986
C2 - Cokefaction/Raffinage	0.985
C3 - Fabrication électronique/ETI-GE	0.883
C3 - Fabrication électronique/TPE	0.989
C3 - Fabrication électronique/PME	0.983
C4 - Fabrication matériel de transport/ETI-GE	0.986
C4 - Fabrication matériel de transport/TPE	0.983
C4 - Fabrication matériel de transport/PME	0.967
C5 - Autres produits indus/ETI-GE	0.99
C5 - Autres produits indus/TPE	0.974
C5 - Autres produits indus/PME	0.988
DE - Extraction + Energie/ETI-GE	0.978
DE - Extraction + Energie/TPE	0.974
DE - Extraction + Energie/PME	0.977

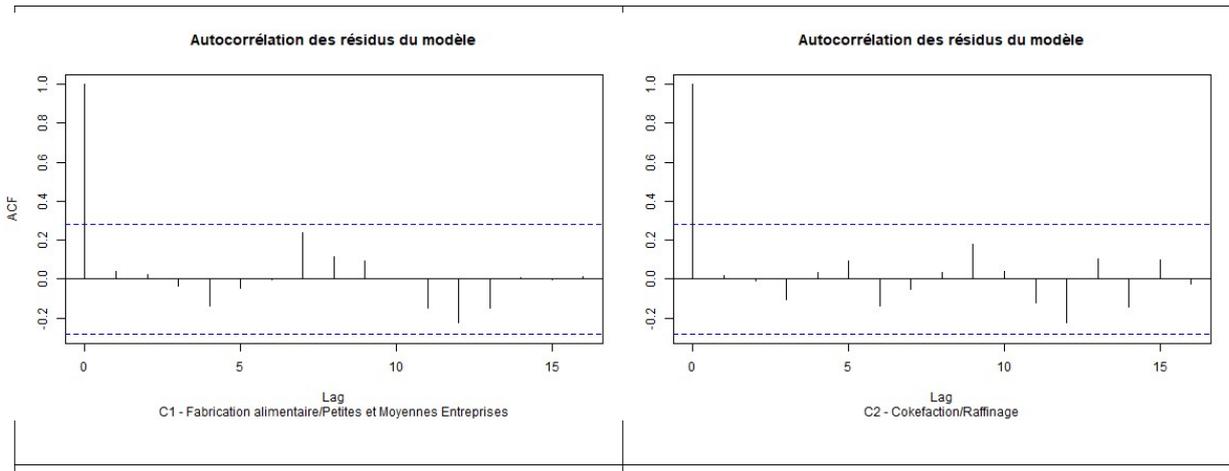
FZ - Construction/ETI-GE	0.99
FZ - Construction/TPE	0.981
FZ - Construction/PME	0.838
GZ - Commerce/ETI-GE	0.975
GZ - Commerce/TPE	0.977
GZ - Commerce/PME	0.989
HZ - Transport/ETI-GE	0.934
HZ - Transport/TPE	0.986
HZ - Transport/PME	0.971
IZ - Restaurants / Hotels/ETI-GE	0.977
IZ - Restaurants / Hotels/TPE	0.783
IZ - Restaurants / Hotels/PME	0.992
JZ - TIC/ETI-GE	0.957
JZ - TIC/TPE	0.994
JZ - TIC/PME	0.987
LZ - Immobilier/ETI-GE	0.983
LZ - Immobilier/TPE	0.976
LZ - Immobilier/PME	0.968
MN - Scientifique / Technique/ETI-GE	0.341
MN - Scientifique / Technique/TPE	0.987
MN - Scientifique / Technique/PME	0.976
OQ - Santé/TPE	0.24
OQ - Santé/Plus de 10 salariés	0.97
RU - Autres Services/ETI-GE	0.974
RU - Autres Services/TPE	0.941
RU - Autres Services/PME	0.991

3. Autocorrélation des résidus

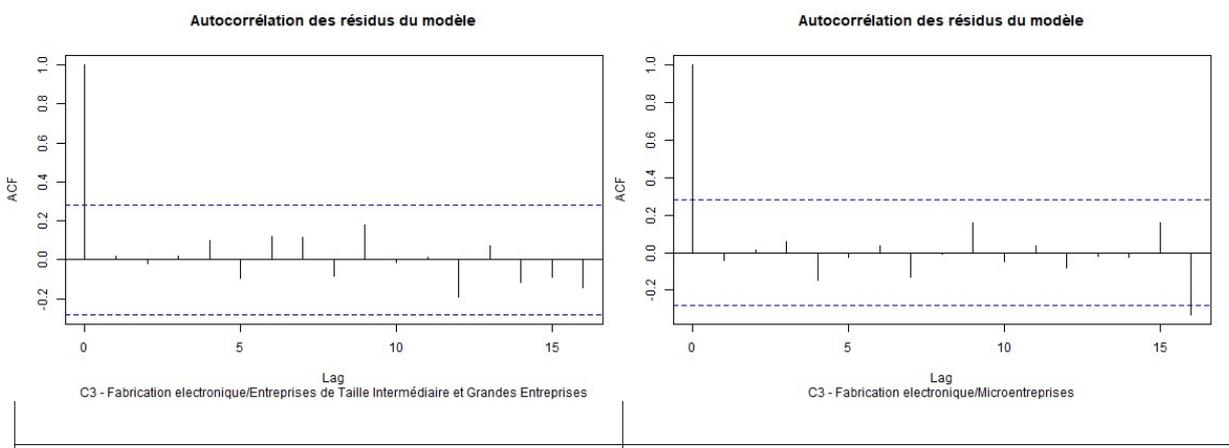
Les tests d'autocorrélations des résidus montrent bien que les résidus des équations *secteur x taille* ne sont pas significativement auto-corrélés.

Tableau D.3 – Autocorrélation des résidus par série secteur x taille

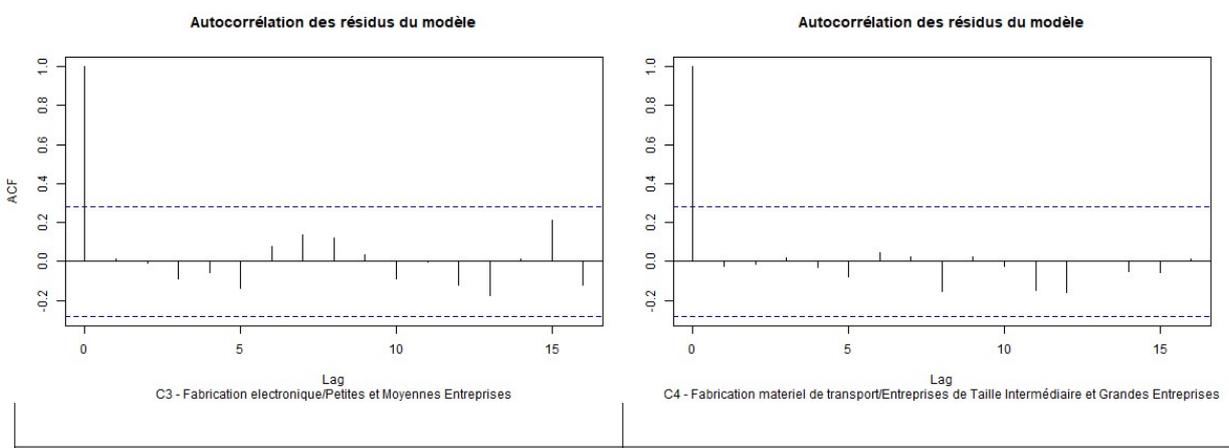
<p>AZ – Agri/Peche/TPE</p> <p style="text-align: center;">Autocorrélation des résidus du modèle</p> <p style="text-align: center;">Lag AZ - Agri/Peche/Microentreprises</p>	<p>AZ – Agri/Peche/Plus de 10 salariés</p> <p style="text-align: center;">Autocorrélation des résidus du modèle</p> <p style="text-align: center;">Lag AZ - Agri/Peche/Plus de 10 salariés</p>
<p>C1 – Fabrication alimentaire/ETI-GE</p> <p style="text-align: center;">Autocorrélation des résidus du modèle</p> <p style="text-align: center;">Lag C1 - Fabrication alimentaire/Entreprises de Taille Intermédiaire et Grandes Entreprises</p>	<p>C1 – Fabrication alimentaire/TPE</p> <p style="text-align: center;">Autocorrélation des résidus du modèle</p> <p style="text-align: center;">Lag C1 - Fabrication alimentaire/Microentreprises</p>
<p>C1 – Fabrication alimentaire/PME</p>	<p>C2 – Cokefaction/Raffinage</p>



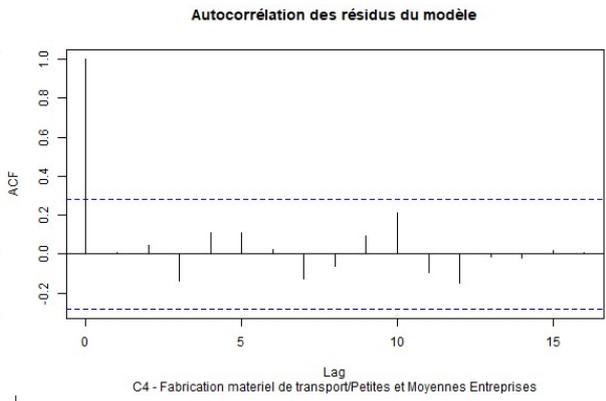
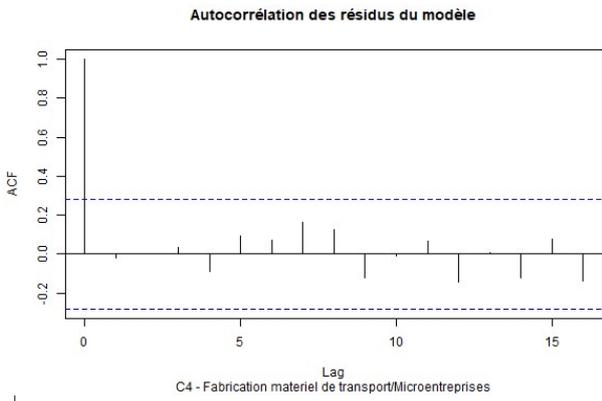
C3 – Fabrication électronique/ETI-GE	C3 – Fabrication électronique/TPE
---	--



C3 – Fabrication électronique/PME	C4 – Fabrication matériel de transport/ETI-GE
--	--

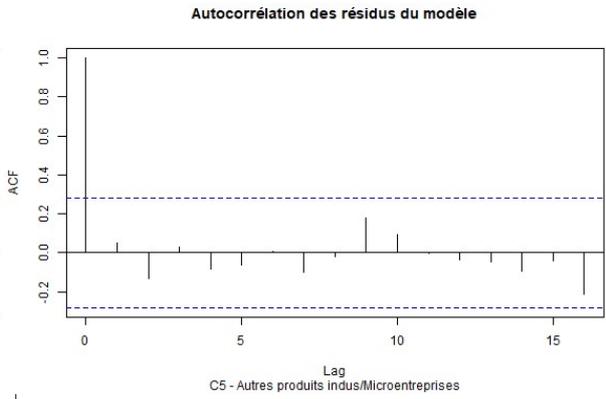
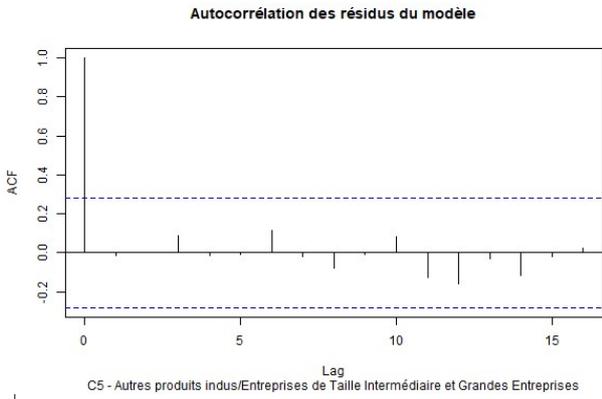


C4 – Fabrication matériel de transport/TPE	C4 – Fabrication matériel de transport/PME
---	---



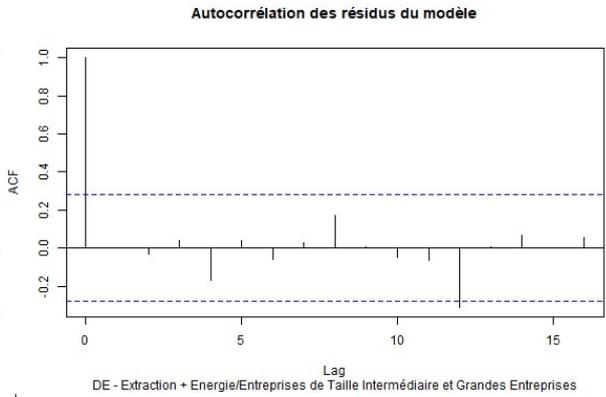
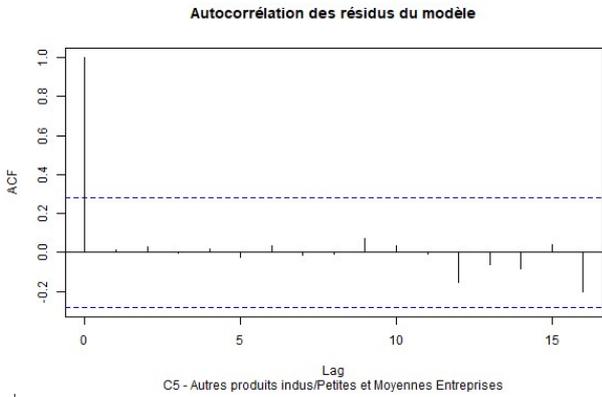
C5 – Autres produits indus/ETI-GE

C5 – Autres produits indus/TPE



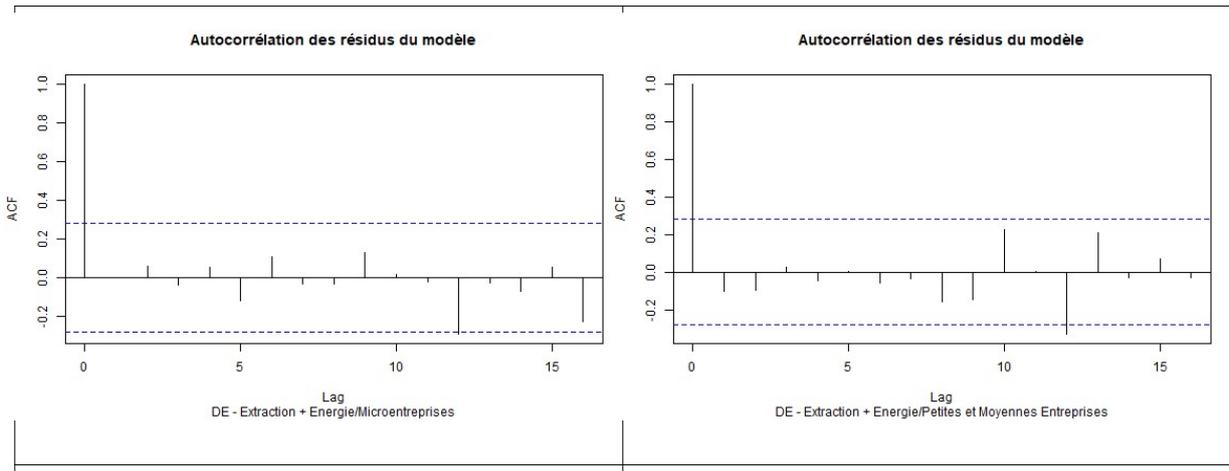
C5 – Autres produits indus/PME

DE – Extraction + Energie/ETI-GE



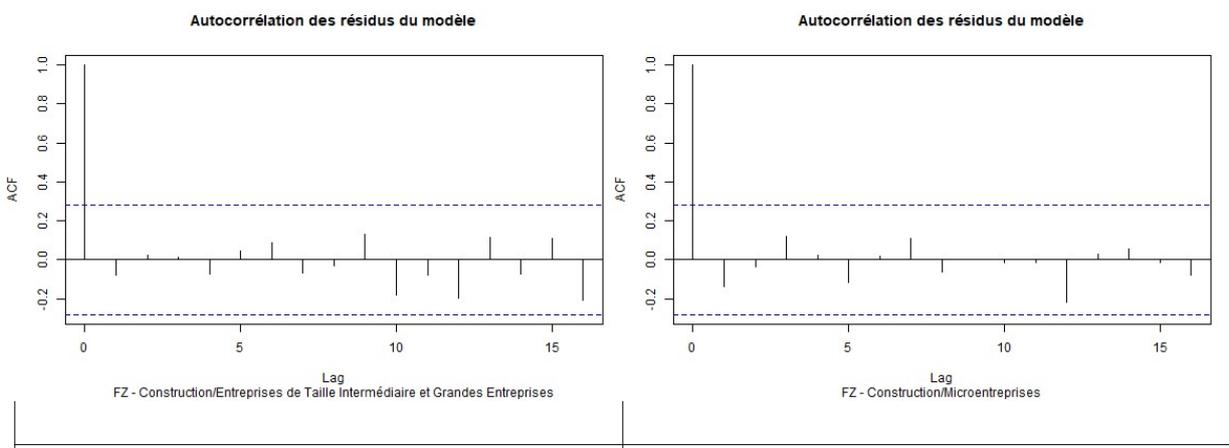
DE – Extraction + Energie/TPE

DE – Extraction + Energie/PME



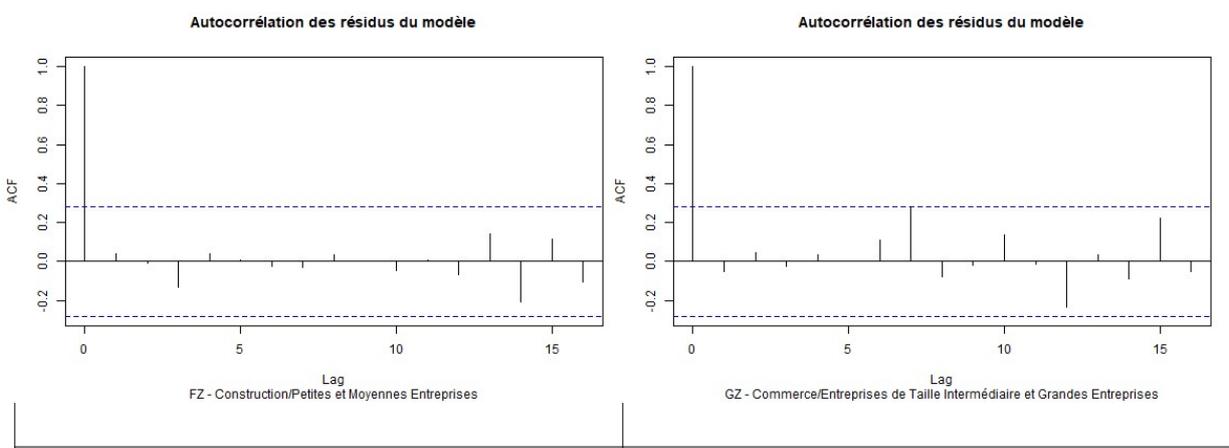
FZ – Construction/ETI-GE

FZ – Construction/TPE



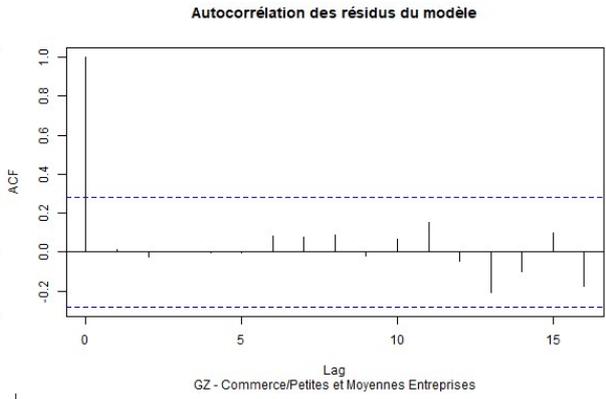
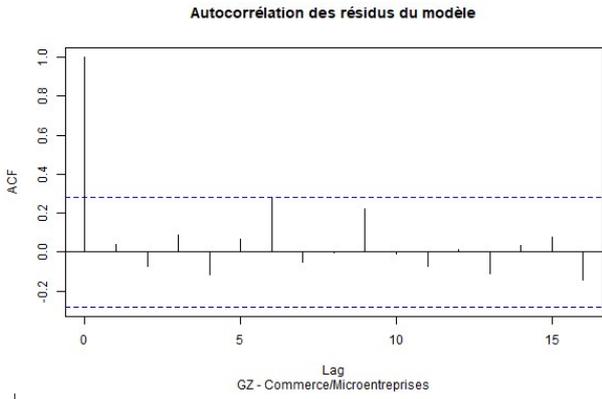
FZ – Construction/PME

GZ – Commerce/ETI-GE



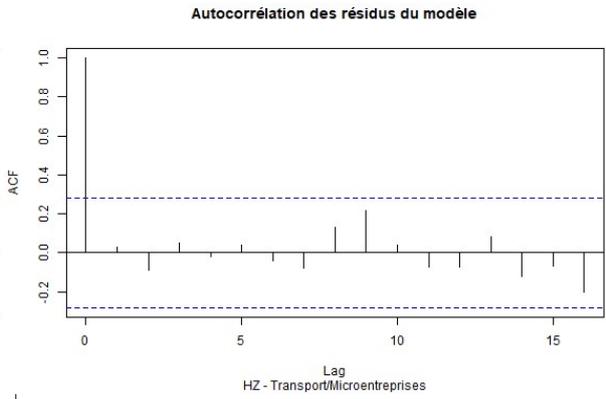
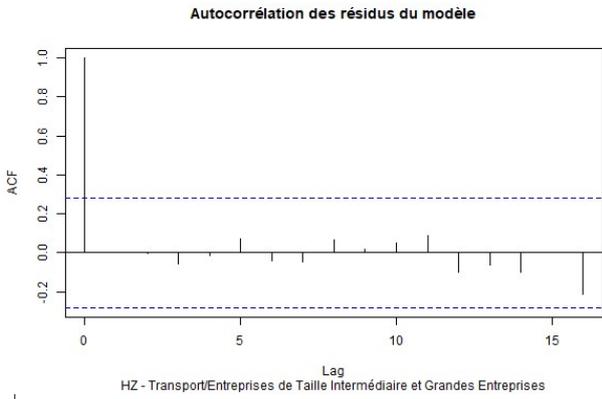
GZ – Commerce/TPE

GZ – Commerce/PME



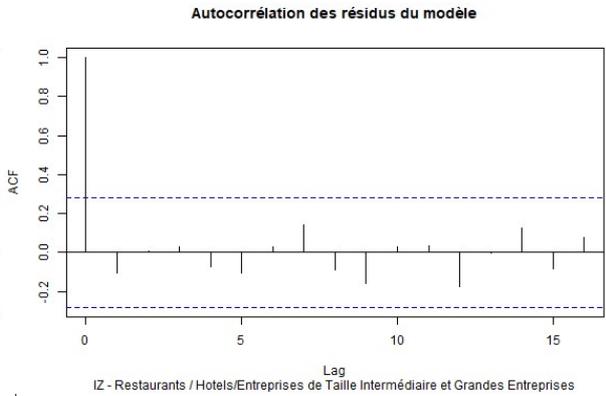
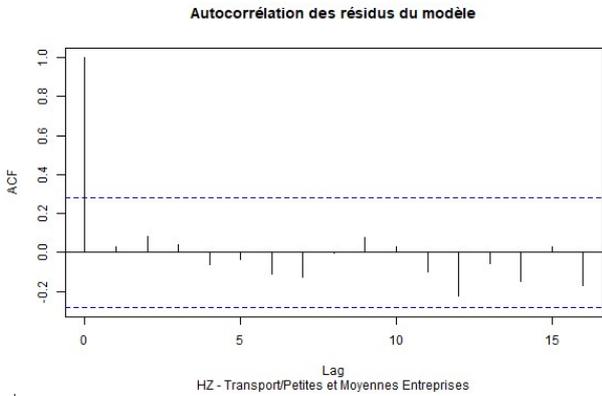
HZ – Transport/ETI-GE

HZ – Transport/TPE



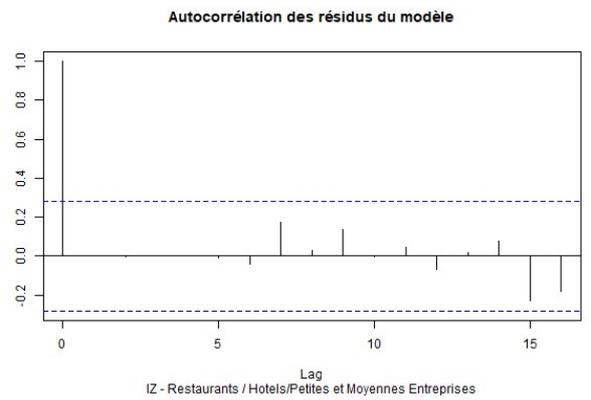
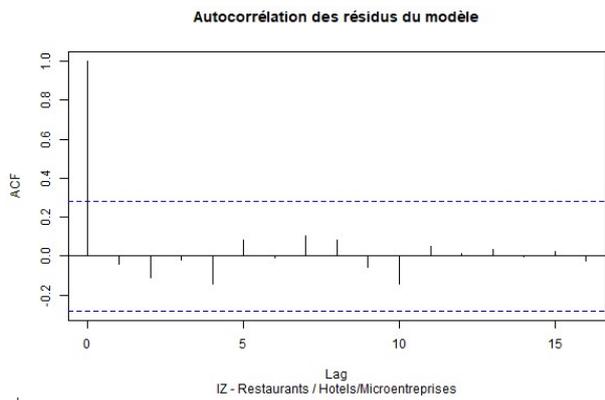
HZ – Transport/PME

IZ – Restaurants / Hotels/ETI-GE



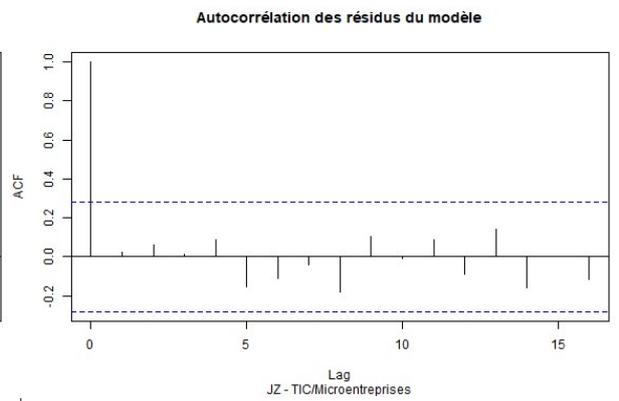
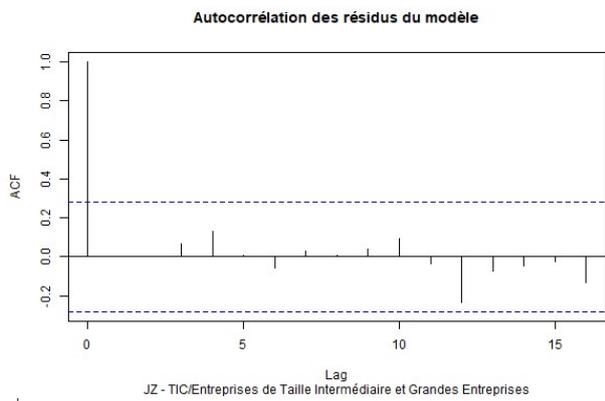
IZ – Restaurants / Hotels/TPE

IZ – Restaurants / Hotels/PME



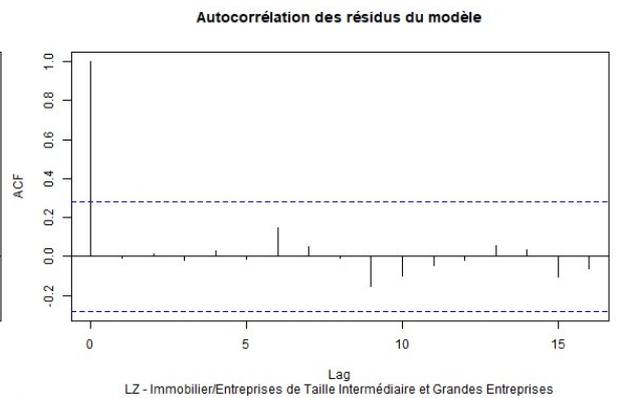
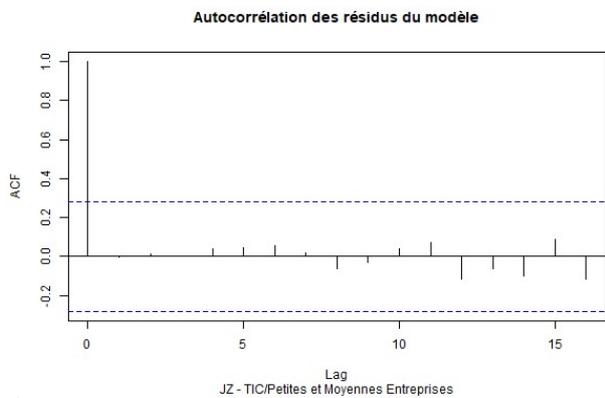
JZ – TIC/ETI-GE

JZ – TIC/TPE



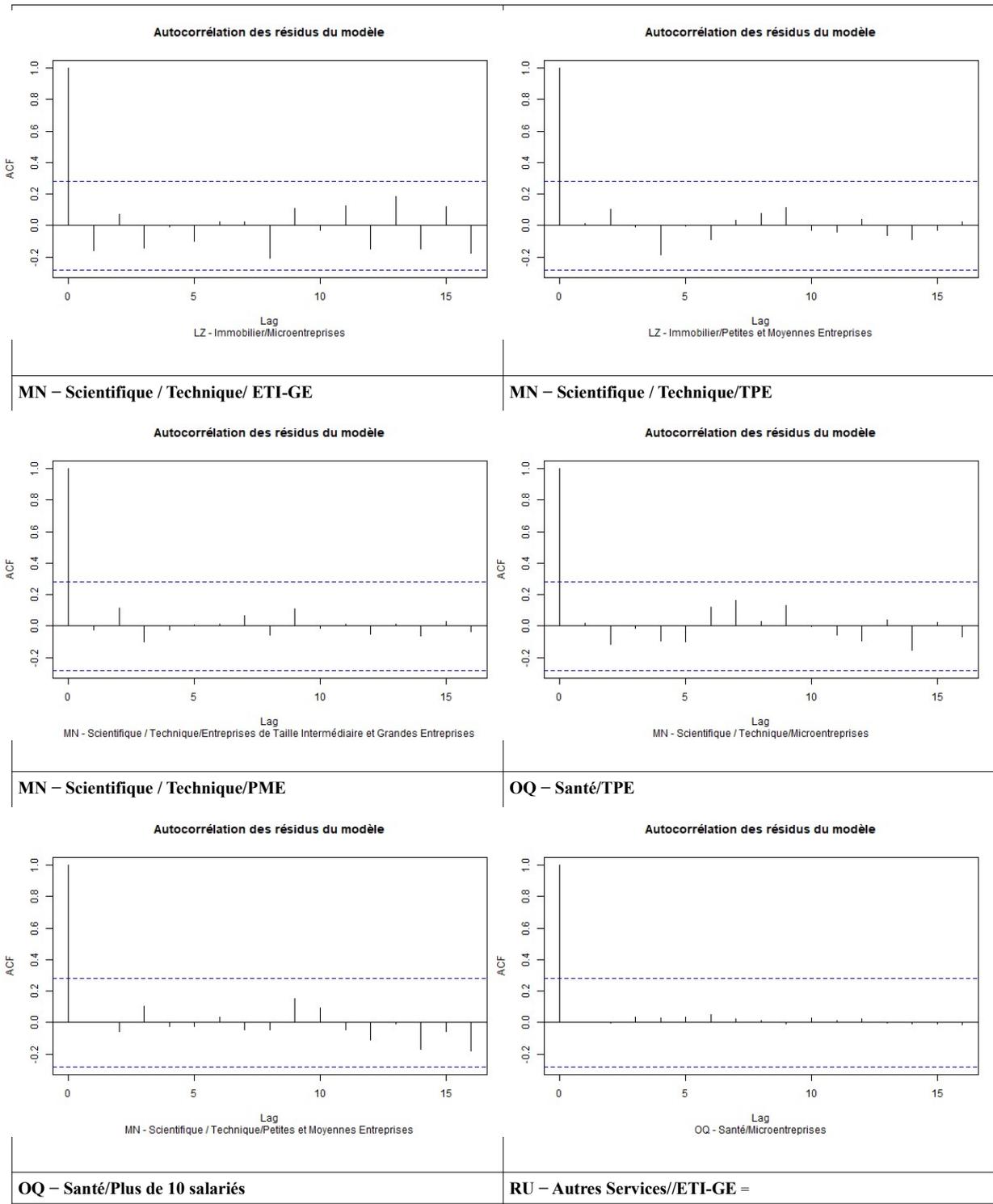
JZ – TIC/PME

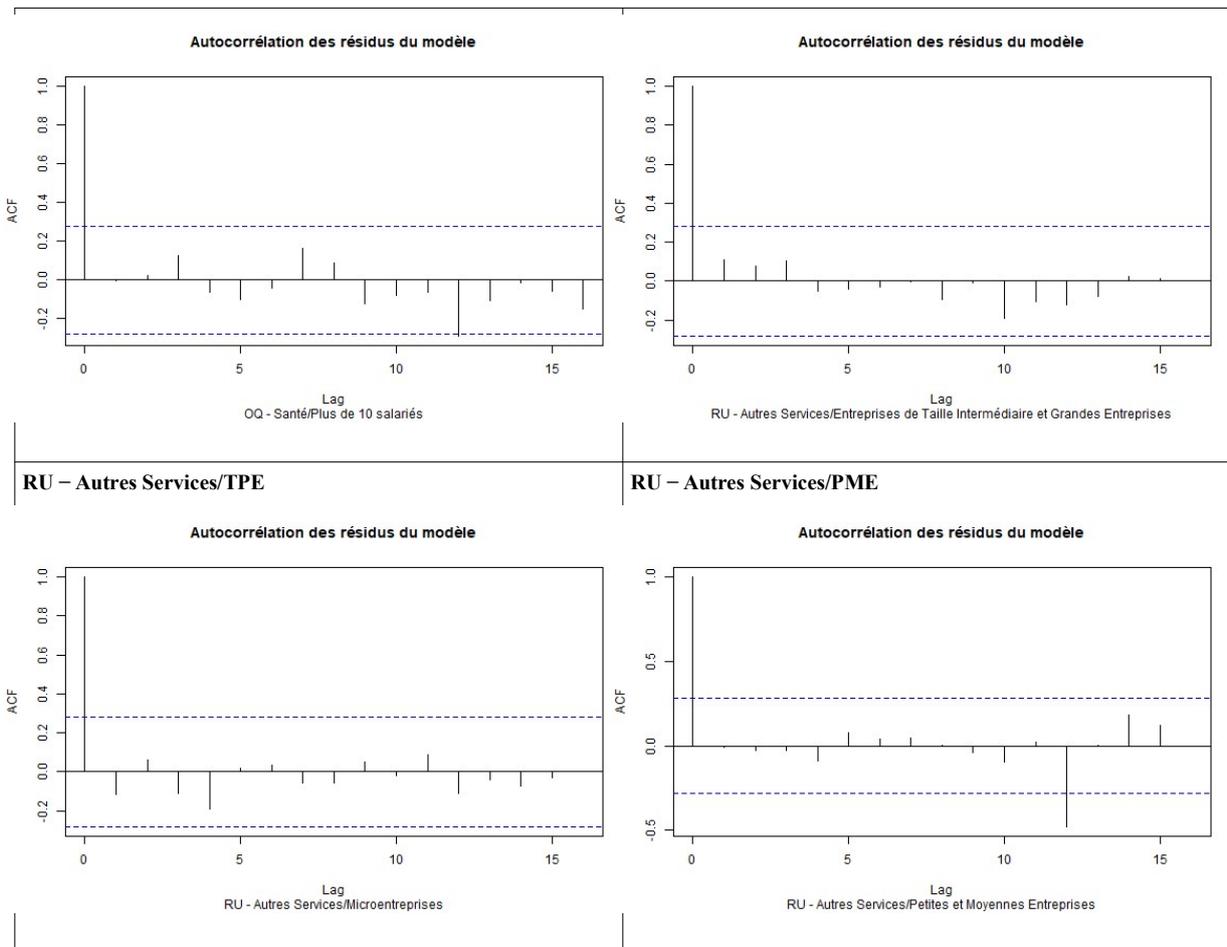
LZ – Immobilier/ETI-GE



LZ – Immobilier/TPE

LZ – Immobilier/PME





4. Blancher des résidus

La blancheur des résidus est testée grâce à un test de Box-Pierce. L'hypothèse nulle de ce test est l'indépendance temporelle des résidus. Les p-valeurs des différentes équations sont toutes largement au-dessus de 0,5 (Tableau D.4). Ce résultat s'interprète comme une preuve robuste de l'indépendance des résidus.

Tableau D.4 – P-valeur du test de Box-Pierce par série secteur x taille

Groupe secteur/taille	p-value-Box-Pierce test
AZ - Agri/Peche/TPE	0.878
AZ - Agri/Peche/Plus de 10 salariés	0.697
C1 - Fabrication alimentaire/ETI-GE	0.972
C1 - Fabrication alimentaire/TPE	0.861
C1 - Fabrication alimentaire/PME	0.763
C2 - Cokefaction/Raffinage	0.892
C3 - Fabrication électronique/ETI-GE	0.883

C3 - Fabrication electronique/TPE	0.759
C3 - Fabrication electronique/PME	0.919
C4 - Fabrication materiel de transport/ETI-GE	0.852
C4 - Fabrication materiel de transport/TPE	0.885
C4 - Fabrication materiel de transport/PME	0.961
C5 - Autres produits indus/ETI-GE	0.929
C5 - Autres produits indus/TPE	0.705
C5 - Autres produits indus/PME	0.928
DE - Extraction + Energie/ETI-GE	0.993
DE - Extraction + Energie/TPE	0.987
DE - Extraction + Energie/PME	0.456
FZ - Construction/ETI-GE	0.575
FZ - Construction/TPE	0.314
FZ - Construction/PME	0.769
GZ - Commerce/ETI-GE	0.697
GZ - Commerce/TPE	0.784
GZ - Commerce/PME	0.942
HZ - Transport/ETI-GE	0.997
HZ - Transport/TPE	0.823
HZ - Transport/PME	0.842
IZ - Restaurants / Hotels/ETI-GE	0.455
IZ - Restaurants / Hotels/TPE	0.765
IZ - Restaurants / Hotels/PME	0.98
JZ - TIC/ETI-GE	0.984
JZ - TIC/TPE	0.865
JZ - TIC/PME	0.987
LZ - Immobilier/ETI-GE	0.962
LZ - Immobilier/TPE	0.253
LZ - Immobilier/PME	0.939
MN - Scientifique / Technique/ETI-GE	0.851
MN - Scientifique / Technique/TPE	0.882
MN - Scientifique / Technique/PME	0.977
OQ - Santé/TPE	0.998
OQ - Santé/Plus de 10 salariés	0.986
RU - Autres Services/ETI-GE	0.432
RU - Autres Services/TPE	0.41
RU - Autres Services/PME	0.936

5. Significativité des coefficients

La significativité des coefficients associés aux variables de chacune des équations est estimée avec un test Z. Ainsi, la p-valeur de chaque coefficient est donnée par :

$$p = 1 - 2 * \Phi \left(\left| \frac{\hat{\phi}_i}{\sqrt{\hat{V}(\hat{\phi}_i)}} \right| \right)$$

Avec $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite et $\hat{\phi}_i$ le coefficient estimé par maximum de vraisemblance.

Les coefficients de l'ensemble des équations présentent une significativité statistique. La majorité des variables étant significatives au seuil de 1 %.

Tableau D.5 – Coefficients estimés par série secteur x taille

AZ – Agri/Pêche/TPE					AZ – Agri/Pêche/Plus de 10 salariés					C1 – Fabrication alimentaire/ETI-GE					
	Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)	
ar1	-0.092848	0.160178	-0.5797	0.5621483	ar1	1.906758	0.0120171	158.6704	< 2.2e-16	ar1	0.4795105	0.1512946	3.1694	0.001528	
ma1	0.512606	0.132968	3.8551	0.0001157	***					**					
***					ar2	-0.9977808	0.0072929	-136.8152	< 2.2e-16	***	0.3714110	0.1349141	2.7529	0.005906	
ma2	0.347856	0.082873	4.1975	2.699e-05	***					**					
***					ma1	-1.8931964	0.0978072	-19.3564	< 2.2e-16	***	0.8236353	0.0361042	22.8127	< 2.2e-16	
ma3	0.512554	0.143343	3.5757	0.0003493	***					***					
***					ma2	0.9990032	0.1015204	9.8404	< 2.2e-16	***	-0.7642770	0.1172213	-6.5199	7.033e-11	
ma4	0.999959	0.174313	5.7366	9.662e-09	***					***					
***					intercept	0.0284855	0.0086729	3.2844	0.001022	***	-0.2539645	0.1566904	-1.6208	0.105060	
intercept	0.050915	0.019214	2.6499	0.0080510	**					***	-0.7357831	0.137523	-5.3502	8.784e-08	
**										***					
										***	-0.0983585	0.112616	-0.8734	0.382448	
										***	-0.7052039	0.132675	-5.3152	1.065e-07	

										***	0.9811655	0.150399	6.5237	6.859e-11	

										***	intercept	0.0225145	0.004093	5.4997	3.804e-08

C1 – Fabrication alimentaire/TPE					C1 – Fabrication alimentaire/PME					C2 – Cokefaction/Raffinage					
	Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)	
ar1	-0.234703	0.145387	-1.6143	0.1064552	ar1	0.1097463	0.1454419	0.7546	0.45051	ar1	1.516000	0.133277	11.3748	< 2.2e-16	
ar2	-0.315515	0.124328	-2.5378	0.0111563 *	***					***					
ar3	0.564608	0.124092	4.5499	5.367e-06	***					***	-0.244161	0.25729	-0.9490	0.3426375	
***					ar3	0.3046247	0.1632187	1.8664	0.06199	***	-0.337066	0.14128	-2.3857	0.0170483	
ar4	0.530869	0.146166	3.6320	0.0002813	***					*					
***					ar4	-0.1714702	0.1161535	-1.4762	0.13988	***	-0.999999	0.06582	-15.1911	< 2.2e-16	
ma1	0.211659	0.202966	1.0428	0.2970265	***					***					
ma2	0.769657	0.189244	4.0670	4.762e-05	***					***	intercept	0.06626	0.01957	3.3847	0.000712
***					ma1	-0.3361479	0.1575976	-2.1330	0.03293 *	***					
ma3	-0.361462	0.234458	-1.5417	0.1231491	***					***					
ma4	-0.361440	0.230596	-1.5674	0.1170180	***					***					
					intercept	0.04437	0.005767	7.6941	1.425e-14	***					

ma5 0.769321 0.282600 2.7223 0.0064831 ** ma6 0.211523 0.218332 0.9688 0.3326388 ma7 0.999771 0.213681 4.6788 2.886e-06 *** intercept 0.082243 0.016931 4.8577 1.188e-06 ***		
C3 – Fabrication electronique/ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.900711 0.151007 5.9647 2.451e-09 *** ar2 -0.452874 0.139065 -3.2566 0.001128 ** ma1 -1.124005 0.075841 -14.8206 < 2.2e-16 *** ma2 0.999951 0.102245 9.7799 < 2.2e-16 *** intercept 0.020482 0.013003 1.5752 0.115208	C3 – Fabrication electronique/TPE Estimate Std. Error z value Pr(> z) ar1 1.8700443 0.0642233 29.1178 < 2.2e-16 *** ar2 -0.9602691 0.0541984 -17.7177 < 2.2e-16 *** ma1 -1.8592509 0.2309842 -8.0493 8.330e-16 *** ma2 0.9998149 0.2446844 4.0861 4.386e-05 *** intercept 0.0633176 0.008727 7.2551 4.013e-13 ***	C3 – Fabrication electronique/PME Estimate Std. Error z value Pr(> z) ar1 1.3016826 0.2363190 5.5082 3.626e-08 *** ar2 -1.4341199 0.3233410 -4.4353 9.194e-06 *** ar3 1.5651283 0.3505161 4.4652 7.999e-06 *** ar4 -0.7541223 0.4334688 -1.7397 0.08190 . ar5 0.6128206 0.3328966 1.8409 0.06564 . ar6 -0.3946308 0.2879702 -1.3704 0.17057 ar7 -0.1203786 0.2171655 -0.5543 0.57936 ma1 -1.6390897 0.1993145 -8.2236 < 2.2e-16 *** ma2 2.4656365 0.3327266 7.4104 1.259e-13 *** ma3 -2.4412708 0.4005039 -6.0955 1.091e-09 *** ma4 1.3869631 0.3549946 3.9070 9.345e-05 *** ma5 -0.7719228 0.1926918 -4.0060 6.176e-05 *** intercept 0.0416761 0.003010 13.8429 < 2.2e-16 ***
C4 – Fabrication materiel de transport/ETI-GE Estimate Std. Error z value Pr(> z) ar1 -1.254298 0.185146 -6.7746 1.247e-11 *** ar2 -0.610955 0.162519 -3.7593 0.0001704 *** ma1 1.545078 0.178200 8.6705 < 2.2e-16 *** ma2 1.121906 0.275217 4.0764 4.573e-05 *** ma3 0.609871 0.298399 2.0438 0.0409724 * ma4 0.719909 0.272668 2.6402 0.0082846 ** ma5 0.584666 0.151933 3.8482 0.0001190 *** intercept 0.039862 0.014909 2.6736 0.0075039 **	C4 – Fabrication materiel de transport/TPE Estimate Std. Error z value Pr(> z) ar1 0.551388 0.145108 3.7998 0.0001448 *** ar2 -1.358210 0.141857 -9.5745 < 2.2e-16 *** ar3 0.743878 0.205209 3.6250 0.0002890 *** ar4 -0.461855 0.150331 -3.0723 0.0021245 ** ar5 0.451862 0.140393 3.2186 0.0012884 ** ma1 -0.559718 0.127849 -4.3780 1.198e-05 *** ma2 2.029675 0.253306 8.0127 1.122e-15 *** ma3 -0.559572 0.154585 -3.6198 0.0002948 *** ma4 0.998892 0.245654 4.0663 4.777e-05 *** intercept 0.096884 0.014096 6.8733 6.275e-12 ***	C4 – Fabrication materiel de transport/PME Estimate Std. Error z value Pr(> z) ar1 -1.1242274 0.0329112 -34.1594 < 2.2e-16 *** ar2 -0.9944219 0.0148249 -67.0780 < 2.2e-16 *** ma1 1.0763065 0.0836259 12.8705 < 2.2e-16 *** ma2 0.9998995 0.0851872 11.7377 < 2.2e-16 *** intercept 0.0485431 0.005969 8.1319 4.227e-16 ***
C5 – Autres produits indus/ETI-GE Estimate Std. Error z value Pr(> z)	C5 – Autres produits indus/TPE Estimate Std. Error z value Pr(> z)	C5 – Autres produits indus/PME Estimate Std. Error z value Pr(> z)

ar1 0.866121 0.111672 7.7559 8.769e-15 *** ma1 -1.290825 0.189211 -6.8222 8.968e-12 *** ma2 0.460317 0.237123 1.9413 0.05223 . ma3 0.562109 0.218955 2.5672 0.01025 * ma4 -0.331905 0.169153 -1.9622 0.04974 * intercept 0.025028 0.014486 1.7277 0.08405 .	ar1 0.6266390 0.1419788 4.4136 1.017e-05 *** ma1 -0.9721226 0.1331499 -7.3010 2.857e-13 *** ma2 0.7128178 0.1385506 5.1448 2.678e-07 *** intercept 0.0549480 0.008590 6.3965 1.590e-10 ***	ar1 -0.733335 0.272824 -2.6879 0.0071893 ** ar2 0.011297 0.155427 0.0727 0.9420593 ar3 0.472152 0.159465 2.9609 0.0030678 ** ar4 0.425453 0.143505 2.9647 0.0030296 ** ma1 0.517501 0.281975 1.8353 0.0664655 . intercept 0.033188 0.009943 3.3378 0.0008443 ***
DE – Extraction + Energie/ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.900553 0.116198 7.7502 9.178e-15 *** ar2 -1.592230 0.133561 -11.9214 < 2.2e-16 *** ar3 1.324314 0.192054 6.8955 5.367e-12 *** ar4 -0.753301 0.156960 -4.7993 1.592e-06 *** ar5 0.687912 0.137339 5.0089 5.476e-07 *** ma1 -0.646210 0.130379 -4.9564 7.181e-07 *** ma2 1.942627 0.207683 9.3538 < 2.2e-16 *** ma3 -0.646209 0.180247 -3.5851 0.0003369 *** ma4 0.999997 0.203935 4.9035 9.414e-07 *** intercept 0.041245 0.024904 1.6562 0.0976838 .	DE – Extraction + Energie/TPE Estimate Std. Error z value Pr(> z) ar1 0.902086 0.072879 12.3778 < 2.2e-16 *** ma1 -1.246525 0.156867 -7.9464 1.92e-15 *** ma2 0.419973 0.206334 2.0354 0.04181 * ma3 0.152629 0.145971 1.0456 0.29574 intercept 0.082586 0.021225 3.8911 9.98e-05 ***	DE – Extraction + Energie/PME Estimate Std. Error z value Pr(> z) ar1 0.3517905 0.1240537 2.8358 0.0045712 ** ar2 0.6567616 0.1285911 5.1074 3.267e-07 *** ar3 0.8954276 0.0694451 12.8940 < 2.2e-16 *** ar4 -0.4383465 0.1252372 -3.5001 0.0004650 *** ar5 -0.5648885 0.1229497 -4.5945 4.339e-06 *** ma1 -0.6918953 0.1365307 -5.0677 4.027e-07 *** ma2 -0.5912433 0.1243683 -4.7540 1.995e-06 *** ma3 -0.6917322 0.1882036 -3.6754 0.0002374 *** ma4 0.9998804 0.1591274 6.2835 3.310e-10 *** intercept 0.0760304 0.003924 19.3724 < 2.2e-16 ***
FZ – Construction/ETI-GE Estimate Std. Error z value Pr(> z) ar1 1.621192 0.138948 11.6676 < 2.2e-16 *** ar2 -0.701693 0.133065 -5.2733 1.34e-07 *** ma1 -1.762473 0.086708 -20.3266 < 2.2e-16 *** ma2 0.999996 0.085641 11.6766 < 2.2e-16 *** intercept 0.044121 0.019730 2.2362 0.02534 *	FZ – Construction/TPE Estimate Std. Error z value Pr(> z) ar1 1.8734821 0.0793515 23.6099 < 2.2e-16 *** ar2 -0.9229688 0.0766744 -12.0375 < 2.2e-16 *** ma1 -1.816927 0.2175687 -8.3511 < 2.2e-16 *** ma2 0.8169383 0.2073787 3.9394 8.17e-05 *** intercept 0.1039736 0.0021762 47.7778 < 2.2e-16 ***	FZ – Construction/PME Estimate Std. Error z value Pr(> z) ar1 0.407190 0.163325 2.4931 0.012662 * ma1 0.451109 0.147924 3.0496 0.002291 ** intercept 0.024028 0.042198 0.5694 0.569072
GZ – Commerce/ETI-GE Estimate Std. Error z value Pr(> z) ar1 -0.239631 0.085806 -2.7927 0.0052270 ** ar2 0.056504 0.096680 0.5844 0.5589224 ar3 0.912980 0.082151 11.1135 < 2.2e-16 *** ma1 0.482854 0.145220 3.3250 0.0008842 *** ma2 0.165520 0.168457 0.9826 0.3258209 ma3 -0.703428 0.139172 -5.0544 4.318e-07 *** intercept 0.037548 0.012319 3.0481 0.0023032 **	GZ – Commerce/TPE Estimate Std. Error z value Pr(> z) ar1 0.9843849 0.1559402 6.3126 2.744e-10 *** ar2 -0.4933423 0.1538758 -3.2061 0.001345 ** ma1 -1.067737 0.0810828 -13.1685 < 2.2e-16 *** ma2 0.999972 0.1060827 9.4263 < 2.2e-16 *** intercept 0.0651243 0.005609 11.6101 < 2.2e-16 ***	GZ – Commerce/PME Estimate Std. Error z value Pr(> z) ar1 0.32103423 0.17267164 1.8592 0.0629962 . ar2 -1.21116509 0.1420644 -8.5255 < 2.2e-16 *** ar3 0.55765643 0.22284976 2.5024 0.0123359 * ar4 -0.34826646 0.22876072 -1.5224 0.1279076 ar5 0.47539632 0.139449 3.4091 0.0006518 ***

		ar6 0.00074438 0.17355062 0.0043 0.9965778 ma1 -0.55153439 0.111013 -4.9682 6.759e-07 *** ma2 2.05193425 0.208606 9.8364 < 2.2e-16 *** ma3 -0.549249 0.1341958 -4.0929 4.260e-05 *** ma4 0.9917431 0.2039035 4.8638 1.152e-06 *** intercept 0.0487264 0.0056767 8.5835 < 2.2e-16 ***
HZ – Transport/ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.342746 0.201295 1.7027 0.0886225 . ma1 -0.333642 0.184964 -1.8038 0.0712586 . ma2 0.374087 0.161934 2.3101 0.0208819 * ma3 0.641710 0.166348 3.8576 0.0001145 *** intercept 0.024958 0.019840 1.2579 0.2084197	HZ – Transport/TPE Estimate Std. Error z value Pr(> z) ar1 0.872933 0.072409 12.0557 < 2.2e-16 *** ma1 -1.445126 0.108437 -13.3269 < 2.2e-16 *** ma2 0.999977 0.138858 7.2014 5.958e-13 *** intercept 0.099035 0.021610 4.5829 4.586e-06 ***	HZ – Transport/PME Estimate Std. Error z value Pr(> z) ar1 1.8999758 0.2853079 6.6594 2.750e-11 *** ar2 -0.8453102 0.5611691 -1.5063 0.13198 ar3 -0.0771111 0.2827162 -0.2728 0.78504 ma1 -3.0663735 0.078492 -39.0658 < 2.2e-16 *** ma2 3.0561168 0.541248 5.6464 1.638e-08 *** ma3 -0.4071508 1.0634442 -0.3829 0.70182 ma4 -1.0481779 0.8141526 -1.2874 0.19794 ma5 0.4666161 0.2198742 2.1222 0.03382 * intercept 0.0620487 0.001537 40.3573 < 2.2e-16 ***
IZ – Restaurants / Hotels/ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.9615422 0.0691323 13.9087 < 2.2e-16 *** ar2 0.8548429 0.1013013 8.4386 < 2.2e-16 *** ar3 -0.894163 0.0705572 -12.6729 < 2.2e-16 *** ma1 -1.0813905 0.172855 -6.2561 3.948e-10 *** ma2 -1.286841 0.2236521 -5.7538 8.728e-09 *** ma3 1.638547 0.2235875 7.3284 2.328e-13 *** ma4 0.5366476 0.1951635 2.7497 0.005964 ** ma5 -0.806954 0.1529250 -5.2768 1.315e-07 *** intercept 0.03592 0.0018136 19.8060 < 2.2e-16 ***	IZ – Restaurants / Hotels/TPE Estimate Std. Error z value Pr(> z) ar1 0.93343 0.07284 12.8148 < 2.2e-16 *** ma1 -0.35094 0.17876 -1.9632 0.049619 * intercept 0.11807 0.03727 3.1681 0.001534 **	IZ – Restaurants / Hotels/PME Estimate Std. Error z value Pr(> z) ar1 0.5721717 0.1639698 3.4895 0.0004839 *** ma1 -0.1207111 0.1819372 -0.6635 0.5070253 ma2 0.1582268 0.1282108 1.2341 0.2171603 ma3 0.0443197 0.1215724 0.3646 0.7154444 ma4 0.5505596 0.1357149 4.0567 4.976e-05 *** intercept 0.066205 0.0084515 7.8336 4.742e-15 ***
JZ – TIC/ETI-GE Estimate Std. Error z value Pr(> z) ar1 -1.0122218 0.1401180 -7.2241 5.046e-13 *** ar2 -0.1516783 0.1421335 -1.0672 0.2859 ma1 0.9999919 0.0598781 16.7005 < 2.2e-16	JZ – TIC/TPE Estimate Std. Error z value Pr(> z) ar1 0.810391 0.099058 8.1810 2.815e-16 *** ma1 -0.796276 0.175903 -4.5268 5.989e-06 ***	JZ – TIC/PME Estimate Std. Error z value Pr(> z) ar1 -0.066612 0.256645 -0.2595 0.7952134 ar2 0.191577 0.127910 1.4977 0.1342009 ar3 0.461799 0.138299 3.3391 0.0008404 ***

*** intercept 0.042453 0.0056516 7.5118 5.832e-14 ***	ma2 0.430933 0.209286 2.0591 0.03949 * ma3 -0.332529 0.232394 -1.4309 0.15246 ma4 -0.180199 0.208914 -0.8625 0.38839 ma5 0.758849 0.178056 4.2619 2.027e-05 *** intercept 0.130644 0.020365 6.4152 1.406e-10 ***	ma1 -0.157174 0.290708 -0.5407 0.5887419 intercept 0.089892 0.011033 8.1475 3.717e-16 ***
LZ – Immobilier/ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.668120 0.135077 4.9462 7.567e-07 *** ar2 -1.081304 0.164996 -6.5535 5.620e-11 *** ar3 0.549950 0.206803 2.6593 0.007830 ** ar4 -0.159658 0.161048 -0.9914 0.321506 ar5 0.359763 0.130288 2.7613 0.005758 ** ma1 -0.485848 NA NA NA ma2 0.988408 NA NA NA intercept 0.044704 0.015164 2.9481 0.003198 **	LZ – Immobilier/TPE Estimate Std. Error z value Pr(> z) ar1 1.8422053 0.0464134 39.6913 < 2.2e-16 *** ar2 -0.9528128 0.0451986 -21.0806 < 2.2e-16 *** ma1 -1.7948829 0.1016303 -17.6609 < 2.2e-16 *** ma2 0.9999865 0.1089090 9.1819 < 2.2e-16 *** intercept 0.068948 0.0084729 8.1376 4.031e-16 ***	LZ – Immobilier/PME Estimate Std. Error z value Pr(> z) ar1 1.1003244 0.1511132 7.2815 3.302e-13 *** ar2 0.5246910 0.1953869 2.6854 0.007244 ** ar3 -0.7524332 0.1951247 -3.8562 0.000115 *** ar4 -0.0062132 0.1513395 -0.0411 0.9672525 ma1 -0.9957955 0.1316753 -7.5625 3.954e-14 *** ma2 -0.9957182 0.1370555 -7.2651 3.728e-13 *** ma3 0.9999650 0.1316511 7.5956 3.064e-14 *** intercept 0.061749 0.0010296 59.9761 < 2.2e-16 ***
MN – Scientifique / Technique/ ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.642516 0.660667 0.9725 0.3308 ma1 -0.533379 0.733395 -0.7273 0.4671 intercept 0.026376 0.030767 0.8573 0.3913	MN – Scientifique / Technique/TPE Estimate Std. Error z value Pr(> z) ar1 1.029634 0.158126 6.5115 7.440e-11 *** ar2 -1.08007 0.17459 -6.1863 6.158e-10 *** ar3 0.610532 0.146794 4.1591 3.195e-05 *** ma1 -1.261748 0.16020 -7.8758 3.387e-15 *** ma2 2.018334 0.414087 4.8742 1.093e-06 *** ma3 -1.260890 0.462780 -2.7246 0.006438 ** ma4 0.998909 0.385779 2.5893 0.009616 ** intercept 0.105269 0.013376 7.8703 3.539e-15 ***	MN – Scientifique / Technique/PME Estimate Std. Error z value Pr(> z) ar1 0.7615778 0.2479058 3.0720 0.002126 ** ma1 -1.290930 0.2717478 -4.7505 2.029e-06 *** ma2 0.568728 0.1456168 3.9056 9.397e-05 *** intercept 0.076841 0.0057756 13.3045 < 2.2e-16 ***
OQ – Santé/TPE Estimate Std. Error z value Pr(> z) ar1 0.891023 0.092830 9.5984 < 2.2e-16 *** ma1 -0.999931 0.055304 -18.0806 < 2.2e-16 *** intercept 0.085479 0.022491 3.8005 0.0001444 ***	OQ – Santé/Plus de 10 salariés Estimate Std. Error z value Pr(> z) ar1 0.8231159 0.0888970 9.2592 < 2.2e-16 *** ma1 -0.9999997 0.0522266 -19.1473 < 2.2e-16 *** intercept 0.045011 0.0013972 32.2147 < 2.2e-16 ***	RU – Autres Services/ETI-GE Estimate Std. Error z value Pr(> z) ar1 0.108261 0.169617 0.6383 0.52330 ar2 -0.807075 0.086841 -9.2938 < 2.2e-16 *** ar3 -0.384471 0.166375 -2.3109 0.02084 * ma1 -0.442317 0.094368 -4.6872 2.77e-06 *** ma2 0.999963 0.086945 11.5011 < 2.2e-16 *** intercept 0.051704 0.004427 11.6792 < 2.2e-16 ***
RU – Autres Services/TPE	RU – Autres Services/PME	

	Estimate	Std. Error	z value	Pr(> z)		Estimate	Std. Error	z value	Pr(> z)		
ar1	0.965348	0.047405	20.3639	< 2.2e-16	***	ar1	-0.3105171	0.1931964	-1.6073	0.10800	
ma1	-0.672433	0.123223	-5.4570	4.842e-08	***	ar2	-0.6092969	0.1506412	-4.0447	5.239e-05	
intercept	0.089924	0.032194	2.7932	0.005219	**	***	ma1	0.3807224	0.2352394	1.6184	0.10557
						ma2	0.9949056	0.0949719	10.4758	< 2.2e-16	
						***	ma3	0.3936518	0.2325012	1.6931	0.09043
						intercept	0.0523767	0.006399	8.1849	2.724e-16	

Note de lecture : Significativité des coefficients : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

E. Prévisions par série *secteur x taille*

La méthode de prévision permet de disposer d'une série de chiffres d'affaires observés et d'une série de chiffres d'affaires contrefactuels et ce pour chaque groupe *secteur x taille* et à toute date à partir de février 2020.

Les séries contrefactuelles sont obtenues par les modèles de prévision. Pour ces séries, il est possible de calculer un intervalle de confiance des prévisions effectuées en tout point.

En effet, la prévision linéaire optimale, pour tout groupe *secteur x taille* s , peut se réécrire avec l'écriture $MA(\infty)$ de la façon suivante :

$$\hat{X}_{T+h} = \mu + \sum_{i=h}^{\infty} c_i \varepsilon_{T+h-i}$$

L'erreur quadratique moyenne s'écrit alors $MSE = \sum_{i=0}^{h-1} c_i^2 \sigma^2$ avec la relation de récurrence suivante :

$$MSE(h) = MSE(h-1) + c_{h-1}^2 \sigma^2$$

On en déduit un intervalle de confiance de niveau α pour \hat{X}_{T+h} :

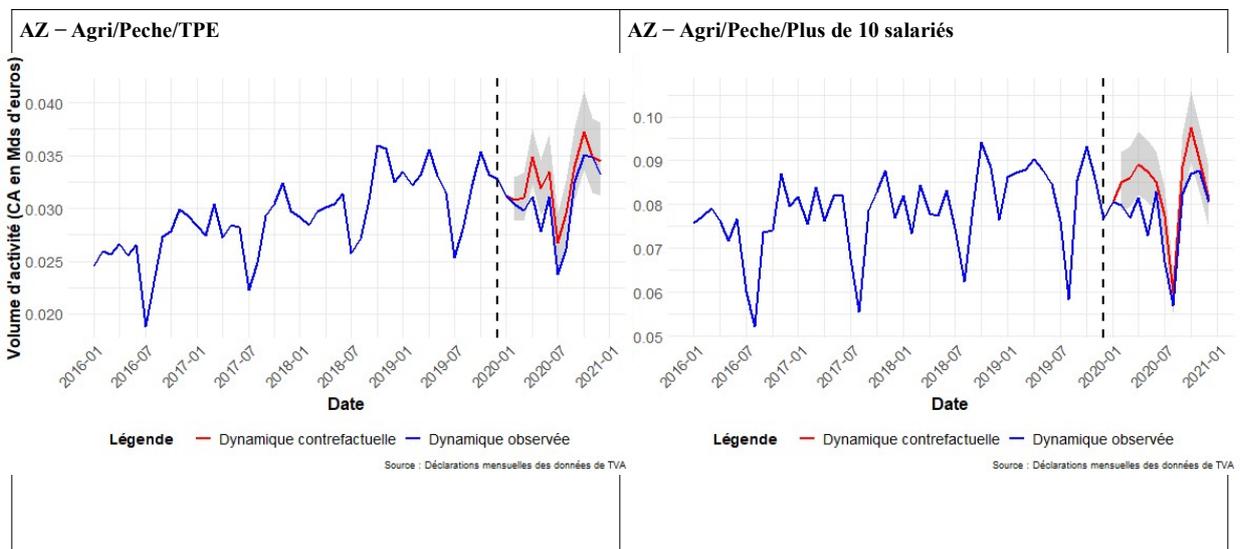
$$IC_{\alpha} = \left[\hat{X}_{T+h} \pm q_{1-\frac{\alpha}{2}} \sqrt{MSE(h)} \right]$$

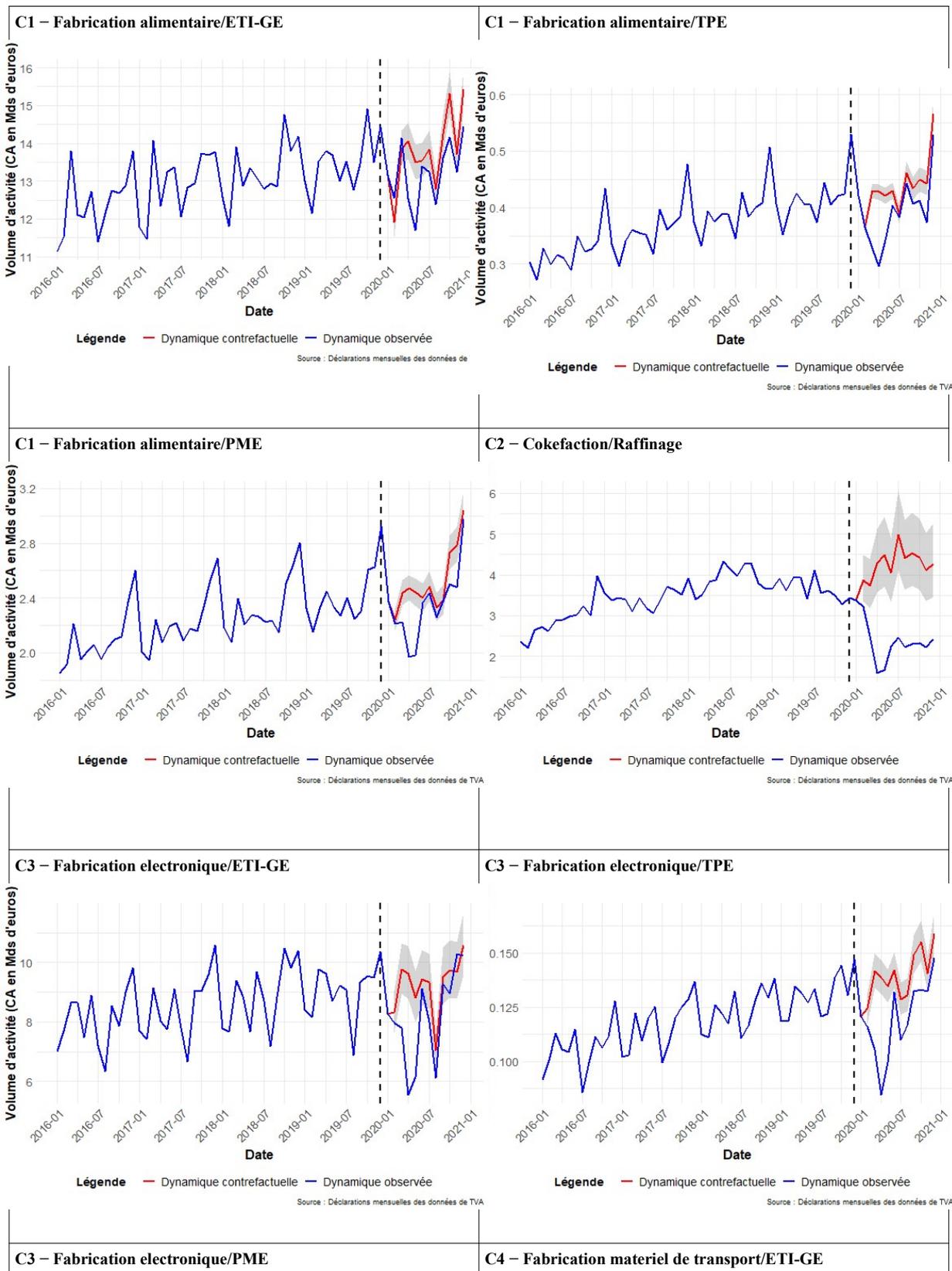
Par croissance de l'exponentielle via $X_t = (1 - B^{12}) Y_t$, on déduit un intervalle de confiance de niveau α pour $(Y_{T+h})_{h \in [1, 10]}$:

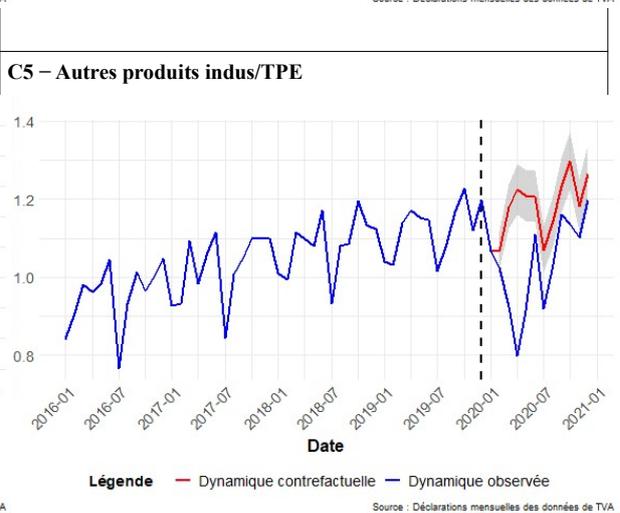
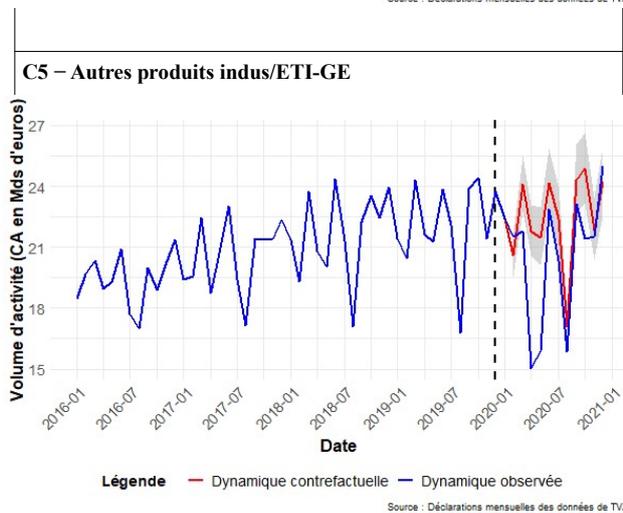
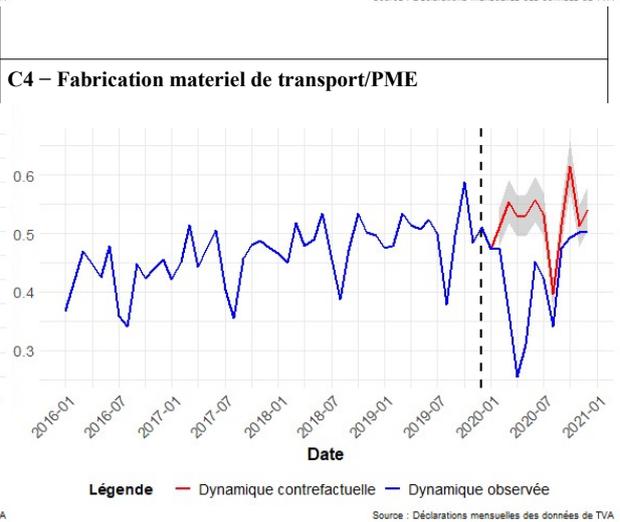
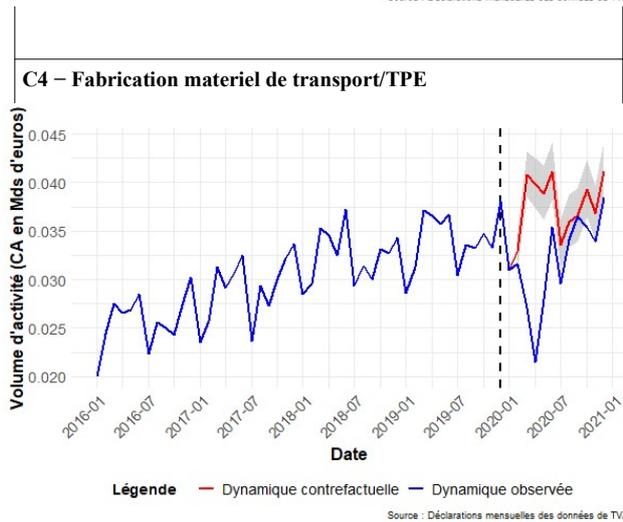
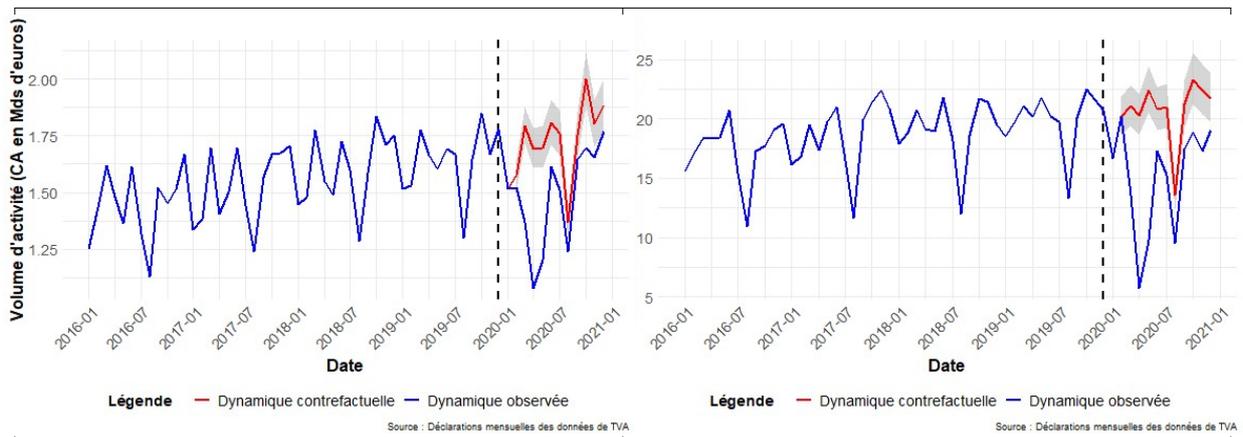
$$IC_{\alpha} = \left[\hat{Y}_{T+h} e^{\pm q_{1-\frac{\alpha}{2}} \sqrt{MSE(h)}} \right], \forall h \in [1, 10]$$

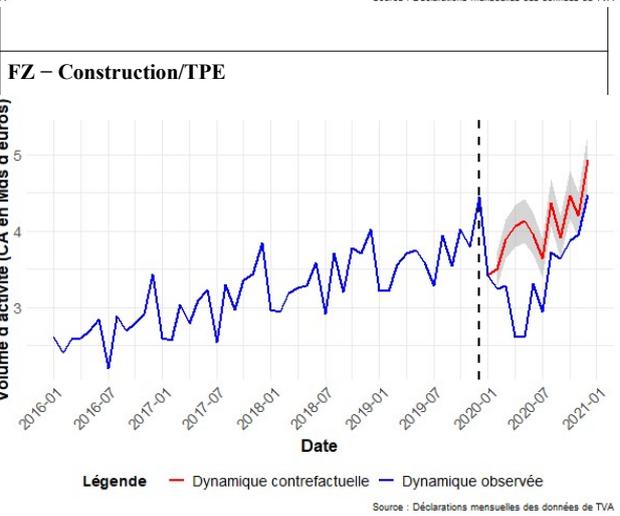
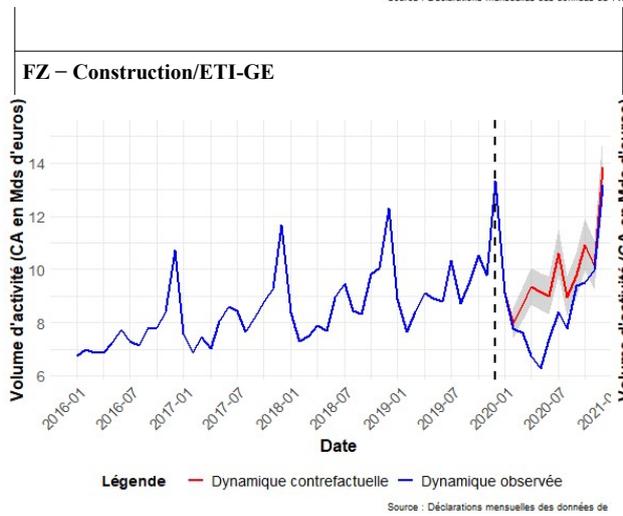
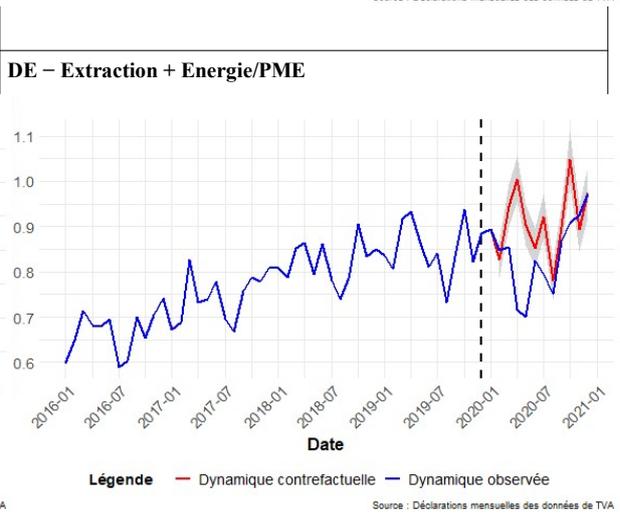
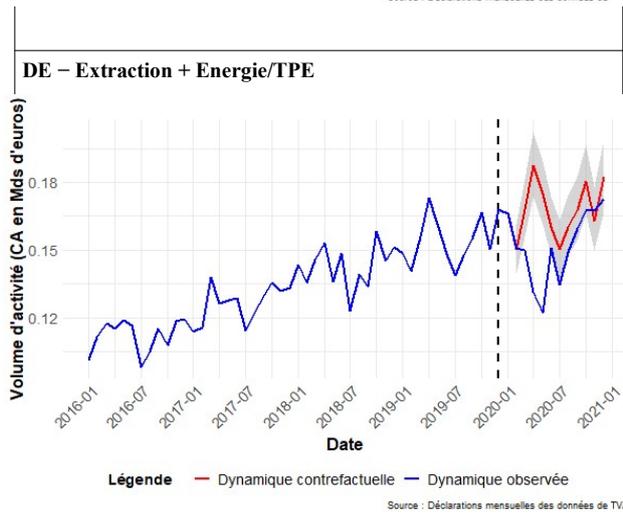
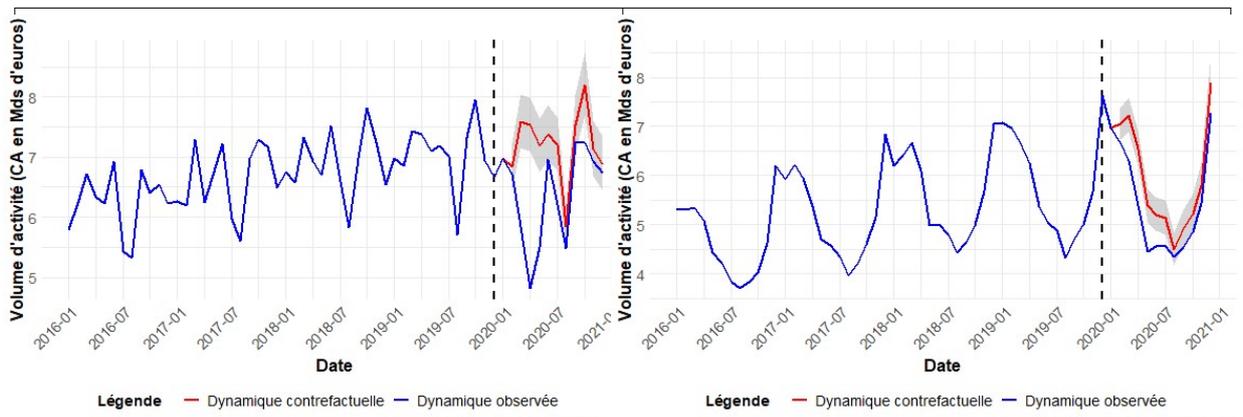
Dans le Tableau E.1, les intervalles de confiance de niveau $\alpha = 5\%$ sont représentés par les zones grisées.

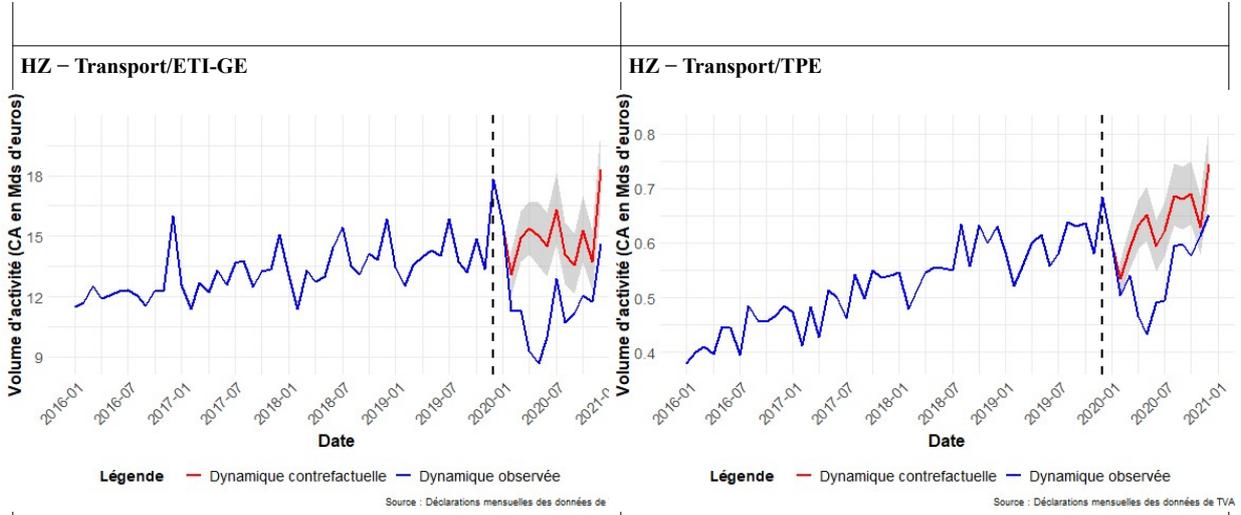
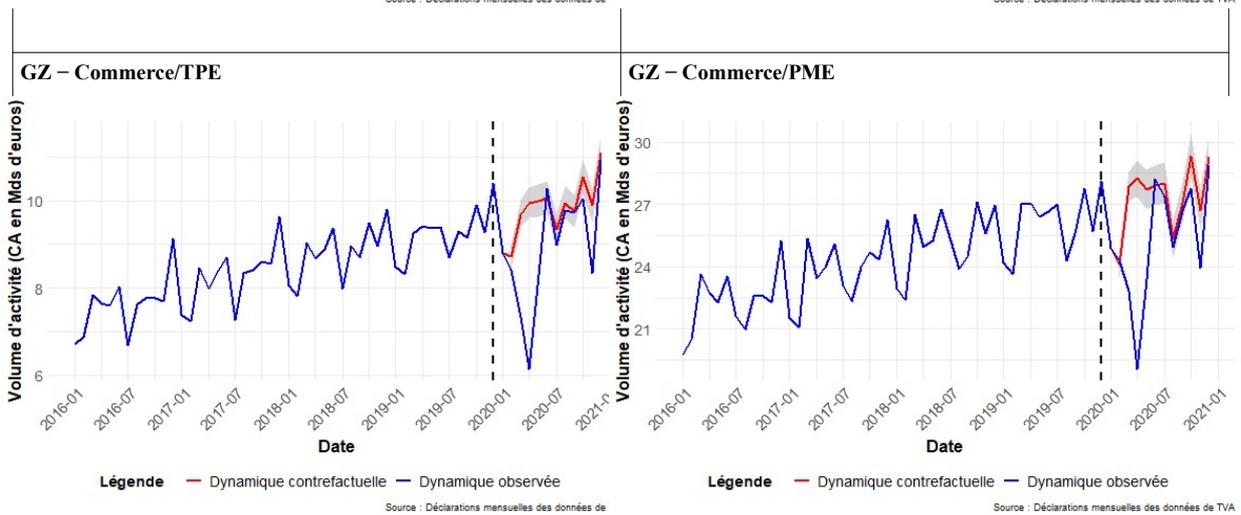
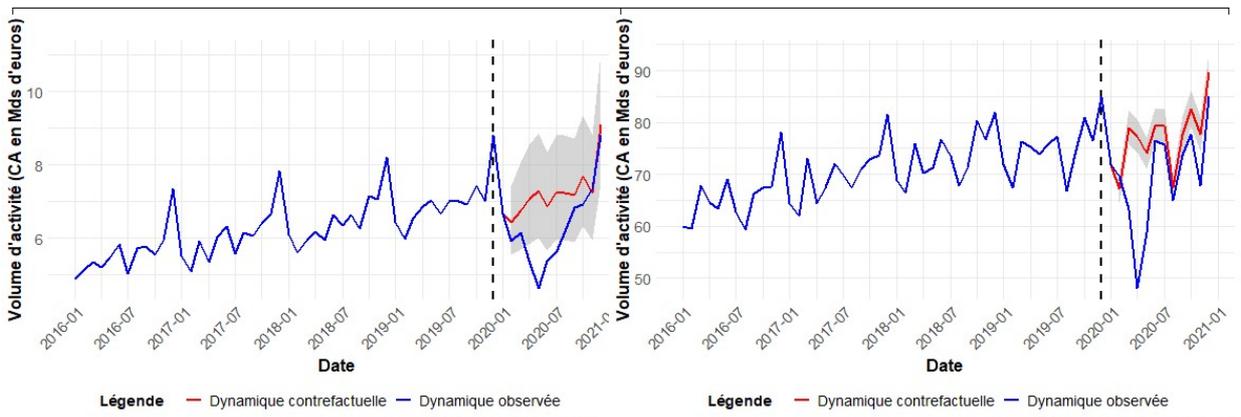
Tableau E.1 – Séries prédites et observées par groupe secteur x taille

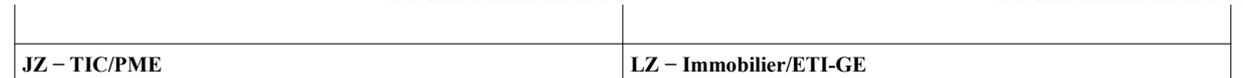
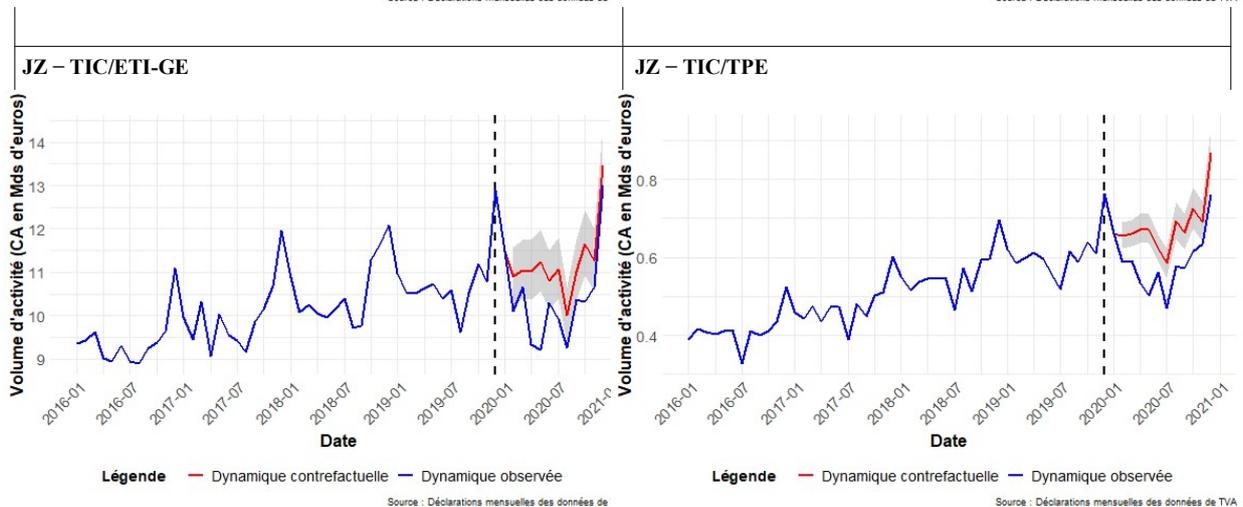
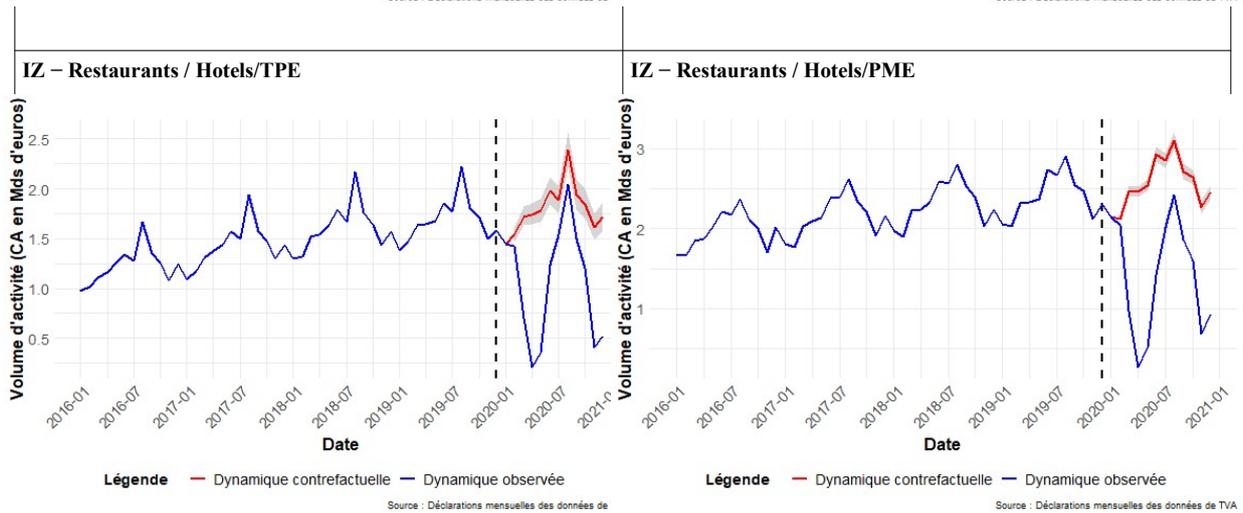
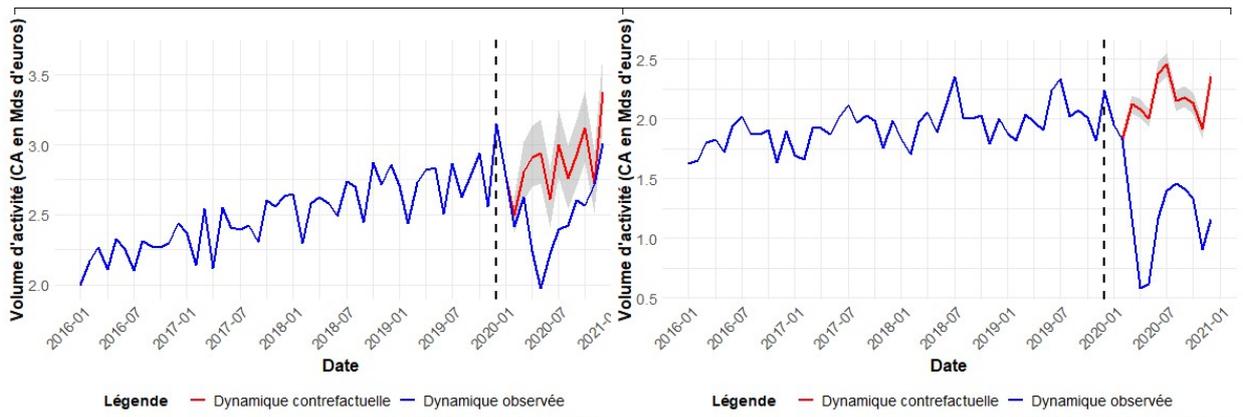


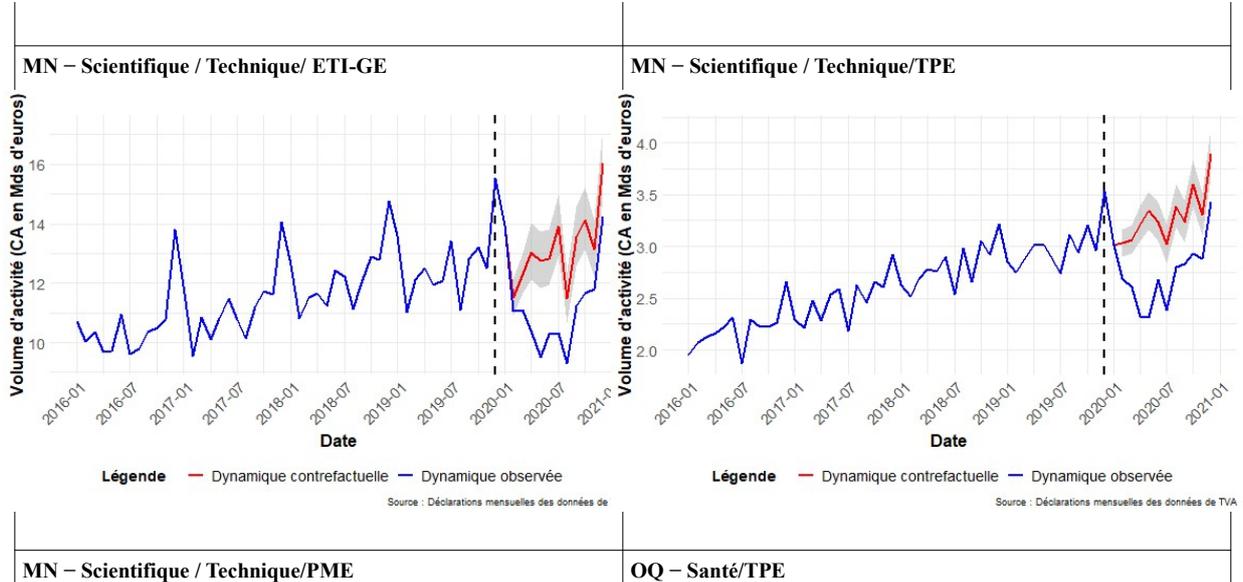
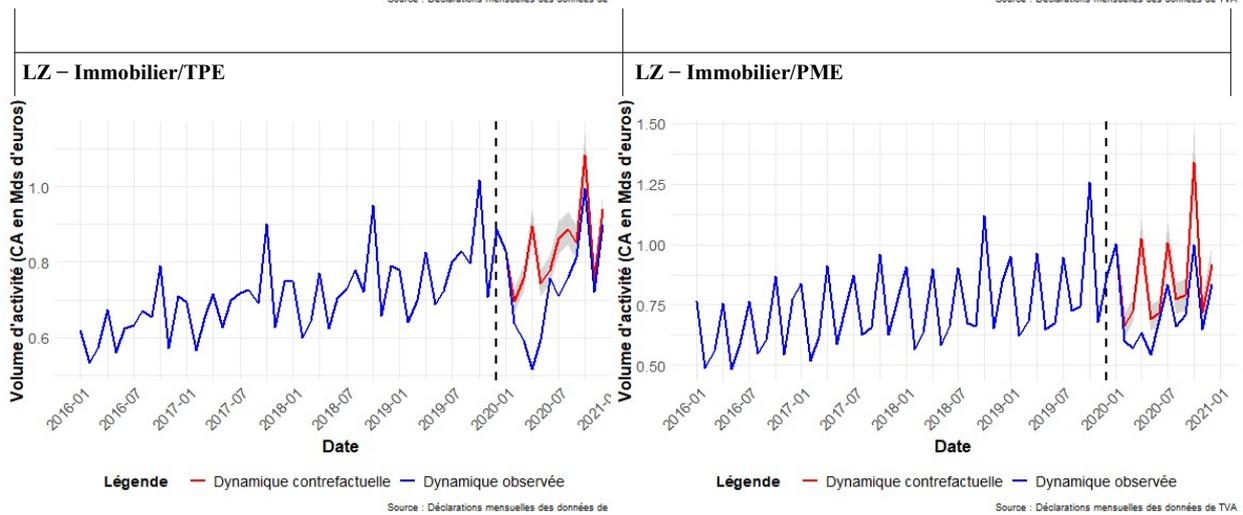
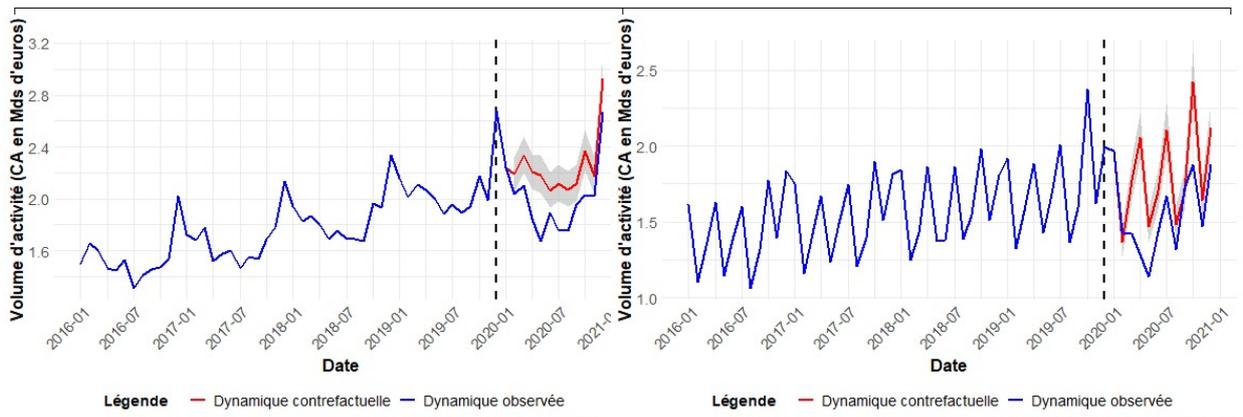


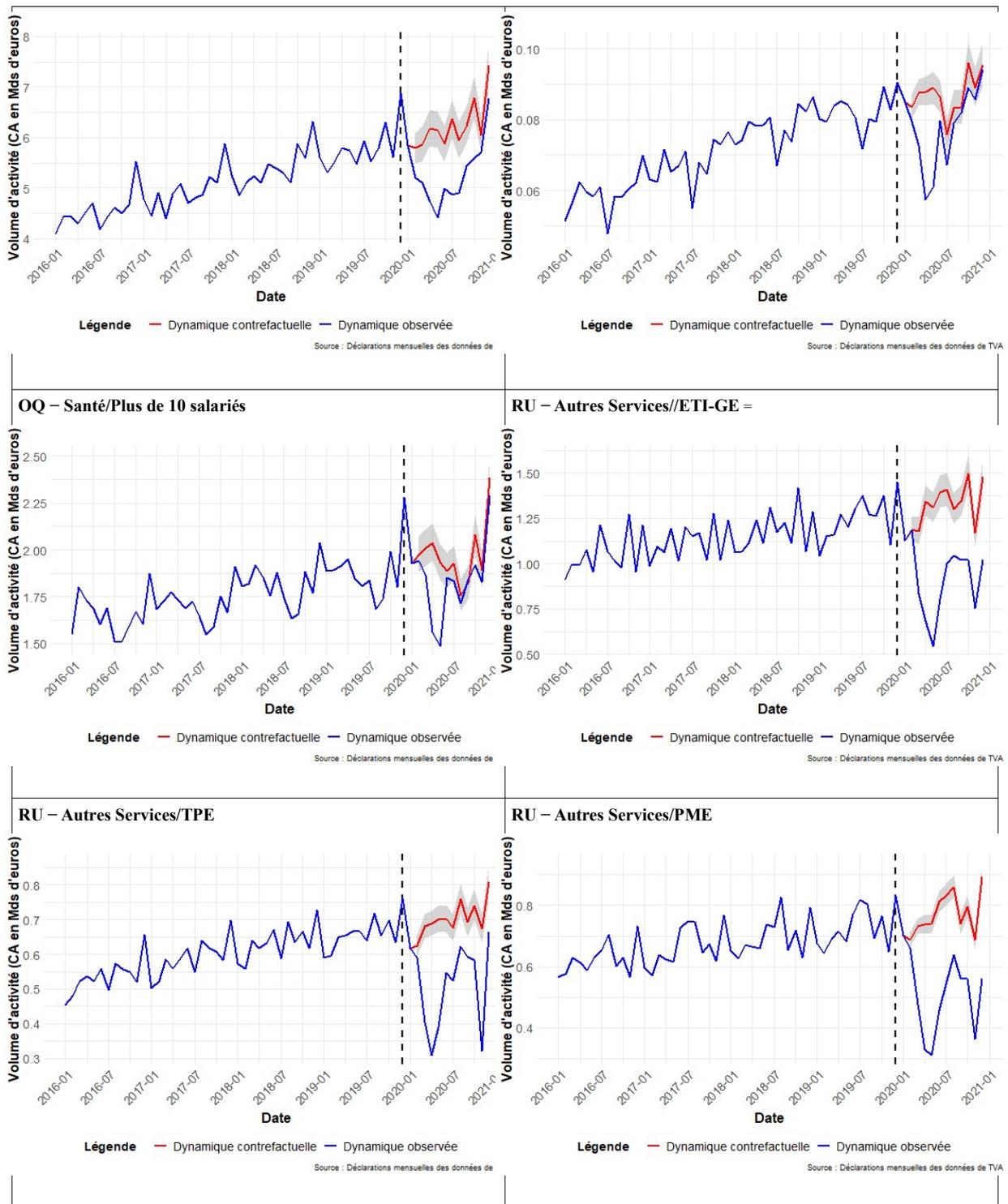










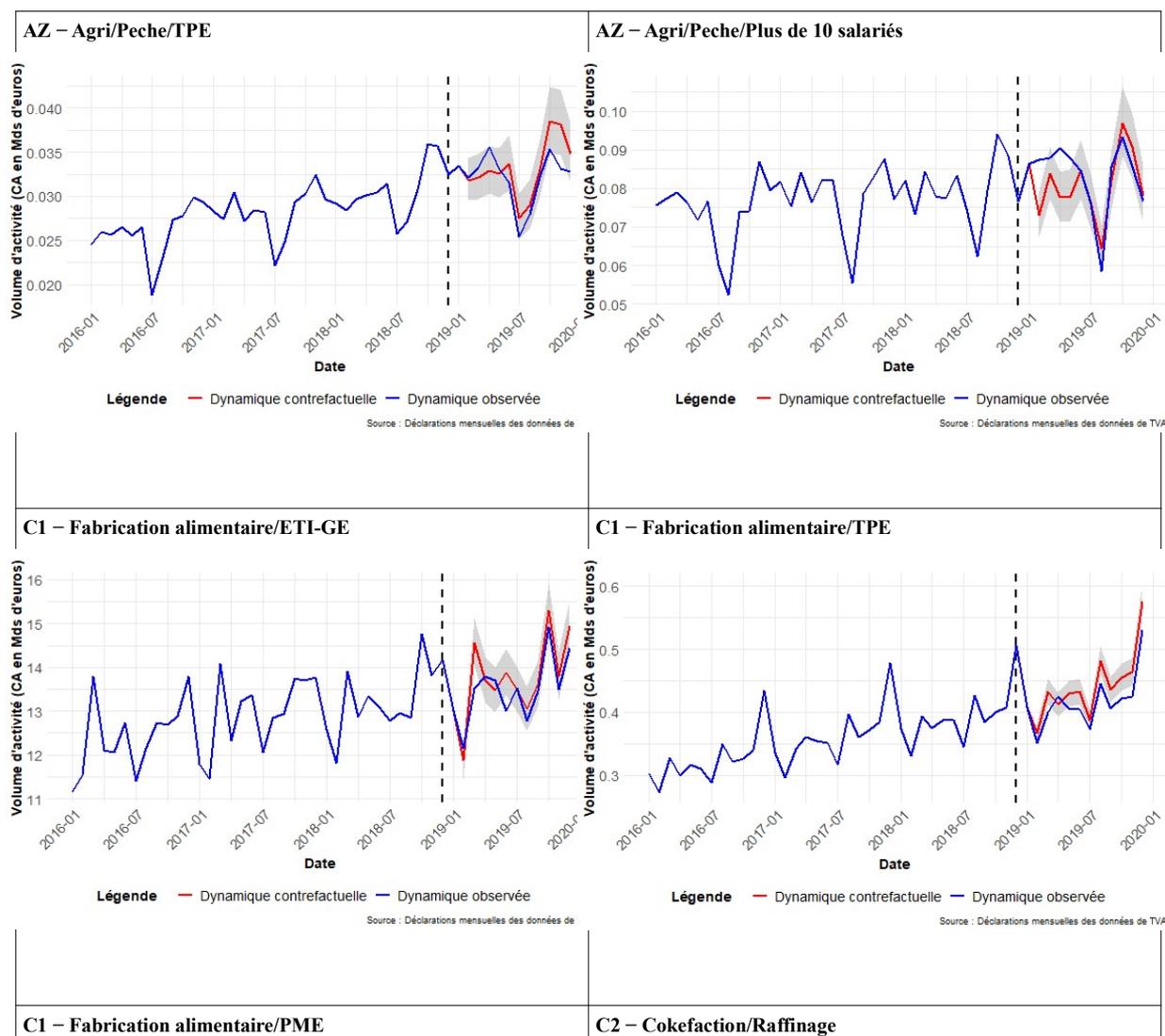


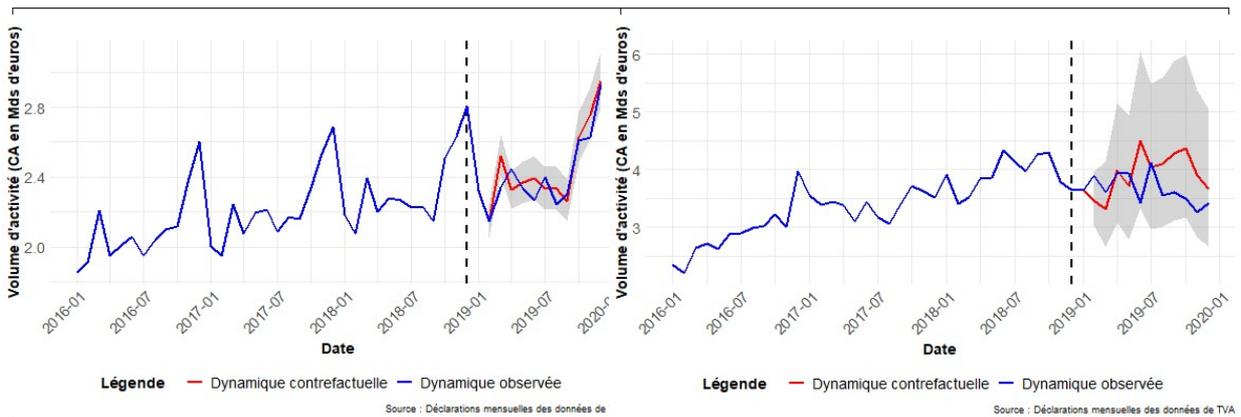
F. Robustesse des chocs d'activité estimés

1. Reproduction en pseudo-temps réel sur une année sans crise

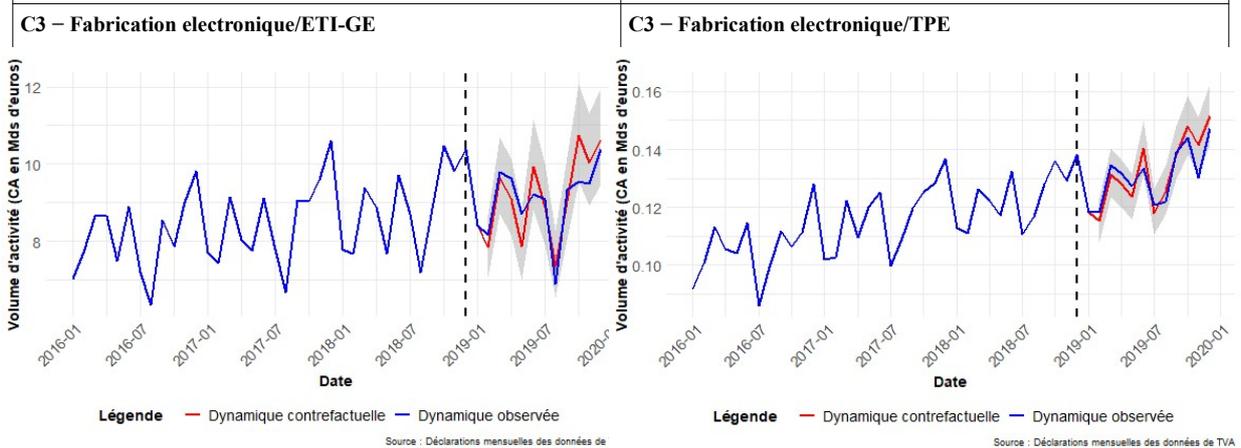
a. Séries agrégées

Figure F.1 – Séries prédites et observées par groupe secteur x taille une année sans crise

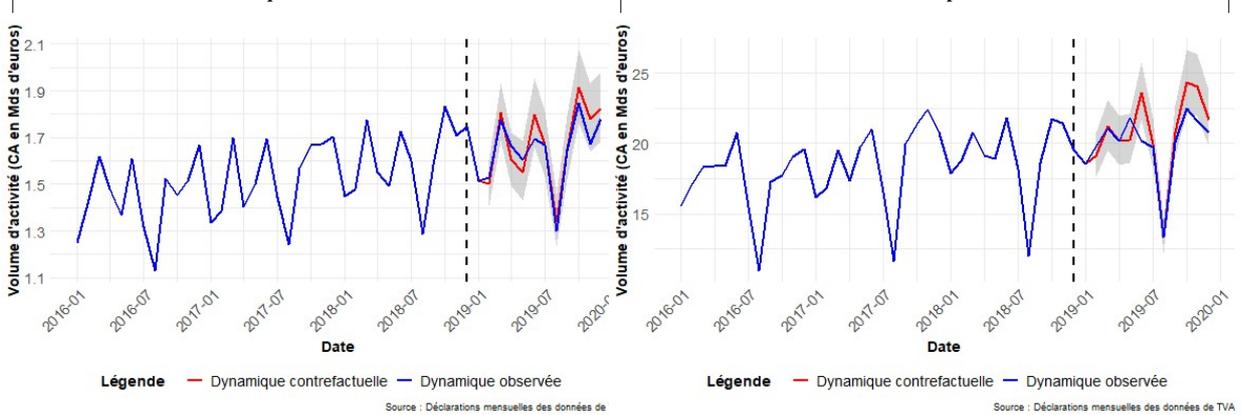




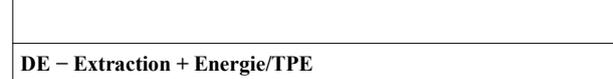
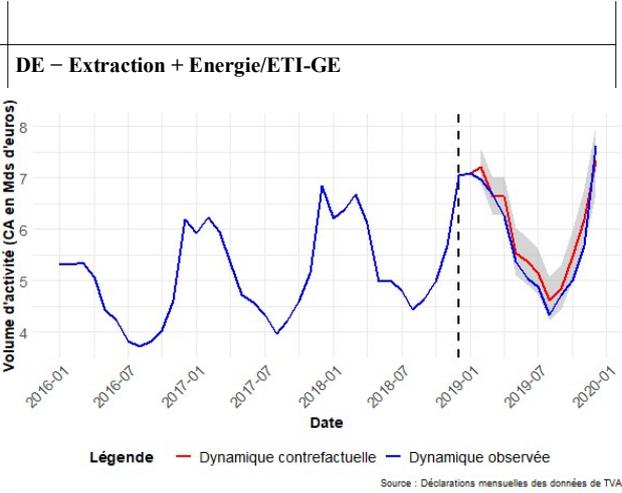
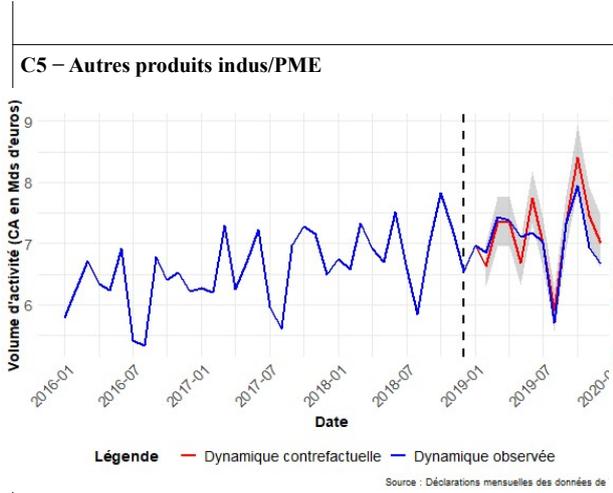
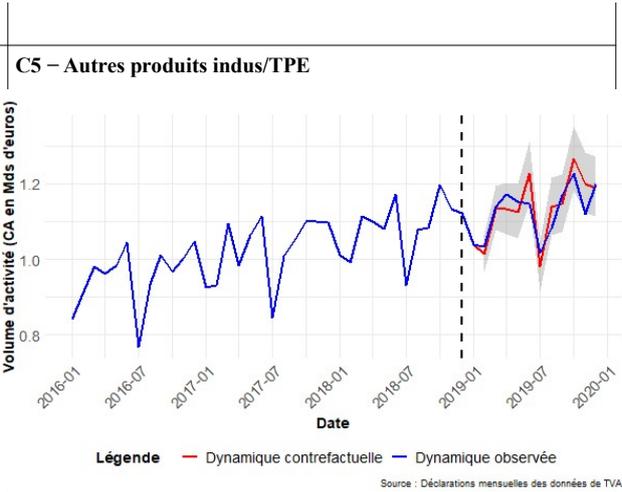
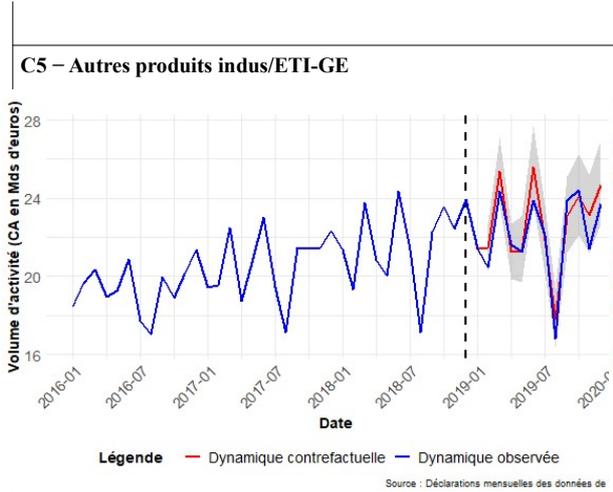
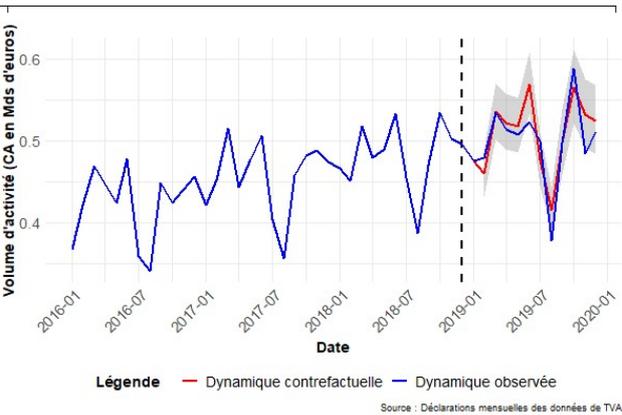
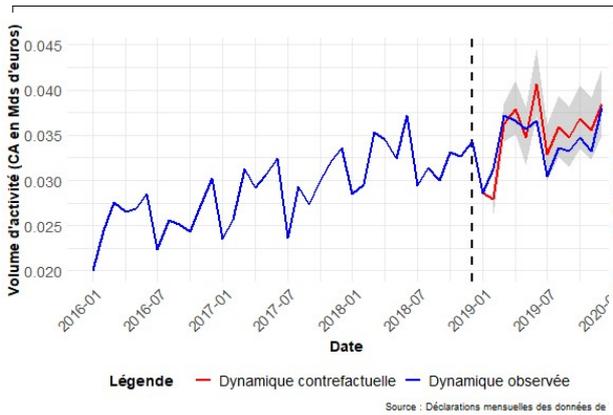
C3 – Fabrication électronique/ETI-GE **C3 – Fabrication électronique/TPE**

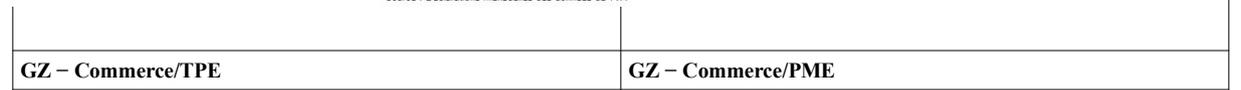
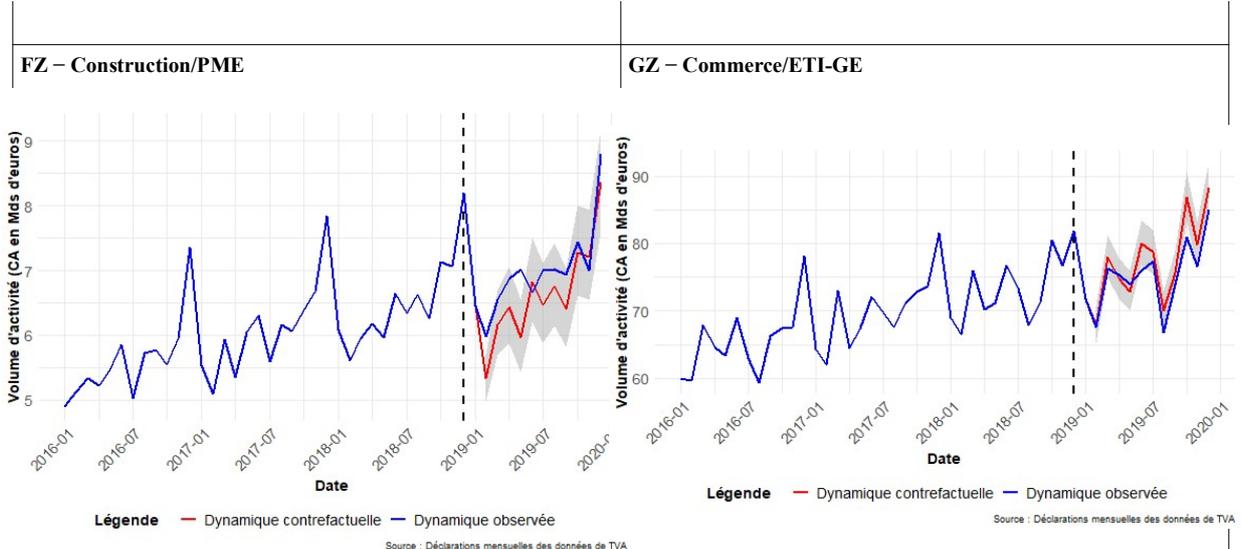
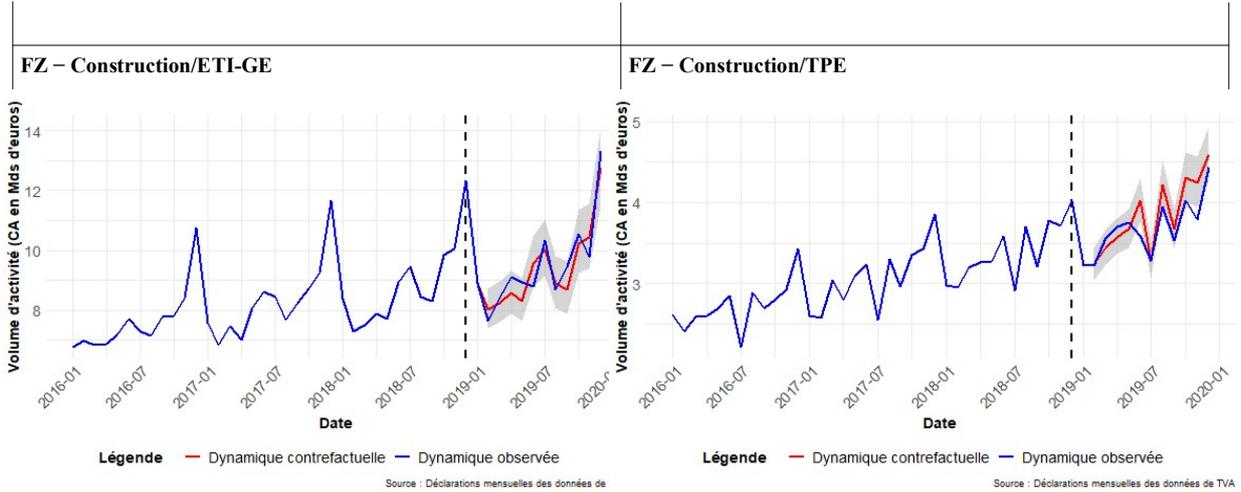
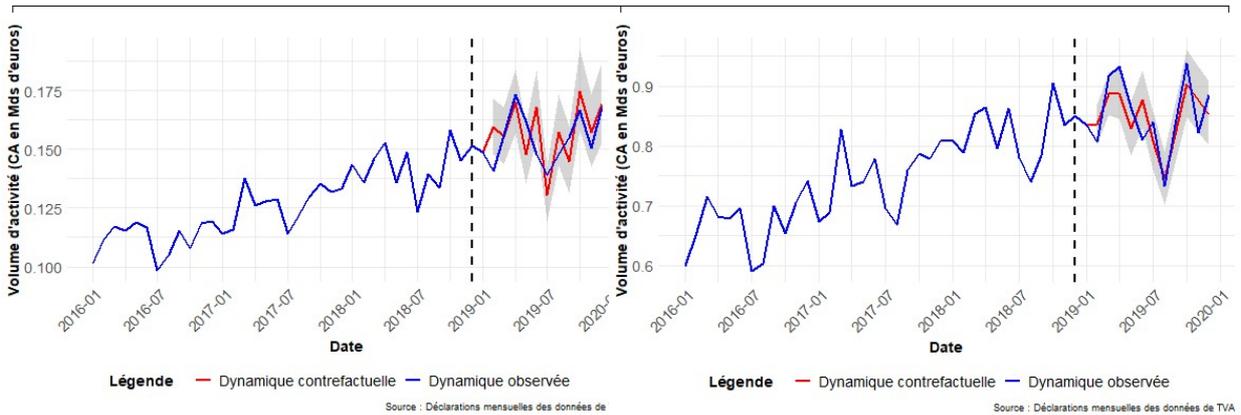


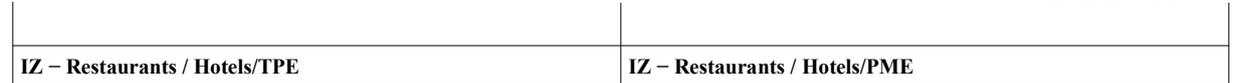
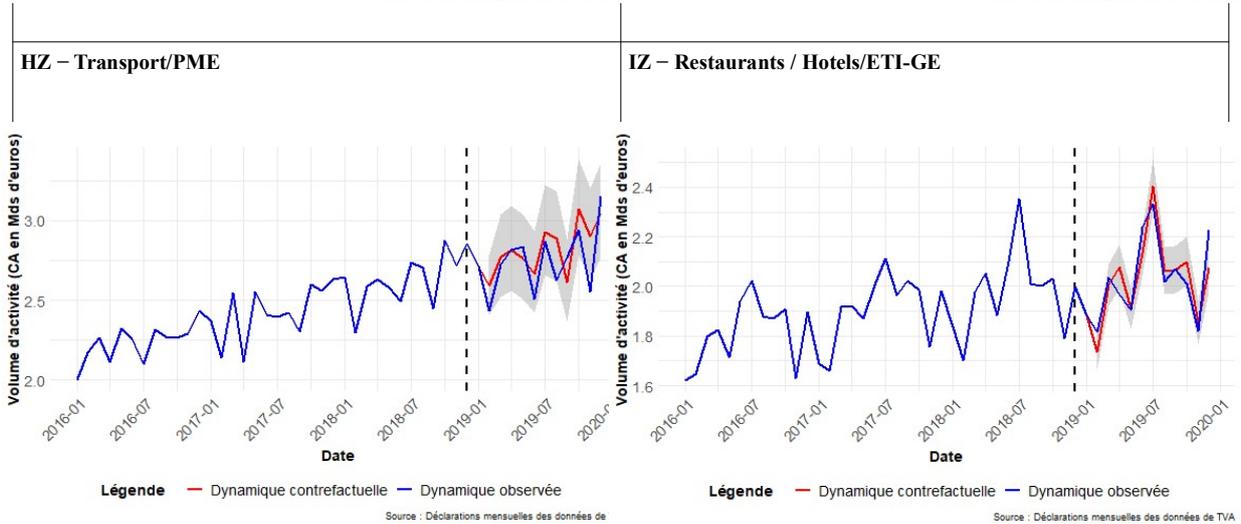
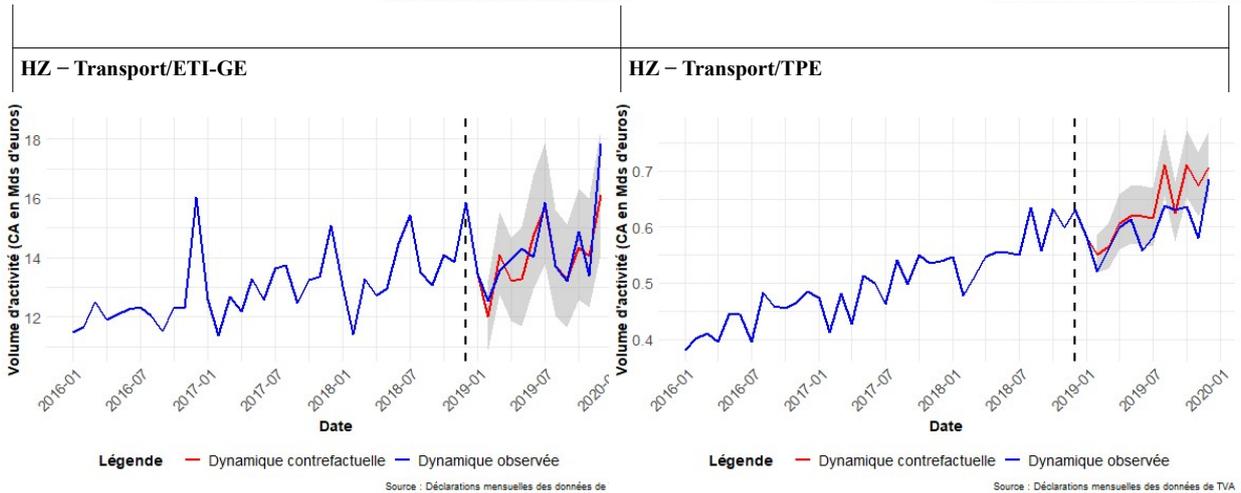
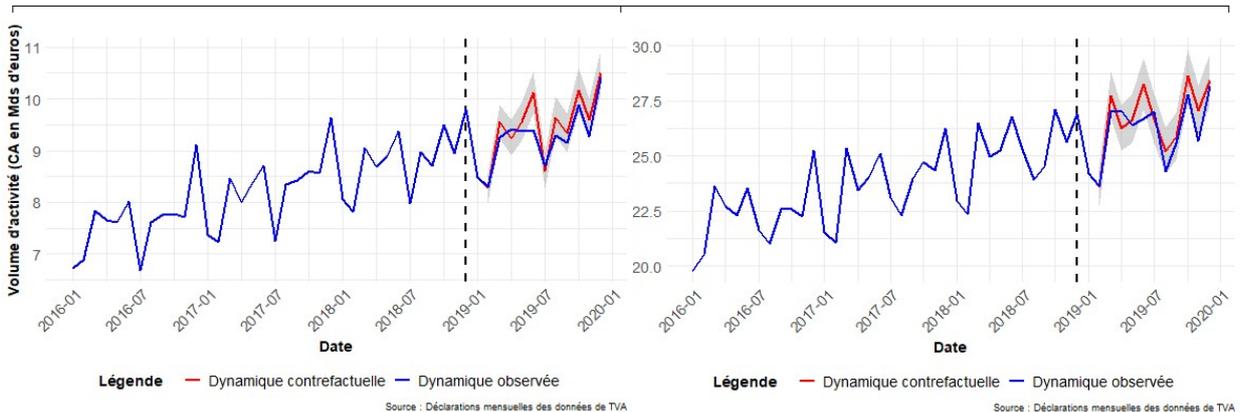
C3 – Fabrication électronique/PME **C4 – Fabrication matériel de transport/ETI-GE**

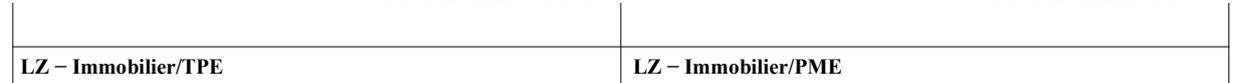
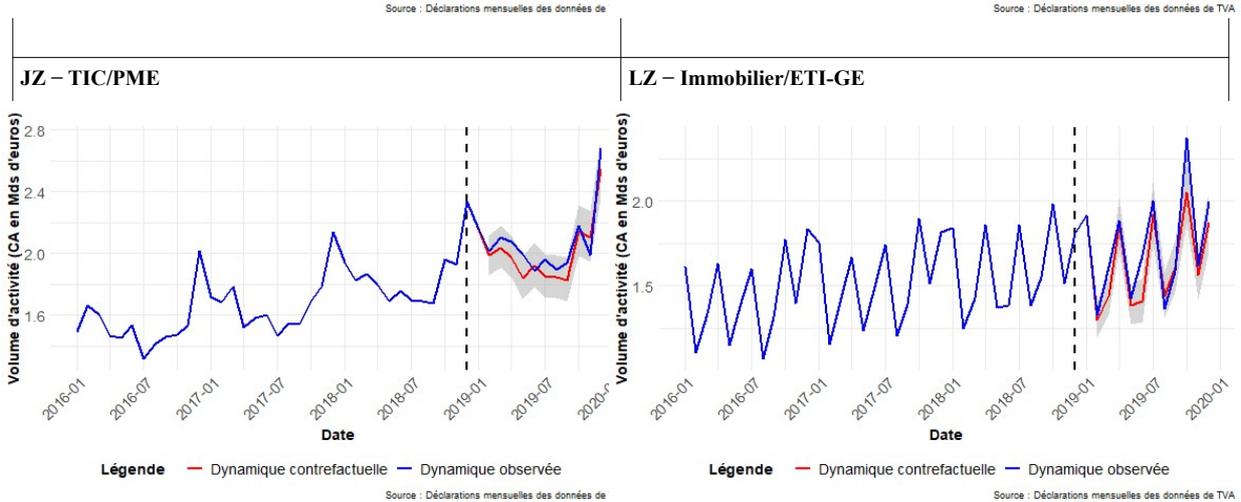
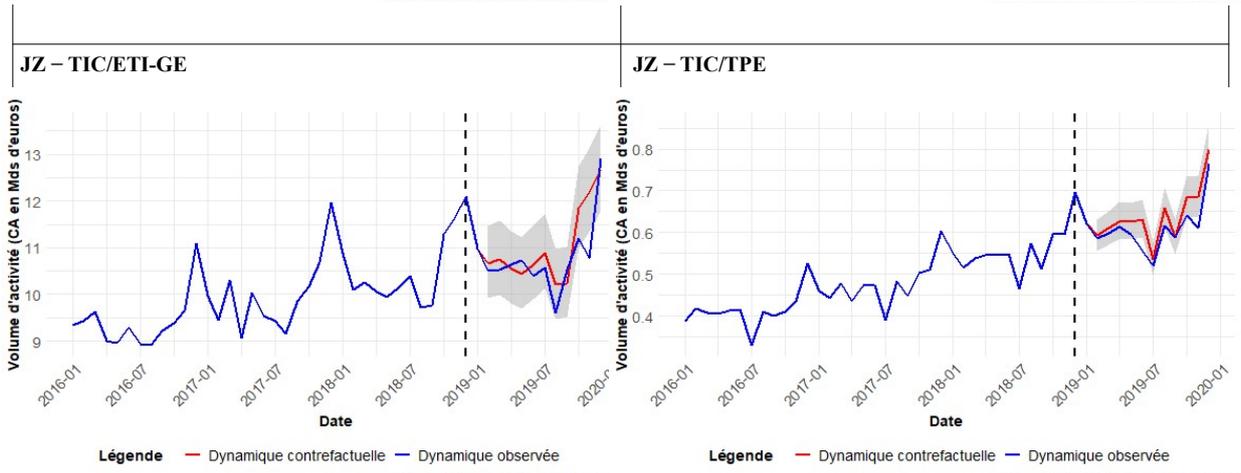
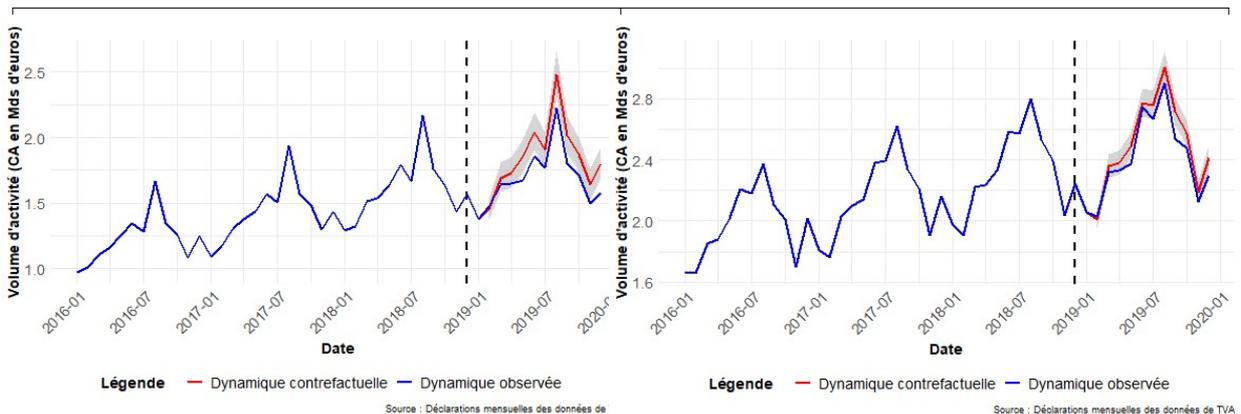


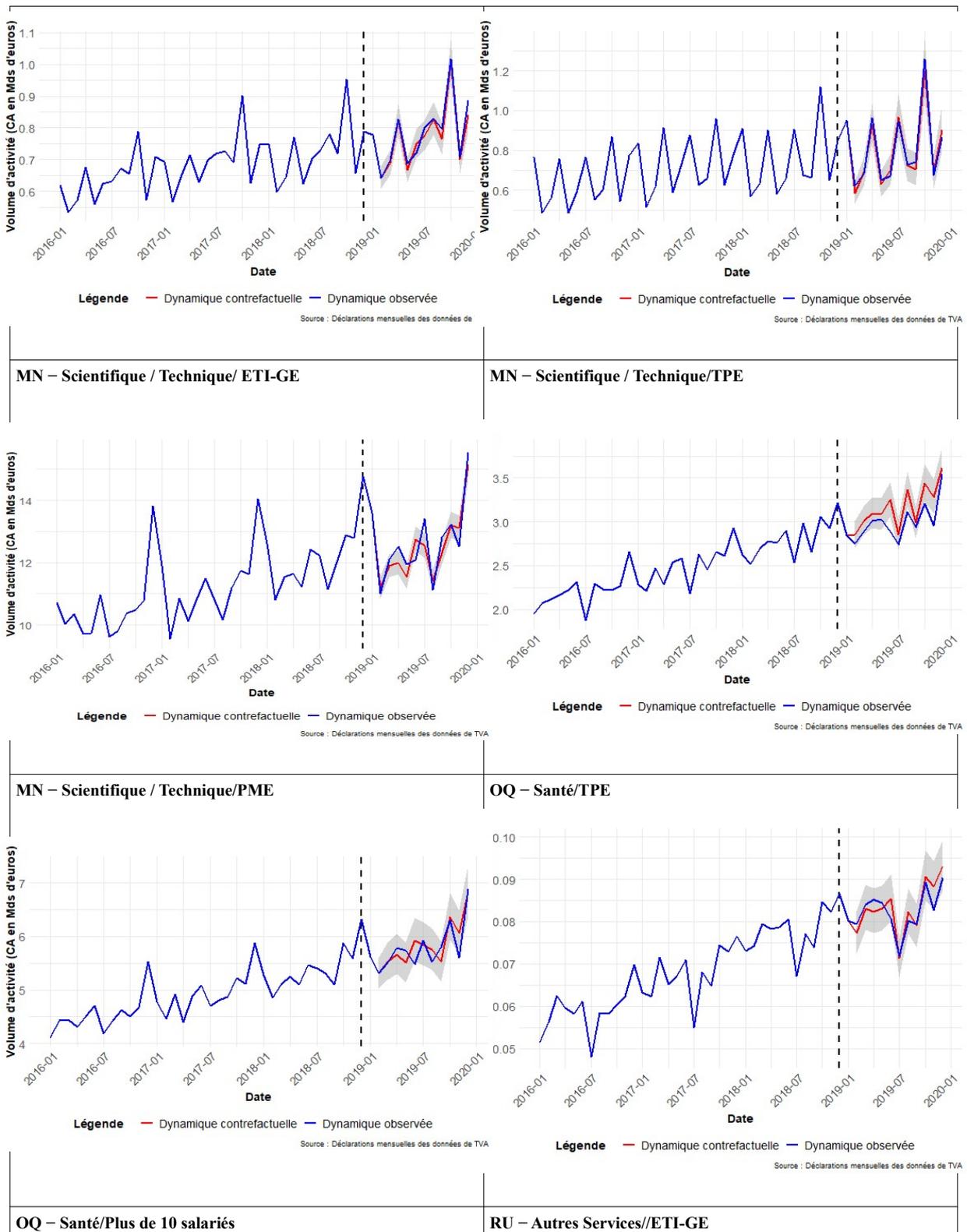
C4 – Fabrication matériel de transport/TPE **C4 – Fabrication matériel de transport/PME**

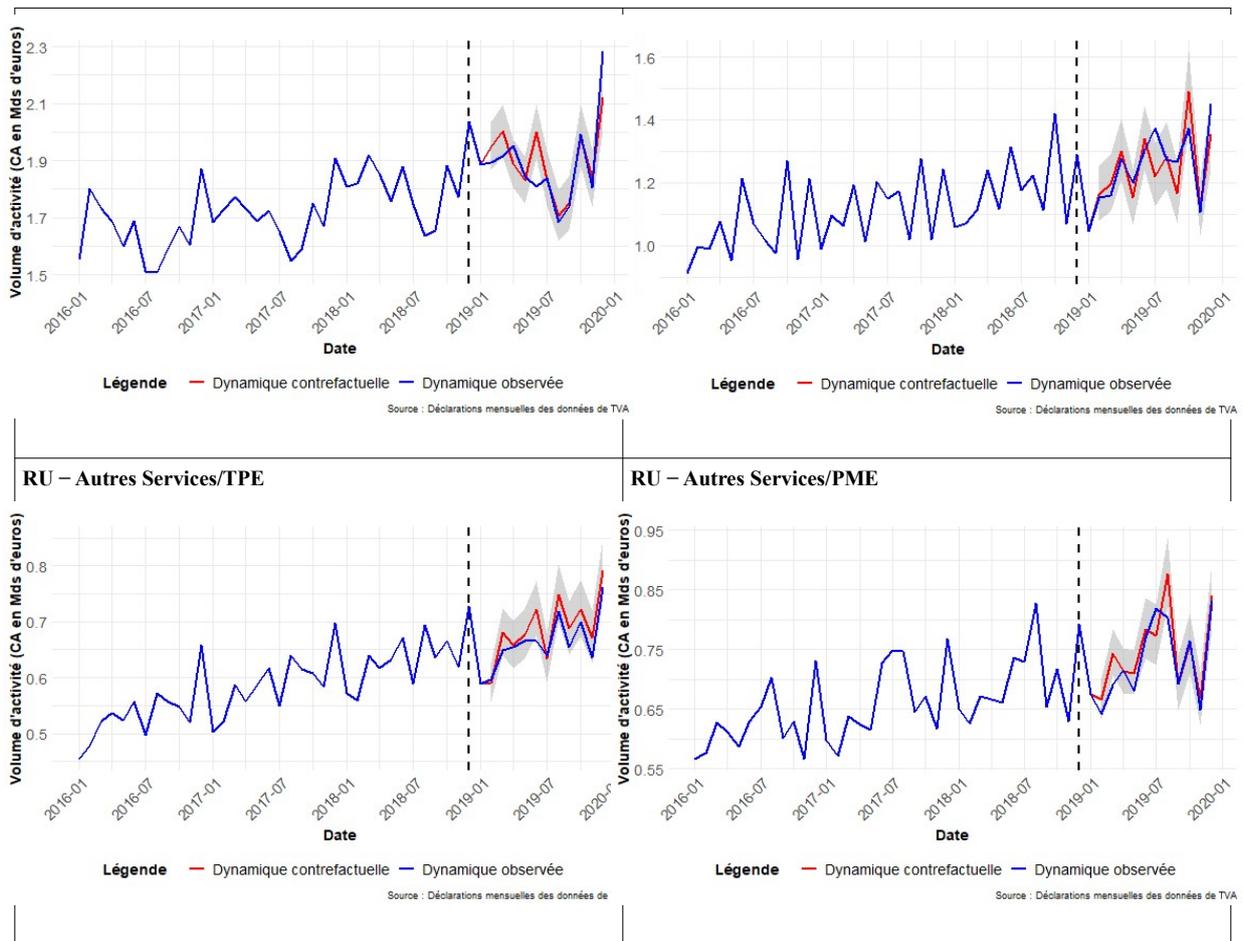








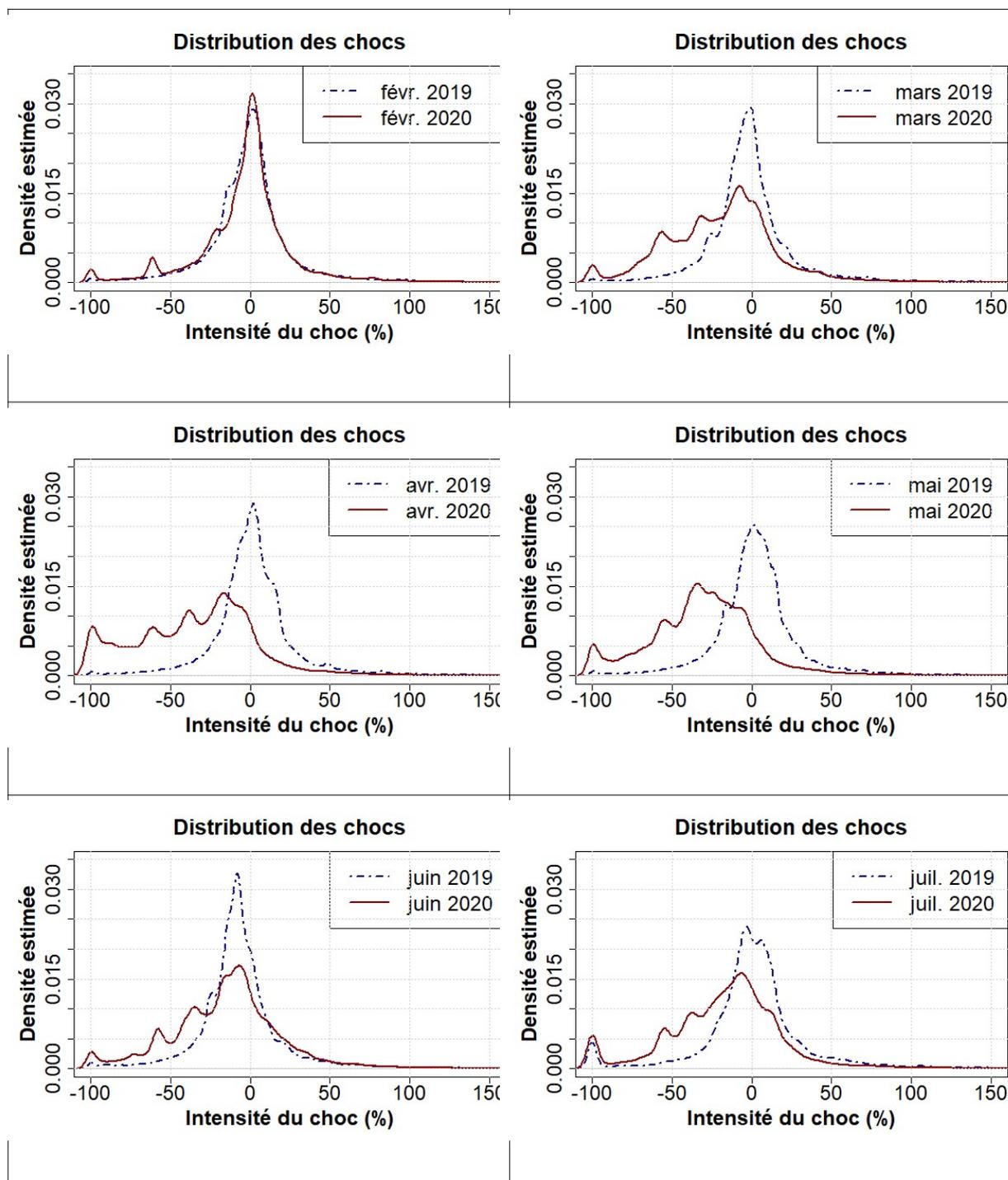


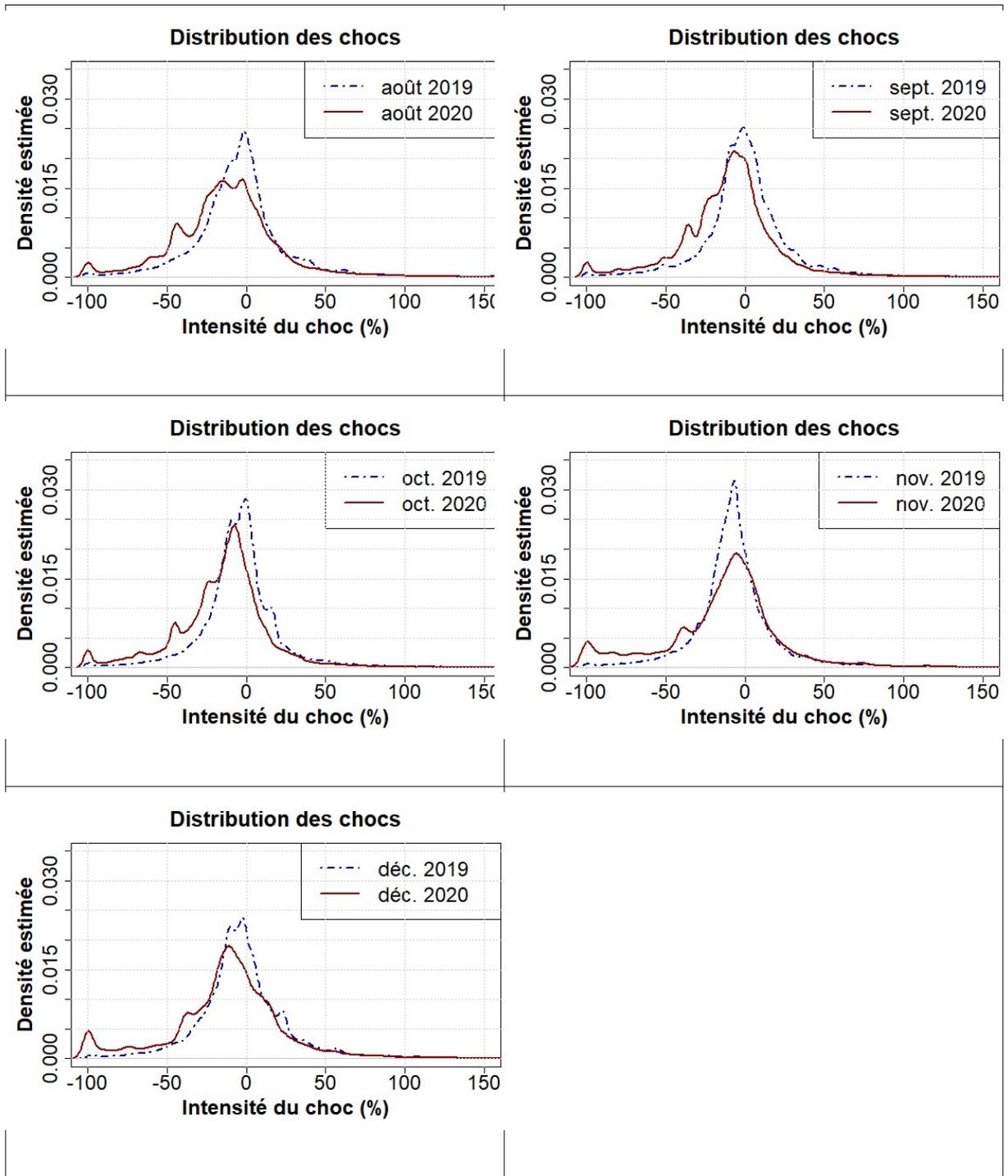


b. Chocs individuels

i. Chocs exprimés en taux de variation

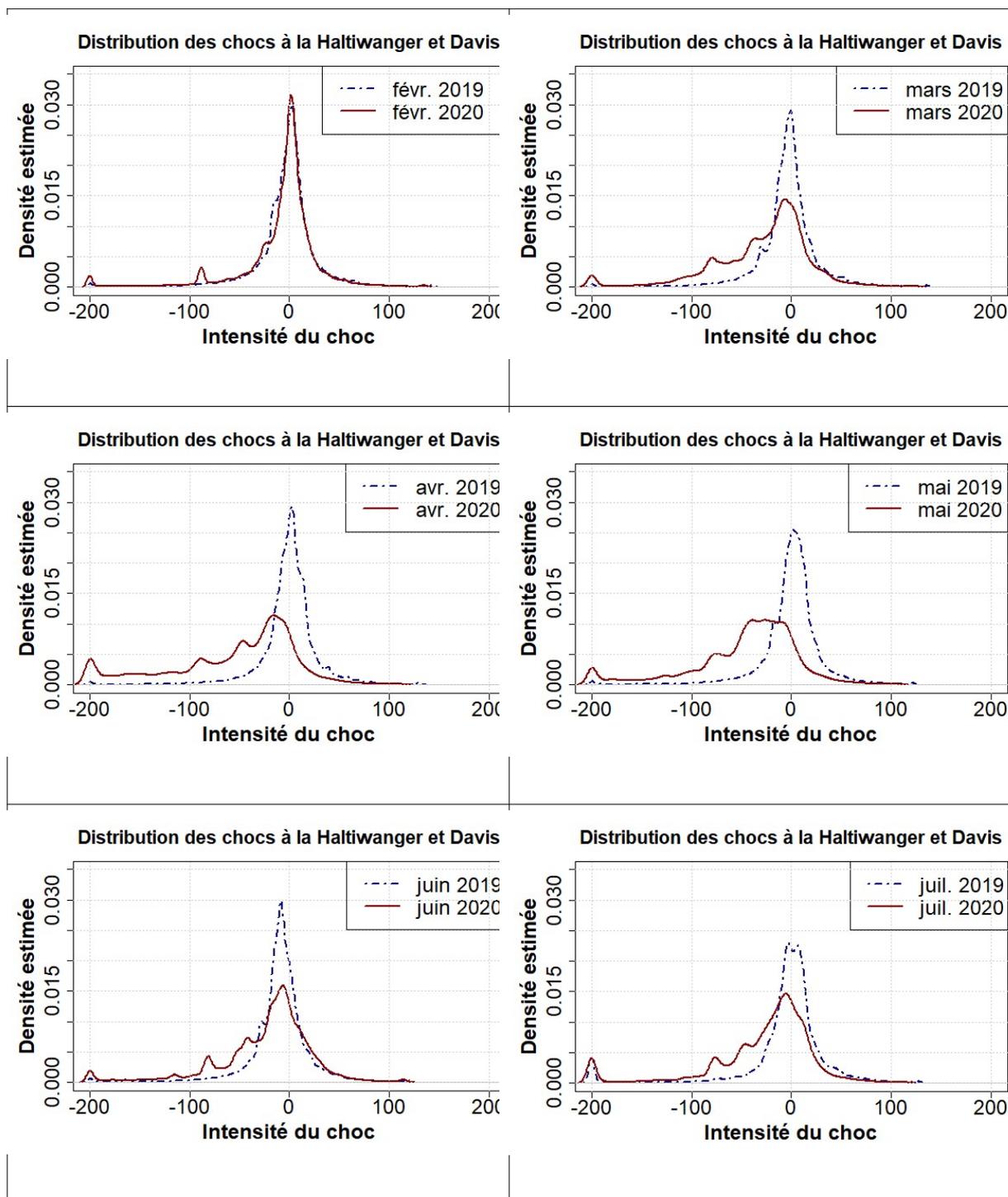
Figure F.2 – Distribution des chocs exprimés en taux de variation par mois

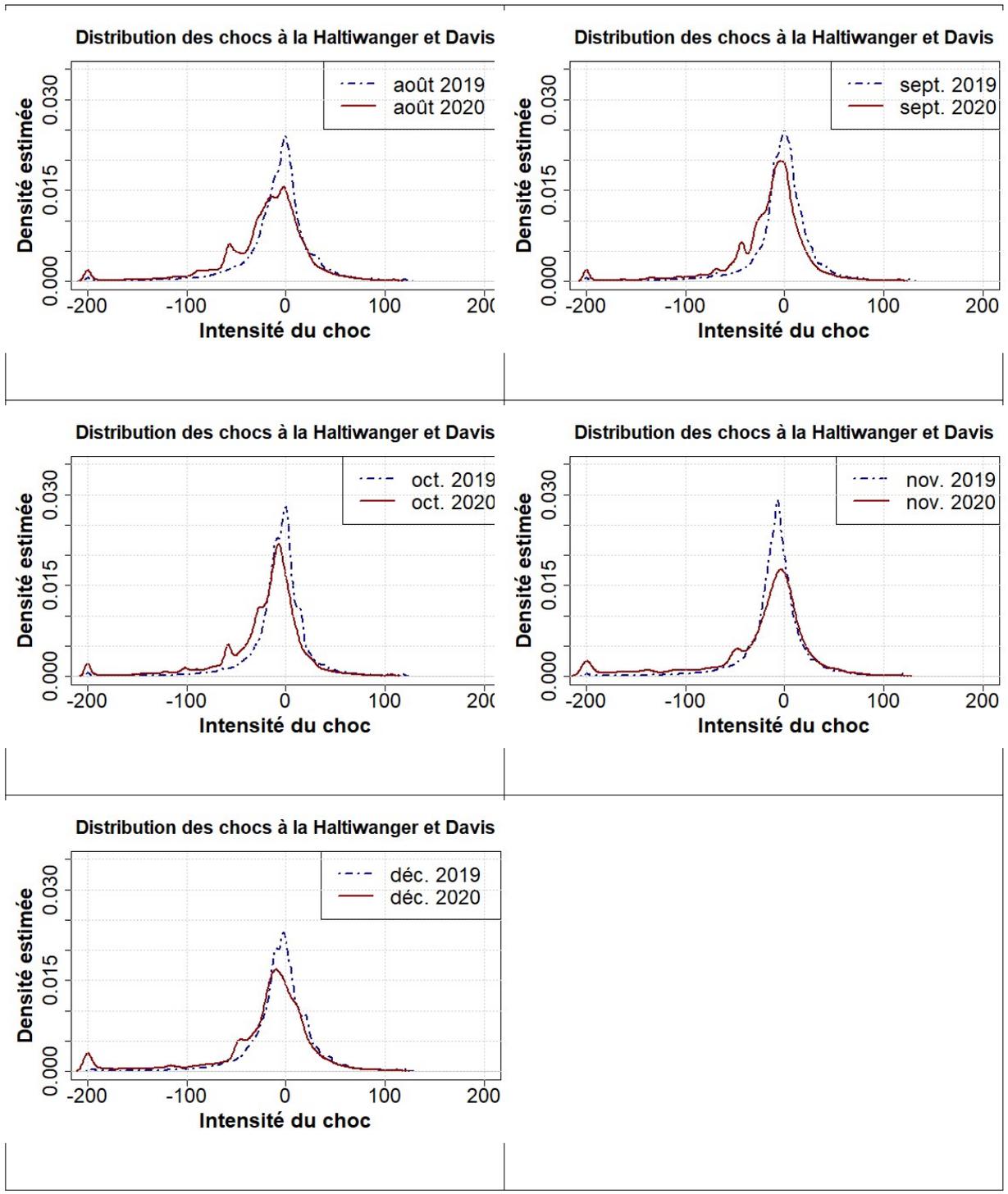




ii. Chocs à la Haltiwanger et Davis

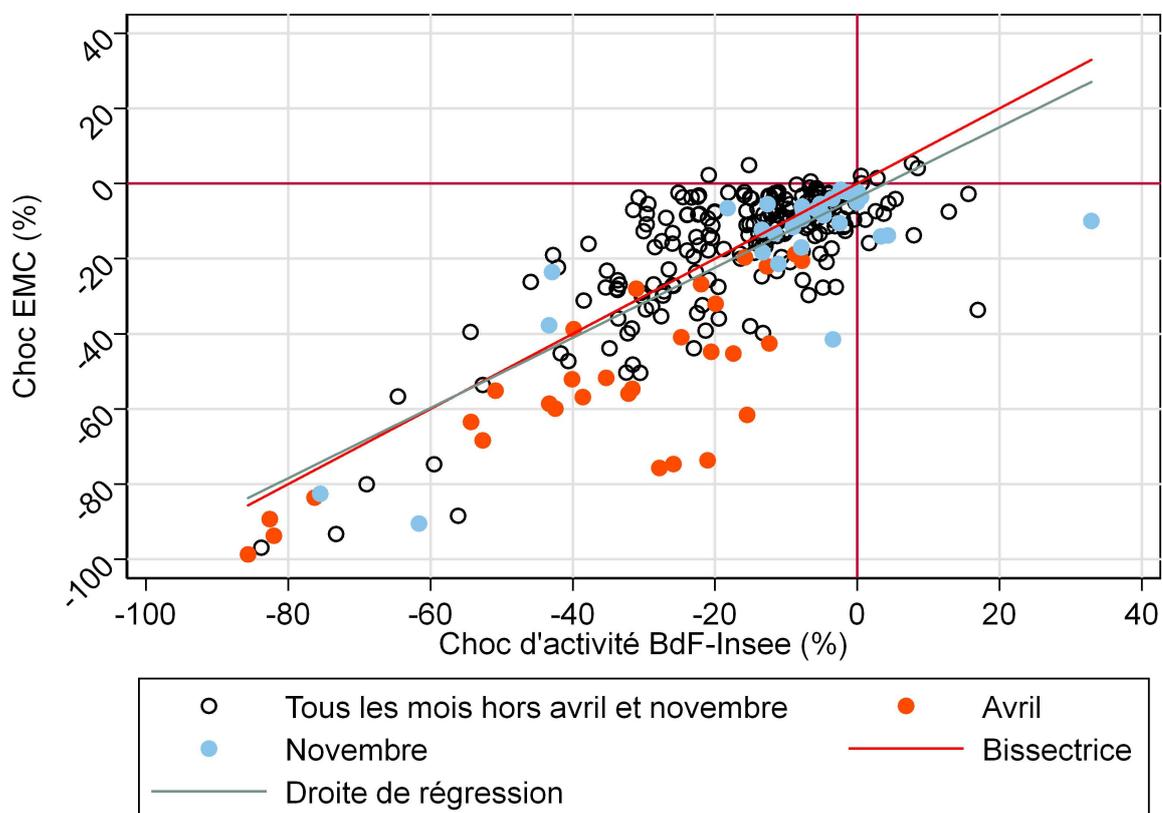
Figure F.3 – Distribution des chocs à la Haltiwanger et Davies par mois





2. Cohérence des chocs estimés avec les données d'enquête

Figure F.4 – Cohérence entre les chocs d'activité de l'étude et les chocs de l'Enquête Mensuelle de Conjoncture (EMC) de la Banque de France



Source : Déclarations TVA (DGFIP) ; Résultats de l'enquête EMC (Banque de France).

3. Cohérence des fluctuations estimées de la valeur ajoutée avec les données des comptes nationaux

Les données des comptes nationaux renseignent les montants de valeur ajoutées pour l'ensemble des sociétés non-financières. Sous l'hypothèse que notre échantillon couvre environ 71 % du total de la valeur ajoutée de l'ensemble des sociétés non-financières et ce de manière représentative, il est possible d'obtenir un montant de valeur ajoutée issu des comptes nationaux pour les années 2018, 2019 et 2020, respectivement 869, 905 et 829 milliards. La valeur ajoutée contrefactuelle issue des comptes nationaux est obtenue en appliquant un taux de croissance contrefactuel au montant de la VA correspondant à l'année 2020. A cette fin, deux taux de croissance contrefactuels sont utilisés : le premier en répliquant le taux de croissance observé entre 2018 et 2019 dans les comptes nationaux soit +4,5 %, le second en extrapolant la croissance observée entre 2017 et 2019 à l'année 2020 :

$$Taux_{2020-2019} = \frac{Taux_{2019-2018}}{Taux_{2018-2017}} \cdot Taux_{2019-2018} .$$

En ce qui concerne la valeur ajoutée issue de l'enquête, elle est estimée à partir des données Fare. La valeur de la VA en 2018 vaut le montant de la VA des sociétés non-financières dans Fare 2018 multiplié par le taux de couverture de l'échantillon. On obtient ensuite une valeur estimée en 2019 en multipliant la valeur obtenue en 2018 par le taux de croissance observé entre 2018 et 2019 dans les comptes nationaux. La valeur ajoutée de l'année 2020, qu'elle soit contrefactuelle ou observée, est obtenue en sommant la valeur ajoutée de l'ensemble des entreprises, cette dernière étant calculée selon une méthode de microsimulation financière détaillée dans [Bureau et al, 2021a] et [Bureau et al, 2021b], . *In fine*, la comparaison entre la valeur ajoutée issue de l'étude et des comptes nationaux est rassurante, avec des taux de croissances proches entre 2019 et 2020 (Tableau F.4).

Tableau F.4 – Cohérence de la valeur ajoutée entre les comptes nationaux et l'étude

	2018	2019	2020	2020-Contrefactuel
VA Etude	831,9Md	869Md	770Md	926Md
VA Comptes nationaux (périmètre équivalent)	865Md	905Md	829Md	<ul style="list-style-type: none"> • 946Md (scenario 1) • 963Md (scenario 2)
Taux de croissance annuel (Etude)		+4,5 %	-11,4 %	+6,3 %
Taux de croissance annuel (Comptes nationaux)		+4,5 %	-8,4 %	<ul style="list-style-type: none"> • +4,5 % (scenario 1)

				• +6,4 % (scenario 2)
--	--	--	--	-----------------------------

G. Décomposition de la variance : démonstration

Soit $Choc_{i,s}$ désignant le choc subi par la firme i appartenant au secteur s . Soit n le nombre de firmes dans l'échantillon, S le nombre de secteurs $s=1, \dots, S$ et n_s le nombre de firme dans le secteur s :

$n = n_1 + \dots + n_k + \dots + n_S$. On omet la dimension temporelle sans perte de généralité. On écrit la variance globale :

$$V = Var(Choc_{i,s}) = \frac{1}{n} \sum_i (Choc_{i,s} - \overline{Choc})^2$$

Avec :

$$\overline{Choc} = \frac{1}{n} \sum_i Choc_{i,s}$$

la moyenne des chocs sur l'ensemble de l'échantillon.

En notant :

$$\overline{Choc}_s = \frac{1}{n_s} \sum_{i \in s} Choc_{i,s}$$

la moyenne sectorielle de chocs pour chacun des secteurs s , l'équation se réécrit :

$$\begin{aligned} V &= \frac{1}{n} \sum_i (Choc_{i,s} - \overline{Choc})^2 = \frac{1}{n} \sum_i (Choc_{i,s} - \overline{Choc}_s + \overline{Choc}_s - \overline{Choc})^2 \\ &\Leftrightarrow V = \frac{1}{n} \sum_i \left[(Choc_{i,s} - \overline{Choc}_s)^2 + 2 \cdot (Choc_{i,s} - \overline{Choc}_s) \cdot (\overline{Choc}_s - \overline{Choc}) + (\overline{Choc}_s - \overline{Choc})^2 \right] \\ &\Leftrightarrow V = \underbrace{\frac{1}{n} \sum_i (Choc_{i,s} - \overline{Choc}_s)^2}_{\text{Terme I}} + \frac{2}{n} \cdot \sum_i (Choc_{i,s} - \overline{Choc}_s) \cdot (\overline{Choc}_s - \overline{Choc}) + \underbrace{\frac{1}{n} \sum_i (\overline{Choc}_s - \overline{Choc})^2}_{\text{Terme B}} \end{aligned}$$

Soit le terme central :

$$\sum_i (Choc_{i,s} - \overline{Choc}_s) \cdot (\overline{Choc}_s - \overline{Choc}) = \sum_i (Choc_{i,s} - \overline{Choc}_s) \cdot \overline{Choc}_s - \sum_i (Choc_{i,s} - \overline{Choc}_s) \cdot \overline{Choc}$$

Ce terme vaut 0. Pour le montrer, il suffit de montrer que :

$$\sum_i (Cho c_{i,s} - \overline{Cho c_s}) = 0$$

En effet :

$$\sum_i (Cho c_{i,s} - \overline{Cho c_s}) = \sum_s \sum_{i \in s} (Cho c_{i,s} - \overline{Cho c_s})$$

$$\sum_s \sum_{i \in s} Cho c_{i,s} - \sum_s n_s \cdot \overline{Cho c_s} = \sum_s \sum_{i \in s} Cho c_{i,s} - \sum_s \sum_{i \in s} Cho c_{i,s} = 0$$

Donc : $V = I + B$

Soit le terme I :

$$I = \frac{1}{n} \sum_i (Cho c_{i,s} - \overline{Cho c_s})^2 = \frac{1}{n} \sum_s \sum_{i \in s} (Cho c_{i,s} - \overline{Cho c_s})^2 = \frac{1}{n} \sum_s n_s \cdot \frac{1}{n_s} \sum_{i \in s} (Cho c_{i,s} - \overline{Cho c_s})^2$$

$$\Leftrightarrow I = \sum_s \frac{n_s}{n} \cdot Var_s(Cho c_{i,s})$$

avec :

$$Var_s(Cho c_{i,s}) = \frac{1}{n_s} \sum_i (Cho c_{i,s} - \overline{Cho c_s})^2$$

la variance intra-sectorielle de chaque secteur s .

Autrement dit, le terme I n'est rien d'autre que la moyenne, sur l'ensemble des secteurs $s=1, \dots, S$, des va-

riances intra-sectorielle pondérée par le poids de chaque secteur $\frac{n_s}{n}$.

Soit le terme B :

$$B = \frac{1}{n} \sum_i (\overline{Cho c_s} - \overline{Choc})^2 = \frac{1}{n} \sum_s \sum_{i \in s} (\overline{Cho c_s} - \overline{Choc})^2 = \sum_s \frac{n_s}{n} \cdot (\overline{Cho c_s} - \overline{Choc})^2$$

$$\Leftrightarrow B = \sum_s \frac{n_s}{n} \cdot (\overline{Cho c_s} - \overline{Choc})^2$$

Autrement dit, le terme B n'est rien d'autre qu'une forme de variance des moyennes sectorielles (pondérée du

poids de chaque secteur $\frac{n_s}{n}$).

In fine, on obtient donc cette décomposition de la variance globale :

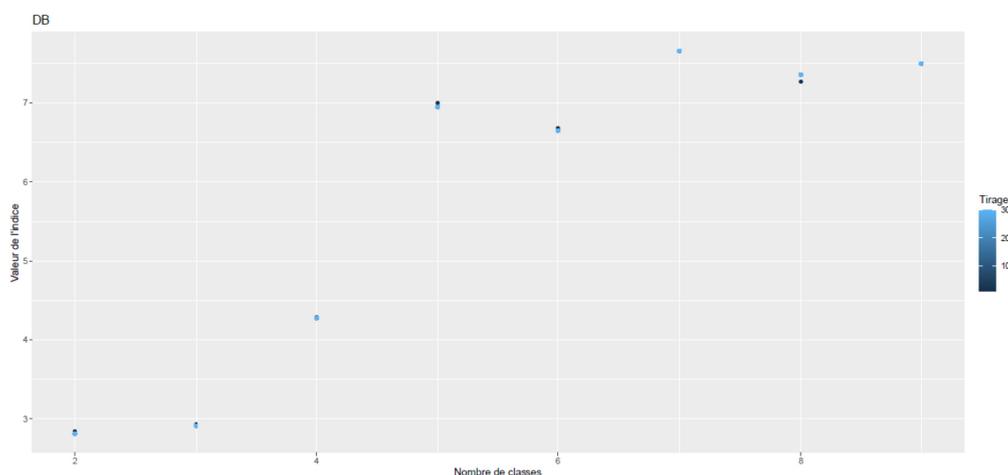
$$V = Var(Cho c_{i,s}) = I + B = \underbrace{\sum_s \frac{n_s}{n} \cdot Var_s(Cho c_{i,s})}_{\text{Variance intra-classe}} + \underbrace{\sum_s \frac{n_s}{n} \cdot (\overline{Cho c_s} - \overline{Choc})^2}_{\text{Variance inter-classe}}$$

H. Choix du nombre de classes

Les tests de qualité de la classification de courbe reposent sur le calcul d'une matrice de distance entre tous les éléments de la population, ce qui s'avère trop coûteux avec un échantillon de travail aussi important (matrice de dimension 645 000 x 645 000). Afin de pouvoir déterminer le nombre de classes le mieux adapté aux données, le recours à un bootstrap a donc été nécessaire : 30 échantillons de 30 000 entreprises ont été tirés, et des partitions de 2 à 9 classes ont été réalisées pour chaque tirage.

Une fois ces partitions effectuées, elles ont été comparées, pour chaque tirage, selon le critère de validation de Davies-Bouldin³⁰ – moyenne des rapports maximaux entre la distance d'une série au centre de sa classe et la distance entre les centres de classes [Davies et Bouldin, 1979]. Le nombre de classes pour chaque tirage a été sélectionné par critère du coude : le nombre de classes choisi est le plus grand nombre de classes permettant de satisfaire le critère de validation.

Figure H.1 – Choix du nombre de classes par critère de Davies-Bouldin



Note : Pour chacun des 30 échantillons de 30 000 entreprises tirés dans l'échantillon de travail, des partitions en 2 à 9 classes ont été réalisées. Ce graphique présente l'indice de Davies-Bouldin calculé pour les différents nombres de classes, avec un point par nombre de classe et par tirage.

Lecture : Pour tous les tirages, l'indice de Davies-Bouldin d'une partition en 4 classes est aux alentours de 4,3.

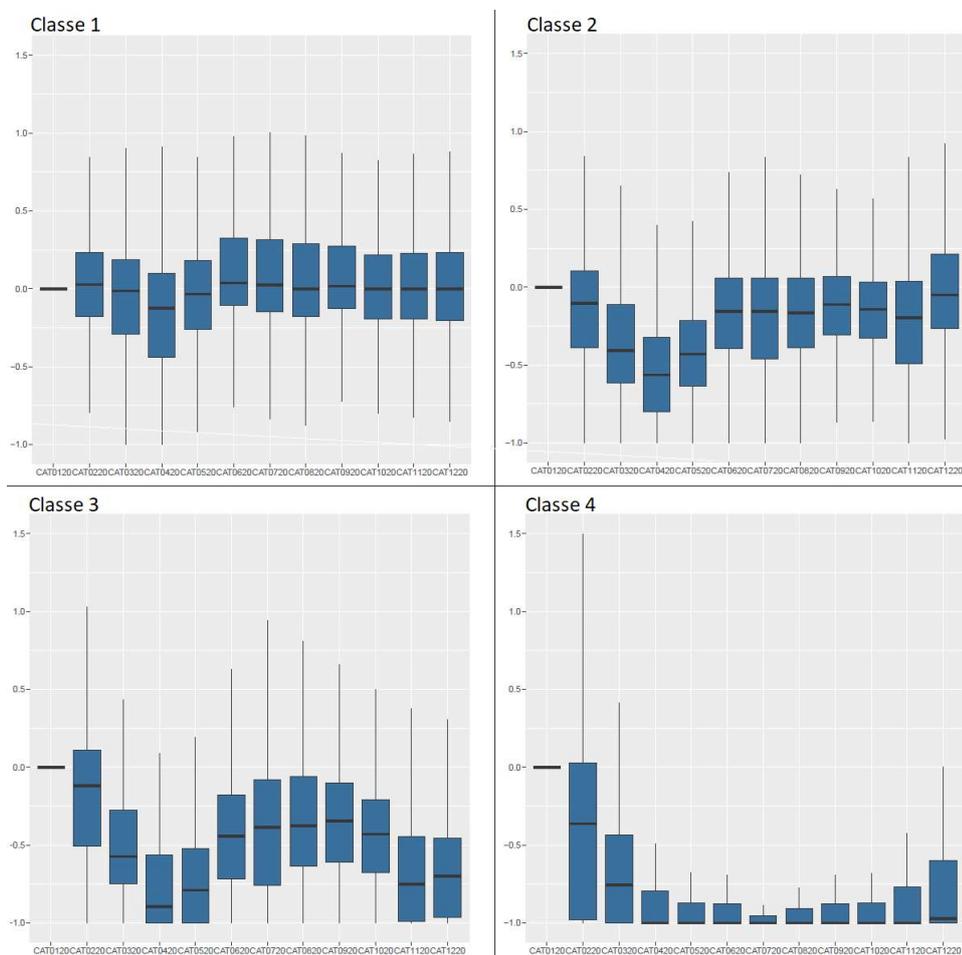
Les indices calculés pour des partitions en 2 et 3 classes sont presque égaux. Les indices calculés pour une partition en 4 classes sont légèrement plus élevés mais restent comparables à ceux d'un partition en 3 classes, on préfère donc une partition en 4 classes. En revanche, l'indice présente un saut important, soit une forte perte de qualité, entre les partitions en 4 classes et celles en 5 classes. Une partition en 4 classes donc est préférée ici car elle permet d'obtenir une partition plus fine tout en présentant une qualité de clustering satisfaisante.

30 Les critères de validation de la silhouette [Rousseeuw, 1987] et du score n'ont pas été retenus ici, car le nombre de classe influençait peu leur valeur.

I. Distribution des chocs au sein des classes

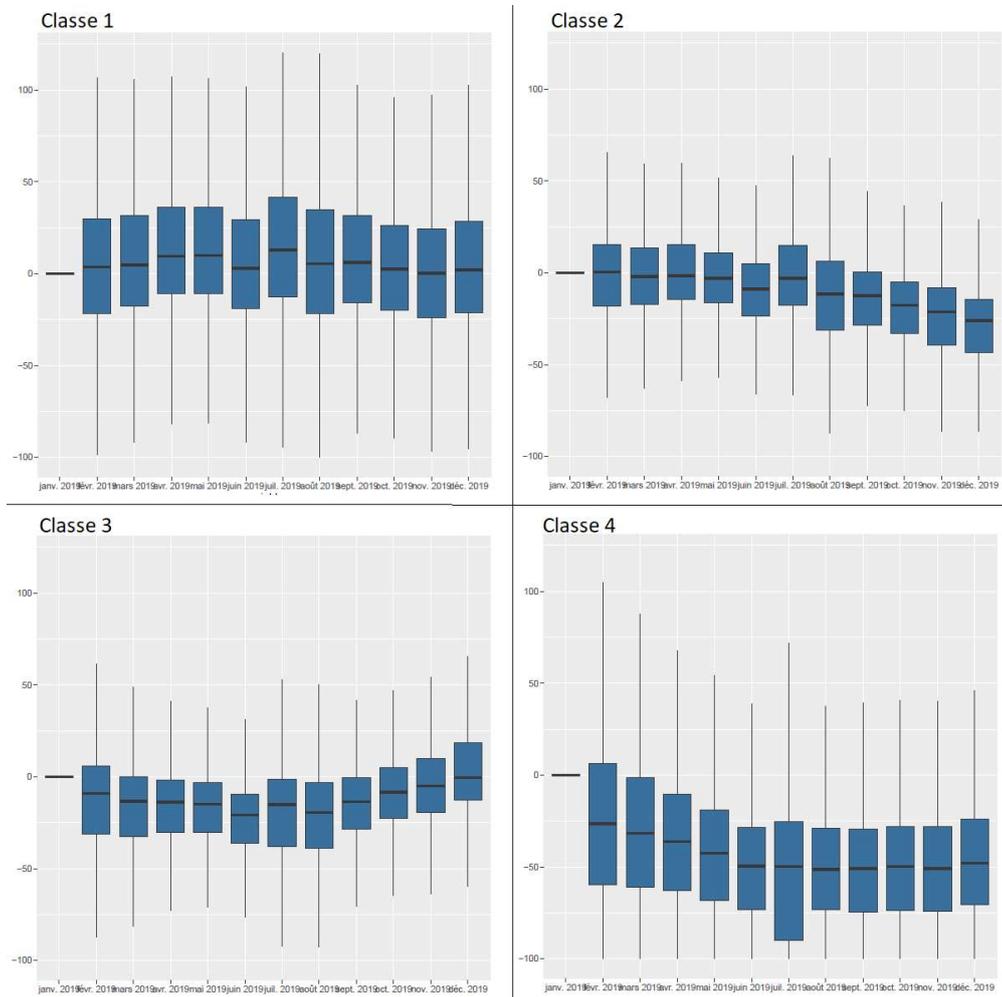
1. En 2020

Figure I.1 – Distribution des chocs mensuels par classes en 2020



2. En 2019

Figure I.2 – Distribution des chocs mensuels par classe, sur l'exercice reproduit en 2019



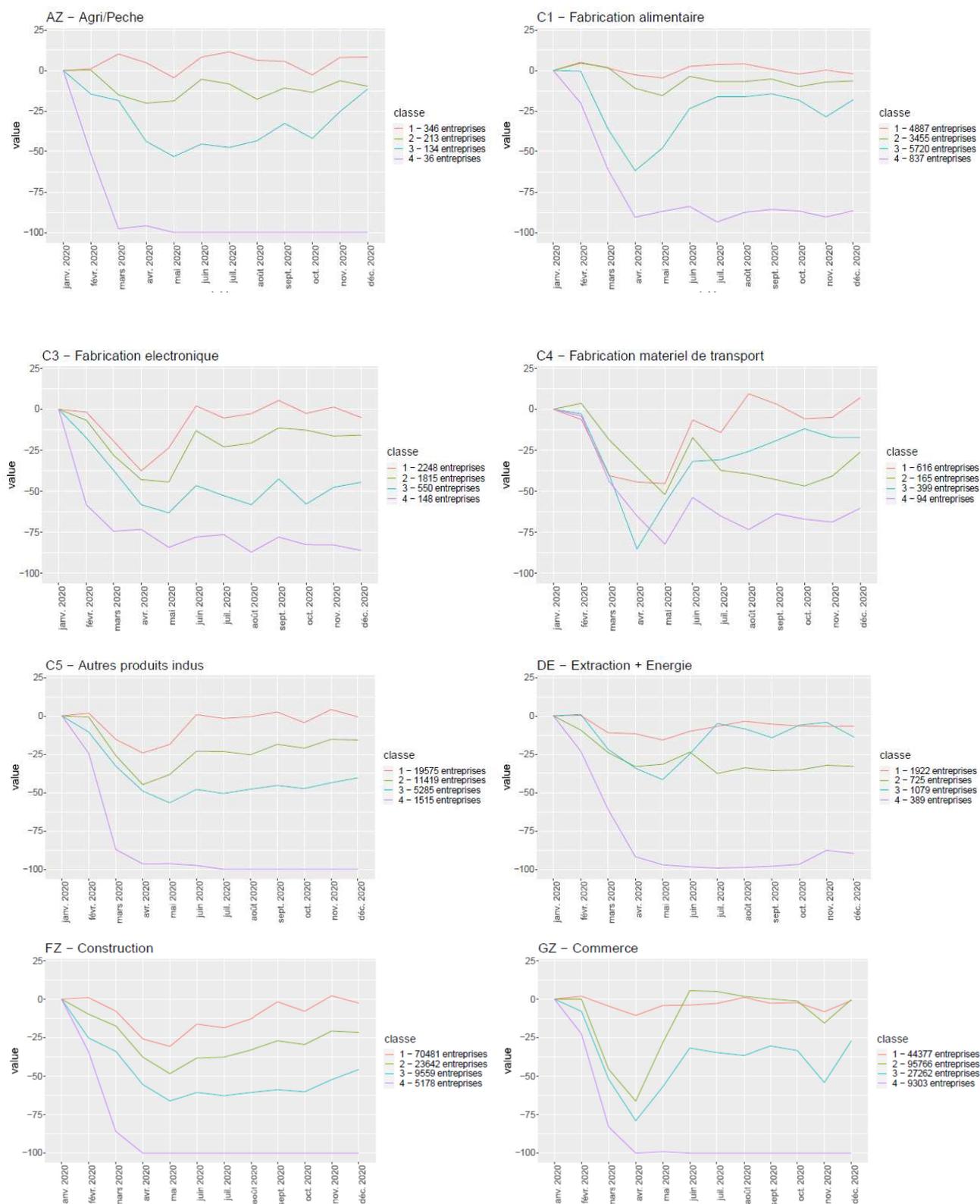
Note : Cette partition a été réalisée sur les chocs issus de la reproduction de l'exercice sur 2019. Elle est indépendante de celle présentée sur 2020.

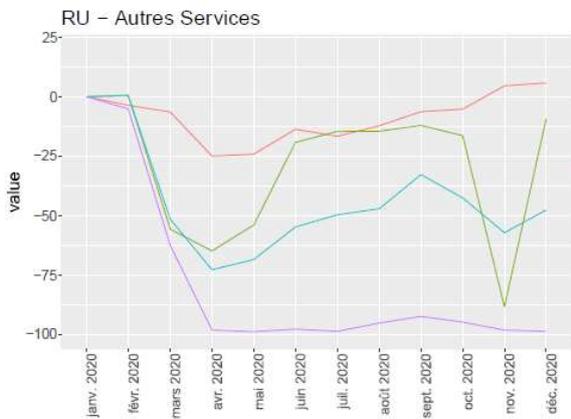
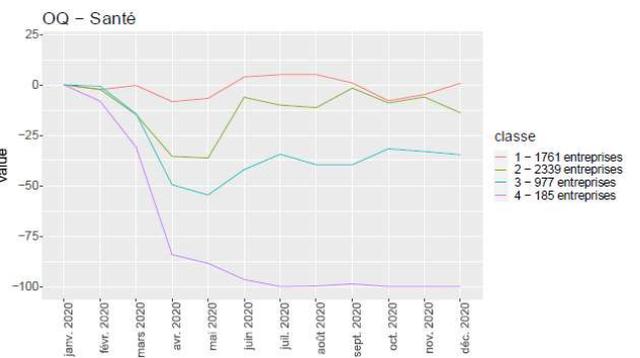
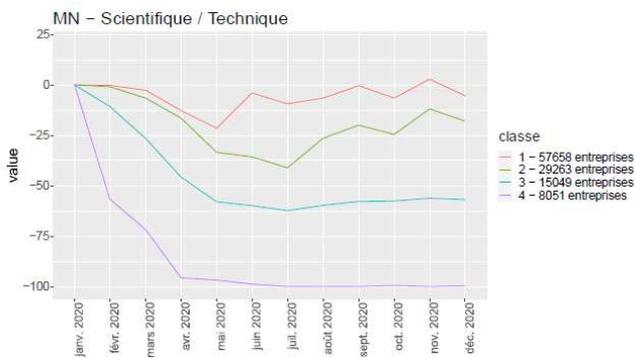
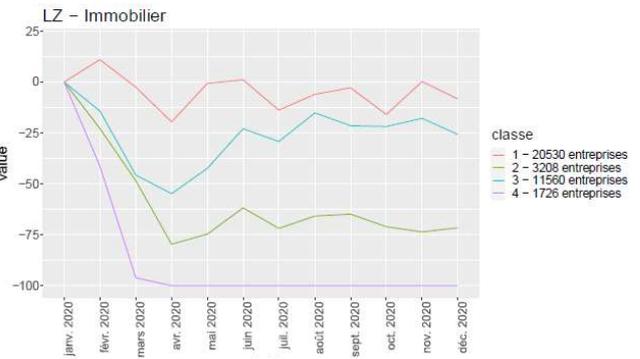
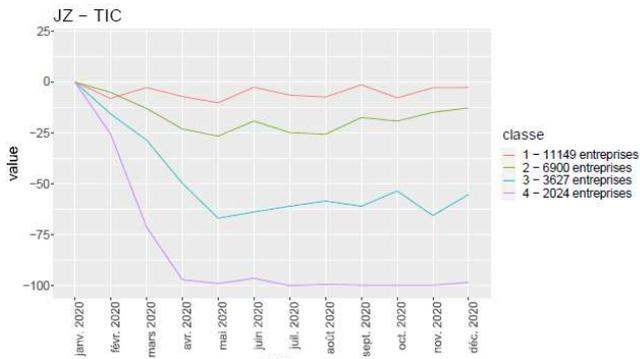
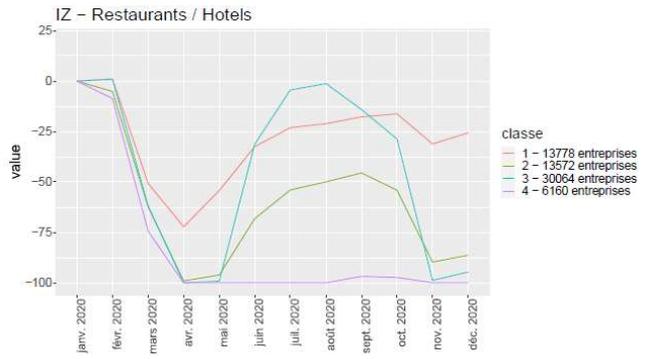
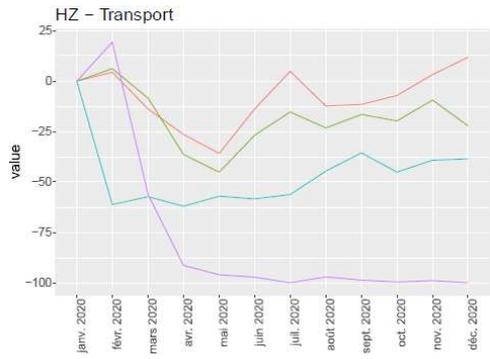
J. Trajectoires médianes de choc au sein des classes estimées sur les secteurs d'activité A17

La partition du secteur C2 - « Cokéfaction-Raffinage » n'est pas présentée ici car ce secteur contient trop peu d'entreprises.

Tableau J.1 – Trajectoires médianes de choc au sein des classes estimées sur les secteurs d'activité

A17





K. Modèle de classification

Le modèle mobilisé pour étudier la répartition des entreprises entre les différents profils de trajectoires identifiés consiste à ramener la ventilation de ces dernières à un problème de choix discret dans lequel chacune d'entre elle se voit attribuée un groupe parmi un ensemble exhaustif d'alternatives non-superposables.

La modélisation par un logit multinomiales permet d'obtenir des probabilités comprises entre 0 et 1 et dont la somme entre toutes les alternatives vaut 1. L'une des propriétés du modèle logit multinomial (MNL) est qu'il peut être représenté par un réseau neuronal qui lui est équivalent, mais qui est établi sous un ensemble beaucoup plus restreint d'hypothèses [Bentz et Merunka, 2000]. Le modèle MNL correspond ainsi à un réseau neuronal feedforward avec une fonction de sortie Softmax, avec des poids partagés et sans couche cachée [Bridle, 1990]. Dans ce type de réseau de neurones, l'entropie, *i.e.* la fonction d'erreur utilisée est identique à la fonction de log-vraisemblance du MNL. Il est donc possible d'estimer un MNL par maximum de vraisemblance en utilisant des réseaux de neurones [Venables et Ripley, 2013]. La méthode d'optimisation utilisée est la méthode de Broyden-Fletcher-Goldfarb-Shanno (BFGS) qui s'appuie sur la descente de gradient. Le nombre d'itérations maximales a été arbitrairement fixée à 300 et est suffisamment élevé pour permettre au modèle de converger. Les écarts-types sont estimés par inversion de la matrice hessienne finale de log-vraisemblance et sont ensuite utilisés pour calculer la significativité statistique de tous les paramètres du modèle. Cette stratégie d'estimation peut être mise en œuvre à l'aide du package R *nnet* [Ripley *et al.*, 2016]. Les variables explicatives utilisées dans le modèle sont le secteur d'activité, la taille des entreprises, leur date de création (en 4 modalités), l'existence d'une activité exportatrice et des variables relatives au développement de la vente en ligne, de systèmes de livraison, de nouveaux produits ou services, la réorganisation de l'activité, la mise en commun de moyens avec d'autres entreprises et la réalisation d'investissements spécifiques dans les nouvelles technologies pendant la crise.

Le Tableau K.1 présente le résultat des estimations. Pour des raisons de mise en page, les coefficients correspondants aux différentes modalités du secteur d'activité ne sont pas reportés sur ce tableau.

Tableau K.1-Résultats de la régression du modèle mlogit

	<i>Variable dépendante :</i>		
	Profil 2	Profil 3	Profil 4
Taille : TPE	0.330 ^{***} (0.027)	0.786 ^{***} (0.044)	-0.178 ^{***} (0.054)
Taille : Petites et Moyennes Entreprises	0.224 ^{***} (0.027)	0.447 ^{***} (0.044)	0.090 [*] (0.054)
Date de création :Avant 1997	0.385 ^{***} (0.009)	0.251 ^{***} (0.013)	-0.020 (0.022)
Date de création :Entre 1998 et 2006	0.481 ^{***} (0.009)	0.780 ^{***} (0.012)	0.091 ^{***} (0.020)
Date de création :Entre 2007 et 2012	0.343 ^{***} (0.009)	-0.098 ^{***} (0.012)	0.229 ^{***} (0.019)
Activité exportatrice	0.108 ^{***} (0.010)	0.103 ^{***} (0.014)	0.566 ^{***} (0.027)
Dév. de la vente en ligne	-0.598 ^{***} (0.013)	-1.375 ^{***} (0.022)	-1.304 ^{***} (0.037)
Dév. de nouveaux systèmes de livraison	-0.525 ^{***} (0.013)	-1.049 ^{***} (0.022)	-0.703 ^{***} (0.036)
Dév. de nouveaux produits/activités/services	0.035 ^{***} (0.011)	0.219 ^{***} (0.016)	-3.410 ^{***} (0.063)
Investissement dans de nouvelles technologies	-0.553 ^{***} (0.020)	-1.141 ^{***} (0.029)	-1.784 ^{***} (0.042)
Réorganisation de l'activité	-0.632 ^{***} (0.013)	-0.797 ^{***} (0.021)	-0.194 ^{***} (0.035)
Mise en commun de moyens	-0.068 ^{***} (0.013)	-1.057 ^{***} (0.022)	0.283 ^{***} (0.026)
Effectifs en télétravail	-0.004 ^{***} (0.0002)	-0.008 ^{***} (0.0002)	-0.009 ^{***} (0.0004)
Constante	-7.878 ^{***} (0.038)	-7.120 ^{***} (0.057)	-7.748 ^{***} (0.072)
Akaike Inf. Crit.	1,199,764.000		

Note:

*p<0.1; ** p<0.05; *** p<0.01

L. Distance de Hellinger

La distance de Hellinger est une mesure permettant de quantifier la similarité entre deux distributions de probabilités. Elle fait partie de la famille plus générale des *f-divergences*.

En notant P et Q les mesures de probabilité absolument continues par rapport à une troisième mesure λ , le carré de la distance de Hellinger entre P et Q vaut :

$$H^2(P, Q) = \frac{1}{2} \int \left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}} \right)^2 d\lambda$$

En prenant pour λ la mesure de Lebesgue, $\frac{dP}{d\lambda}$ et $\frac{dQ}{d\lambda}$ désignent simplement des densités de probabilité. On peut les noter f et g et réécrire le carré de la distance de Hellinger :

$$H^2(f, g) = \frac{1}{2} \int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx = 1 - \int \sqrt{f(x)g(x)} dx$$

La racine carrée de cette distance donne la distance de Hellinger.

Cette distance a été choisie pour plusieurs raisons. D'une part, c'est l'une des formules de référence dans la comparaison de distributions statistiques deux à deux. D'autre part, elle est bornée entre 0 et 1. Son interprétation en est ainsi facilitée. Enfin, sa relative simplicité et facilité de manipulation vont dans le sens d'un arbitrage en sa faveur.

A des fins de robustesse, l'analyse présentée dans l'article avec la distance de Hellinger a été reproduite avec d'autres distances statistiques (divergence de Kullback-Leibler et distance de Bhattacharyya). Les constats d'une plus forte dissimilarité des distributions des chocs entre 2019 et 2020 sur les mois les plus touchés par les mesures de restriction sanitaire sont identiques.