

---

## Note méthodologique Indicateurs Image PME – source TVA

---

Version 2.0.7.5 le 29/06/2022

### Préambule

Les indicateurs Image PME sont diffusés par le CNOEC aux Conseils régionaux de l'Ordre, aux experts-comptables et aux pouvoirs publics, et construits à partir des données de la base Statexpert. Cette base, qui appartient à ECMA est construite à partir des télédéclarations fiscales transmises aux administrations concernées par un portail de télétransmission, et ces télédéclarations sont réalisées par les experts-comptables pour le compte de leurs clients, les TPE-PME.

Cette note méthodologique traite des données issues des informations provenant des déclarations mensuelles et trimestrielles de TVA (EDI-TVA). Cette note est évolutive.

Les données concernent les entreprises de tous types de structures et de tous régimes fiscaux et les données personnelles des dirigeants contenues dans les télédéclarations fiscales transmises par les experts-comptables via les plateformes télé déclaratives. Les données personnelles des dirigeants sont « pseudonymisées » afin de respecter les dispositions légales en matière de confidentialité et de protection des données personnelles.

### 1. Définition des nouveaux indicateurs

Le flux EDI-TVA est alimenté mensuellement par la déclaration de TVA imposée sous le régime de la déclaration réel normal et trimestriel (dit mini-réel).

Dans un premier temps, les informations pertinentes à exploiter concernent le chiffre d'affaires (CA), dans sa composante géographique et sectorielle et sont mises en perspective par rapport à la période précédente (mois, trimestre, année).

#### L'indice de chiffre d'affaires (ICA)

##### Par mois (comparaison période N-1) :

Il s'agit de rapporter le CA moyen (CA total des entreprises du périmètre divisé par le nombre d'entreprises du périmètre) d'un mois par rapport au CA moyen du même mois de l'année précédente.

##### Par trimestre (comparaison période N-1)

Il s'agit de rapporter le CA moyen (CA total des entreprises du périmètre divisé par le nombre d'entreprises du périmètre) d'un trimestre par rapport au CA moyen du même trimestre de l'année précédente. Les détails des calculs sont précisés dans la partie 2. Filtre 8.

#### L'indice de chiffre d'affaires cumulé (ICAC)

Il permet de mesurer l'état d'avancement depuis le début de l'année en cours, par rapport à la même période de l'année précédente.

Il s'agit de cumuler le CA depuis le mois de janvier de l'année en cours jusqu'à celui du dernier mois disponible, de le moyenner par rapport au nombre d'entreprises concernées, puis de le rapporter au cumul du CA des mêmes mois de l'année précédente (moyenné par le nombre d'entreprises concernées). Les détails des calculs sont précisés dans la partie 3.8.

## **2. Méthodologie des filtres et jointures appliquées**

### **Filtre 1 : Neutralisation (par suppression) des SIREN qui présentent plusieurs déclarations de chiffre d'affaires pour un même mois.**

Sont neutralisés les SIREN qui ont fait plus d'une déclaration pour un même mois, au cours des 15 derniers mois, si ces déclarations concernent des montants différents pour un même mois (sous-filtre 1a).

Si une déclaration en doublon porte sur le même montant, une seule des deux lignes sera conservée dans le calcul. Idem pour le mois m-12 comparé (sous-filtre 1b).

Pour les déclarations trimestrielles, si le mois de départ n'est pas le premier mois du trimestre, alors l'entreprise est exclue (sous-filtre 1c).

### **Filtre 2 : Neutralisation (par suppression) des SIREN qui présentent des déclarations à 0 sur 6 mois ou plus + consécutifs au cours des 12 derniers mois**

Ce filtre est activable / désactivable pour la totalité de la base, un territoire ou un secteur. En effet, pour la période de crise sanitaire ou pour prévenir toute période de nouvelle crise sanitaire, confinement, fermeture temporaire d'activité ou tout autre choc économique, il convient de pouvoir réintégrer dans les calculs les entreprises ayant réalisé des déclarations à 0.

Pour le calcul des indices trimestriels, l'entreprise est neutralisée si elle réalise 2 trimestres consécutifs à 0 au cours des 4 derniers trimestres.

Pour le calcul des indices trimestriels des entreprises déclarant mensuellement, l'entreprise est neutralisée si elle réalise des déclarations à 0 sur 6 mois ou plus consécutifs au cours des 12 derniers mois.

### **Filtre 3 : neutralisation (par suppression) des SIREN qui présentent pour le mois analysé ou le même mois en N-1 une déclaration à 0**

Ce filtre est activable / désactivable pour la totalité de la base, un territoire ou un secteur. En effet, pour la période de crise sanitaire ou pour prévenir toute période de nouvelle crise sanitaire, confinement, fermeture temporaire d'activité ou tout autre choc économique, il convient de pouvoir réintégrer dans les calculs les entreprises ayant réalisé des déclarations à 0.

Pour les indices trimestriels, la même règle s'applique si l'entreprise présente une déclaration à 0 sur le trimestre ou sur le même trimestre de l'année précédente auquel il est comparé.

Pour le calcul des indices trimestriels des entreprises déclarant mensuellement, l'entreprise est neutralisée si son CA cumulé du trimestre est à 0.

NB : les filtres 2 et 3 sont neutralisés pour les données actuellement présentées sur le site [www.imagepme.fr](http://www.imagepme.fr)

#### **Opération 4 : Détail de tous les niveaux géographiques**

L'information géographique présente initialement dans la base est le code postal. Cela permet d'obtenir l'information sur :

- Le code postal
- Le département
- La région administrative

Les lieux-dits ne sont pas pris en compte. Les entreprises dont le code postal n'est pas reconnu sont neutralisées.

#### **Opération 5 : Détail de tous les niveaux sectoriels**

L'information sectorielle présente initialement dans la base est le code APE (niveau 5 de la NAF). Il convient d'ajouter l'information sur tous les niveaux sectoriels (Niveaux 1 à 4) avec les codes et les intitulés des activités. L'intitulé du niveau 5 est ajouté car seul le code est présent dans la base initiale.

Les entreprises avec un code NAF 000Z sont neutralisées. La table de nomenclature officielle est disponible sur le site de l'Insee : [www.insee.fr](http://www.insee.fr)

#### **Opération 6 : Calcul du chiffre d'affaires annuel de l'entreprise « TRIG »**

Pour calculer le CA annuel d'une entreprise, nous cumulons le CA des 12 derniers mois. Si certains mois sont manquants, quelle qu'en soit la raison, alors nous appliquons une procédure d'extrapolation statistique des mois manquants telle que définie au point Opération 8 :

#### **Opération 7 : Stratification statistique de la population**

Voici les tranches de CA définies par secteur dans un objectif d'analyse de la situation économique par taille d'entreprise tous secteurs confondus :

- Tranche 1 : de 0 à 149 999€ de CA annuel glissant
- Tranche 2 : de 150 000 à 249 999€ de CA annuel glissant
- Tranche 3 : de 250 000 à 499 999€ de CA annuel glissant
- Tranche 4 : de 500 000 à 999 999€ de CA annuel glissant
- Tranche 5 : 1 million d'euros et plus de CA annuel glissant

Ces tranches sont appliquées pour tous les niveaux de la NAF (niveau 1 à 5).

L'affectation de l'entreprise est effectuée en utilisant le TRIG, c'est-à-dire en cumulant les CA des 12 derniers mois glissants (et en extrapolant les mois manquants si besoin).

## Opération 8 : Calcul des indices sur les données chiffre d'affaires

### L'indice de chiffre d'affaires (ICA)

L'ICA est calculé sur un périmètre d'entreprises présentes lors d'une période analysée et lors de la même période de l'année précédente, qui n'ont pas changé de code NAF et dont le siège social n'a pas changé de département entre les deux périodes.

Un chiffre d'affaires moyen par entreprise est calculé (CA total déclaré par les entreprises qui restent dans le périmètre après filtres, rapporté au nombre d'entreprises de cet échantillon).

Ce chiffre d'affaires moyen est corrigé des jours ouvrés. A savoir, le CA sera divisé par le nombre exact de jours ouvrés du mois concerné et multiplié par la moyenne des jours ouvrés pour le mois concerné observé 10 ans avant et 10 ans après. La méthode s'applique sur toutes les données. Cette correction des jours ouvrés permet de comparer un mois donné par rapport au même mois d'une autre année mais pas de le comparer avec un autre mois de l'année.

L'évolution du CA sera ensuite calculée entre le CA moyen du mois concerné et celui du même mois de l'année précédente pour donner l'ICA.

Pour l'ICA trimestriel, est calculé un ICA agréant les déclarations mensuelles et trimestrielles.

Si des déclarations correctives sont récupérées au cours du temps, elles ne viennent pas modifier les indices précédemment publiés.

$$\text{ICA} = \frac{\text{Moyenne des CA du mois en cours de l'année N}}{\text{Moyenne des CA du mois en cours de l'année N-1}} \times 100$$

$$\text{Moyenne des CA du mois en cours de l'année N} = \frac{\text{Somme des CA de chaque entreprise du mois en cours corrigé des jours ouvrés}}{\text{Nombre d'entreprises concernées}}$$

$$\text{CA de chaque entreprise du mois en cours corrigé des jours ouvrés} = \text{somme des CA du mois en cours corrigé des jours ouvrés}$$

$$\text{CA du mois en cours corrigé des jours ouvrés} = \frac{\text{CA mensuel}}{\text{Nb de jours ouvrés du mois}} \times \text{moyenne des jours ouvrés du mois sur 20 ans}$$

## L'indice de chiffre d'affaires cumulé (ICAC)

L'ICAC permet de mesurer l'état d'avancement par rapport au premier mois ou trimestre de l'année. Il s'agit de cumuler le CA depuis le 1er janvier de l'année en cours, de le moyenniser par rapport au nombre d'entreprises concernées, puis de le rapporter au cumul du CA des mêmes mois de l'année précédente (moyenné par le nombre d'entreprises concernées), avec correction des jours ouvrés.

L'ICAC vaut la valeur de la division de la moyenne des CA cumulés entre janvier de l'année N et le dernier mois disponible de l'année N, par la moyenne des CA cumulé des mêmes mois de l'année N-1, multiplié par 100 et ce, après application de la correction des jours ouvrés et des mois manquants (cf. opération 9). L'ICAC peut aussi être présenté en pourcentage dans les analyses afin de le différencier de l'ICA.

La moyenne des CA cumulés depuis janvier de l'année N est calculée comme la somme des CA de chaque entreprise depuis janvier, corrigée des jours ouvrés et corrigé des mois manquants à laquelle nous divisons le nombre d'entreprises concernées.

La correction par les jours ouvrés correspond pour chaque mois à la division du CA mensuel par le nombre de jour ouvré du mois puis multiplié par la moyenne des jours ouvrés sur 20 ans (10 ans avant/10 ans après).

Le calcul des CA depuis janvier corrigés des mois manquants correspond à la somme des CA depuis janvier corrigée des jours ouvrés que l'on divise par le nombre de mois renseignés puis multiplié par le nombre de mois depuis janvier.

Les mêmes opérations sont effectuées pour le calcul de l'ICAC trimestriel (il cumule également les résultats des déclarations mensuelles et trimestrielles des entreprises).

$$\text{ICAC} = \frac{\text{Moyenne des CA cumulé depuis janvier de l'année N}}{\text{Moyenne des CA cumulé depuis janvier de l'année N-1}} \times 100$$

$$\text{Moyenne des CA cumulé depuis janvier de l'année N} = \frac{\text{Somme des CA de chaque entreprise depuis janvier corrigé des jours ouvrés et corrigé des mois manquants}}{\text{Nombre d'entreprises concernées}}$$

$$\text{CA depuis janvier corrigé des mois manquants et des jours ouvrés} = \frac{\text{Somme des CA depuis janvier corrigé des jours ouvrés}}{\text{Nombre de mois renseignés}} \times \text{Nb de mois depuis janvier}$$

$$\text{CA corrigé des jours ouvrés sur 20 ans} = \frac{\text{CA mensuel}}{\text{Nb de jours ouvrés du mois}} \times \text{moyenne des jours ouvrés du mois sur 20 ans}$$

## Opération 9 : Gestion spécifique des données manquantes

Afin de prendre en compte les données manquantes sur un ou plusieurs mois entre les deux périodes comparées, une opération complémentaire est réalisée pour affecter un CA sur ces données manquantes. S'il manque jusqu'à 6 mois de données au cours des 12 derniers mois, nous affectons une valeur indicative à ces mois manquants pour le calcul du TRIG, du CA trimestriel des entreprises qui déclarent mensuellement ou pour le calcul de l'ICAC. Tous ces calculs sont effectués pour les entreprises qui déclarent au mensuel et au trimestriel.

La règle que nous appliquons pour simuler le montant du CA réalisé est la suivante :

- Nous créons une sous base SIREN « cas complet », c'est-à-dire comprenant tous les SIREN qui n'ont pas de données manquantes sur toute la période d'analyse considérée (soit trimestrielle, soit depuis janvier (ICAC), soit annuelle).
- Pour cette base « cas complet » nous allons commencer par créer la matrice qui va calculer pour chaque mois la part que représente ce mois par rapport à l'ensemble des autres mois de la période et ce pour toutes les entreprises du même secteur au niveau 5 de la NAF. Puis la part que représente chacun des couples de deux mois par rapport à l'ensemble des autres mois. Et ainsi de suite jusqu'à chacun des couples de six mois par rapport à l'ensemble des autres mois.
- Pour extrapoler les valeurs des mois manquants, nous allons alors nous baser sur le pro rata de ce que représente ces mois manquants par rapport à la période considérée pour toutes les entreprises du même secteur au niveau 5 de la NAF (sous-filtre 8b) en utilisant les rapports calculés dans la matrice précédente.

Exemple : il manque les données de novembre pour une pharmacie pour le calcul trimestriel du 4<sup>ème</sup> trimestre N. Nous regardons combien en % représente le mois de novembre par rapport aux cumuls d'octobre et décembre pour toutes les autres pharmacies présentes dans la base. Nous appliquons ce % au mois d'octobre et décembre de mon entreprise pour extrapoler son CA de novembre et donc avoir l'extrapolation de son CA du 4<sup>ème</sup> trimestre.

Ces règles sont valables pour chaque niveau 5 de la NAF si, dans la matrice des « cas complet », il y a au moins 10 entreprises « complètes » pour faire la moyenne. S'il y a moins de 10 entreprises alors une version différente de remplacement des données manquantes est effectuée. Cela ne représente que peu de SIREN, et souvent des SIREN qui sont ensuite exclus des analyses car ne répondant pas au secret statistique de 10 entreprises (cf. opération 11). La règle appliquée est alors la suivante :

- Pour une période considérée de N mois, nous cumulons le CA des N mois disponibles. Nous le divisons par le nombre de déclarations mensuelles réellement effectuées et nous remultiplions par N.

Cependant, pour réaliser le calcul des résultats trimestriels des entreprises qui déclarent mensuellement, une règle hybride a été réalisée afin de laisser le programme plus intelligible et manipulable tout en évitant des biais algorithmiques importants. Si pour le calcul de cette somme le dernier mois composant le trimestre a une valeur renseignée, alors le traitement des éventuelles données manquantes des deux premiers mois du trimestre correspond aux règles de remplacement des données manquantes telles que définies ci-dessus. Cependant, si le dernier mois composant le

trimestre est manquant alors que nous avons une information au mois d'avant ou l'antépénultième (ce sont des cas qui arrivent peu souvent en périmètre constant), alors nous appliquons une formule plus simple pour gérer les mois manquants :

- Pour une période considérée de N mois, nous cumulons le CA des N mois disponibles.
- Nous le divisons par le nombre de déclarations mensuelles réellement effectuées
- Nous remultiplions par N.

### **Filtre 2B : neutralisation (par suppression) des SIREN qui présentent au moins 6 déclarations manquantes depuis le début d'année**

Pour le calculer de l'ICAC à partir des déclarations mensuelles de TVA, une entreprise est neutralisée s'il manque 6 données entre le mois en cours et le mois de janvier :

- Cela permet de ne pas extrapoler un nombre trop important de données manquantes.
- Cela permet d'éviter d'avoir des données extrapolées sur trop peu de valeurs

### **Opération 10 : Déclinaison de l'indice de CA**

L'indice de CA peut être décliné par :

- Secteur (les 5 niveaux de la NAF),
- Territoire (National, régional, départemental, CP),
- Taille (tranche de CA annuel réalisé).

Il est publié :

- Mensuellement
- Trimestriellement

A noter que pour la déclinaison de l'ICA et ICAC par tranche, une entreprise qui change de tranche entre m-12 et m reste, pour le calcul des indicateurs, au sein de sa tranche m-12. Cela explique notamment que les résultats de la tranche 1 soient souvent très positifs car elles incluent les entreprises qui seraient passées entre les 2 périodes de la tranche 1 à une tranche supérieure et qui auraient donc connu de fortes croissances.

### **Opération 11 : Secret Statistique**

Les résultats des indices ne seront pas publiés ou seront masqués s'ils concernent moins de 10 entreprises.

### **Filtre 12 : suppression valeurs extrêmes**

Les valeurs extrêmes sont neutralisées sur toutes les données. A savoir, une entreprise est neutralisée au cours d'un mois donné si la valeur de son chiffre d'affaires dépasse 3 fois l'écart type des CA sur les 15 derniers mois de son secteur au niveau 1 de la NAF. L'entreprise est exclue pour 15 mois glissants avant et après le mois M si elle est neutralisée par ce filtre. Ce filtre peut être désactivé en cas de besoin.

### **Filtre 13 : neutralisation des SIREN avec un CA annuel supérieur à 50 millions d'euros**

Une entreprise est neutralisée si son CA annuel dépasse les 50 millions d'euros. L'entreprise est exclue pour 15 mois glissants avant et après le mois si elle est neutralisée par ce filtre.

### **Opération 14 : Impact neutralisation d'une entreprise par un filtre**

Une entreprise neutralisée par un filtre, l'est sur tous les niveaux sectoriels et tous les niveaux géographiques. Ainsi, le cumul des échantillons des départements doit bien faire le total du régional et du national ; même chose sur tous les niveaux sectoriels.

### **Opération 15 : suppression des entreprises dites « zombies »**

Dans notre échantillon, un nombre assez important d'entreprises n'a pas d'activité renseignée sur une période de 12 mois glissants (environ 30 000 SIREN en 2020). Ces entreprises ont télédéclaré des CERFA 3310CA3 durant l'année mais en ne déclarant jamais de CA ou alors toujours avec un montant à 0. Ces entreprises dites « zombies » sont neutralisées.

### **Opération 16 : Ordre d'application des filtres et opérations méthodologiques :**

- 1) **Opération 4** Détail de tous les niveaux géographiques (filtre appliqué dès la base N1)
- 2) **Filtre 1** Neutralisation (par suppression) des SIREN qui présentent plusieurs déclarations de chiffre d'affaires pour un même mois.
- 3) **Opération 9** Extrapolation des données manquantes
- 4) **Opération 6** Calcul du chiffre d'affaires annuel de l'entreprise « TRIG »
- 5) **Filtre 13** Suppression des SIREN dont le CA annuel « TRIG » est supérieur à 50 millions d'euros
- 6) **Filtre 2** Neutralisation (par suppression) des SIREN qui présentent des déclarations à 0 sur 6 mois ou plus consécutifs au cours des 12 derniers mois
- 7) **Filtre 3** Neutralisation (par suppression) des SIREN qui présentent pour le mois analysé ou le même mois en N-1 une déclaration à 0
- 8) **Opération 5** Détail de tous les niveaux sectoriels
- 9) **Opération 7** Stratification statistique de la population
- 10) **Filtre 12** Suppression des entreprises à valeurs extrêmes
- 11) **Opération 14** Suppression des entreprises dites « zombies »
- 12) **Opération 11** Secret Statistique
- 13) **Opération 14** Impact neutralisation d'une entreprise par un filtre



- 14) **Filtre 2b** : neutralisation (par suppression) des SIREN qui présentent au moins 6 déclarations manquantes depuis le début d'année
- 15) **Opération 8a** Calcul de l'ICA, ICAC
- 16) **Opération 8b** Déclinaison de l'ICA, ICAC par secteur, territoire, taille

### **Opération 17 : Information importante pour les mois de juillet et les mois de confinement.**

Les déclarations à néant, congés payés et reports du crédit de TVA sont pris en compte et passés comme des CA à 0 pour le calcul de l'ICAC pour les mois de juillet et les mois de confinement.

Mais attention, l'ICAC de juillet n'aura pas de sens économique et ne sera pas publié / analysé. Il permet d'avoir un ICAC qui redeviendra représentatif à compter du mois d'août. Le CA de juillet est en réalité réintégré avec la déclaration du mois d'août par les experts-comptables. Un CA à 0 est appliqué en juillet afin de ne pas extrapoler le CA de juillet qui serait vide alors que nous le réintégrons bien dans le calcul de l'ICAC à compter du mois d'août.

Pour l'ICA, cela est fait pour les mois de confinement uniquement, car certaines entreprises ont coché cette case quand elles sont restées fermées. Il convient de les prendre en compte comme des CA à 0 pour rendre compte de la réalité économique et de ne pas les exclure. A titre d'exemple, sur le mois d'avril 2020, cela a concerné près de 13 000 entreprises de coiffure sur les 15 000 présents dans la base.

### **Opération 18 : Sortie regroupée pour les mois de juillet et août**

Pour prendre en compte les particularités des mois de juillet et d'août qui impactent un grand nombre d'entreprises, une sortie cumulée des mois de juillet et août est effectuée.

Les filtres 2, 3, 11 et 12 sont recalculés sur le CA cumulé. Un mois « 7-8 » est créé pour retrouver les SIREN filtrés spécifiquement pour ce cumul.

Le cumul s'effectue en sommant les valeurs des CA de juillet et d'août.

### 3. Contrôle qualité des indicateurs

En complément des travaux de nettoyage, agrégation et filtres appliqués pour fiabiliser les données avant leur communication, une démarche de contrôle de la qualité des séries de données produites a été confiée à un prestataire externe afin d’appréhender leur fiabilité.

8 indicateurs ont été retenus pour mesure la fiabilité de l’ICA. Sont à prendre et à interpréter avec précaution et signalées par une couleur distincte (orange pâle au lieu du bleu foncé), les séries détectées par un ou plusieurs des 8 indicateurs : si a minima un indicateur donne une valeur de 2 ou deux des indicateurs donnent une valeur de 1. L’indicateur sur la taille des échantillons distingue de fait les séries avec un échantillon inférieur à 100 entreprises pour les résultats mensuels et de 50 entreprises pour les résultats trimestriels.

A noter que dans un contexte économique instable comme c’est le cas depuis le mois de mars 2020 avec la crise sanitaire, les résultats comparés entre une période et la même période de l’année précédente peuvent être très volatils et être « flaggés » par les indicateurs de qualité. Mais ils peuvent être pour autant le reflet d’une situation économique réelle. **La mise en lumière de certaines séries par les indicateurs est une invitation à la prudence dans l’interprétation des résultats et à les mettre en perspective avec le contexte.**

#### 3.1 Taille de l’échantillon

Il est important de signaler que l’analyse et les indicateurs de test ne sont totalement pertinents que pour les séries comportant plus de 100 observations pour les résultats mensuels et de 50 observations pour les résultats trimestriels. Les séries avec une taille d’échantillon faible ont un écart type de l’ICA plus important. L’écart type de l’ICA reste acceptable au-dessus d’une centaine d’entreprises dans la série. En dessous de ce seuil, la taille du panel peut à elle seule expliquer des variations importantes de l’ICA et de l’ICAC. Toutes les séries dont l’échantillon est inférieur à 100 entreprises pour les résultats mensuels et de 50 entreprises pour les résultats trimestriels sont donc « flaggées » de fait. Il convient d’être prudent en cas d’utilisation et d’interprétation.

#### 3.2 Stabilité aux évolutions de périmètre

L’ICA est calculé en effectuant un ratio entre le chiffre d’affaires moyen du mois M et du mois M-12<sup>1</sup> sur le même périmètre d’entreprises. Ainsi, pour le mois M-12, deux périmètres d’entreprises sont disponibles pour estimer le chiffre d’affaires moyen :

1. Le périmètre utilisé pour calculer l’ICA pour le mois M
2. Le périmètre utilisé pour calculer l’ICA pour le mois M+12<sup>2</sup>.

Cet indicateur de test consiste dans le calcul de la déviation entre les chiffres d’affaires indiqués par ces deux périmètres. Le tableau suivant présente les seuils retenus pour cet indicateur de test.

Signal	Valeur numérique	Déviations due au changement de périmètre
Pas de signal	0	Comprise entre -15% et 15%
Signal faible	1	Comprise entre -25% et -15% ou entre 15% et 25%
Signal fort	2	Inférieur à -25% ou supérieur à 25%

<sup>1</sup> Le même mois mais de l’année d’avant (ex : janvier 2022 et janvier 2021)

<sup>2</sup> Une même approche est retenue pour l’analyse trimestrielle

### 3.3 Impact des filtres 2 et 3

Les filtres 2 et 3 peuvent potentiellement avoir un impact significatif sur les résultats. Si l'impact de ces filtres est trop élevé, les séries sont considérées comme insuffisamment fiables. L'indicateur test analyse l'impact des filtres 2 et 3 sur l'ICA.

- Filtre 2 : Neutralisation (par suppression) des SIREN qui présentent des déclarations à 0 sur 6 mois ou plus + consécutifs au cours des 12 derniers mois
- Filtre 3 : neutralisation (par suppression) des SIREN qui présentent pour le mois analysé ou le même mois en M-12 une déclaration à 0

Le tableau suivant présente les seuils retenus pour cet indicateur.

Signal	Valeur numérique	Impact des filtres 2 et 3 sur l'ICA
Pas de signal	0	Comprise entre -10 points (pts) et 10 pts
Signal faible	1	Comprise entre -15 pts et -10 pts ou entre 15 pts et 25 pts
Signal fort	2	Inférieur à -15 pts ou supérieur à 15 pts

### 3.4 Homogénéité de l'échantillon

L'ICA est calculé pour le chiffre d'affaires moyen mais également pour le premier et troisième quartile du chiffre d'affaires, ainsi que pour la médiane. Une forte disparité entre ces ICA témoigne d'un comportement différent entre les entreprises de la série. Ainsi, les séries dont l'écart entre ces différents calculs d'ICA est significatif peuvent être insuffisamment fiables ou donner des conclusions erronées ou portant sur un sous-ensemble de l'échantillon concerné.

Le tableau suivant présente les seuils retenus pour cet indicateur.

Signal	Valeur numérique	Différence (absolue) entre les ICA
Pas de signal	0	Inférieure à 25 pts
Signal faible	1	Comprise entre 25 pts et 35 pts
Signal fort	2	Supérieure à 35 pts

### 3.5 Évolution du nombre d'observations

Une baisse importante du nombre d'observations peut être le signe d'un manque de fiabilité de la série (cette baisse peut être liée à un biais de sélection ou de survie qui nuit à la pertinence de la série). Le nombre d'observations d'une série est comparé au nombre d'observations moyen sur les 12 mois passés. Le tableau suivant présente les seuils retenus pour cet indicateur.

Signal	Valeur numérique	Baisse du nombre d'observations
Pas de signal	0	Hausse ou baisse inférieure à 10%
Signal faible	1	Baisse comprise entre 10% et 25%
Signal fort	2	Baisse supérieure à 25%

### 3.6 Disparité des ICA

Pour les séries départementales et régionales, les géographies dont l'ICA est trop éloigné de la distribution des ICA pour le secteur d'activité étudié (c'est à dire, avec l'ensemble des séries géographiques sur le même secteur) peuvent être considérées comme sujet à caution. Ainsi, cet indicateur test signale les séries trop singulières par rapport à la distribution des autres séries du même secteur d'activité.

Le tableau suivant présente les seuils retenus pour cet indicateur.

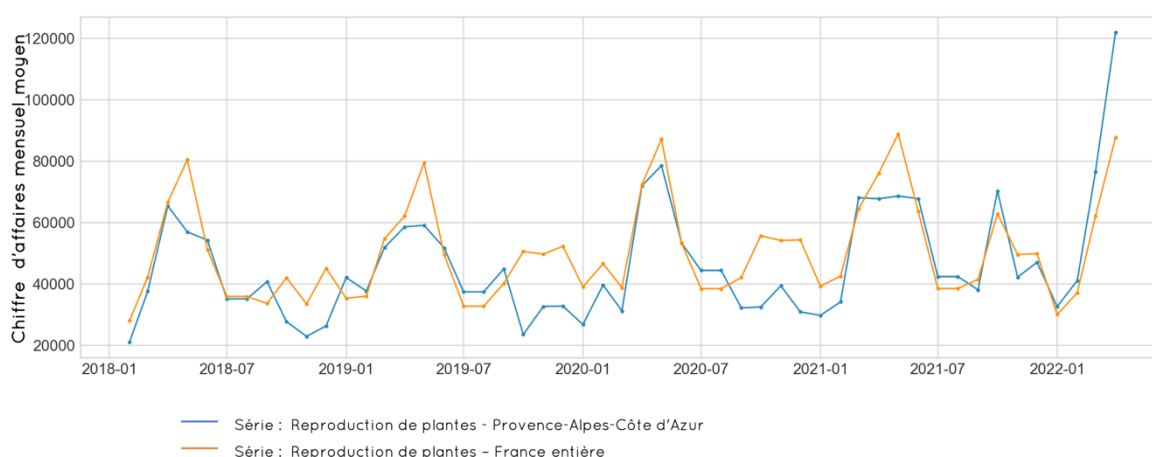
Signal	Valeur numérique	Écart par rapport au secteur d'activité
Pas de signal	0	Déviations inférieures à 2 écarts types
Signal faible	1	Déviations comprises entre 2 et 3 écarts types
Signal fort	2	Déviations supérieures à 3 écarts types

Exemple : pour le secteur d'activité Coiffure, la moyenne et l'écart type des ICA pour l'ensemble des séries géographiques (séries départementales et régionales confondues) sont estimées. Supposons que cette moyenne égale à 105 et cet écart type égal à 10. Une hypothétique série du secteur de la Coiffure en Meuse d'un ICA de 130 aurait un signal faible sur cet indicateur (valeur numérique égale à 1).

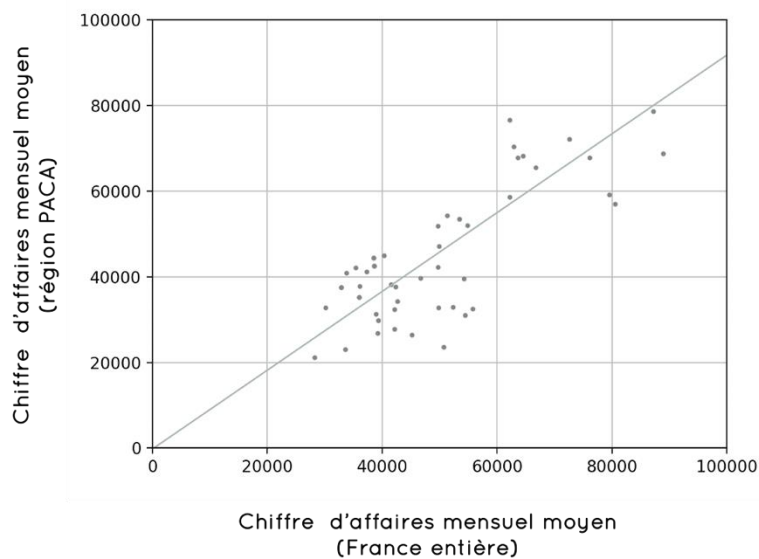
### 3.7 Analyse de la corrélation avec la série nationale

Pour les séries locales (départementales ou régionales), une évolution du chiffre d'affaires au niveau local non expliquée par l'évolution au niveau national est un signal d'une possible anomalie. L'indicateur transcrit les évolutions spécifiques au mois M et au mois M-12. Cet indicateur est construit en deux étapes.

La première étape consiste à déterminer la relation entre la série nationale et la série locale. Dans l'exemple ci-dessous, les chiffres d'affaires moyen des activités de reproduction de plantes en Provence-Alpes-Côte d'Azur ("PACA") et dans la France entière sont liés, comme le démontre le graphique.

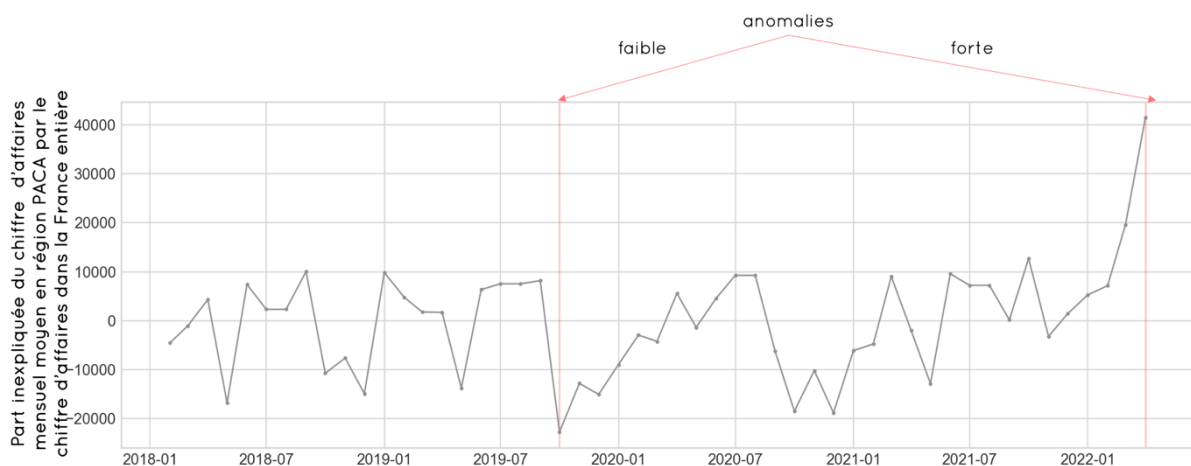


La relation entre les deux séries peut être analysée par une régression linéaire, comme illustré dans le graphique ci-dessous.



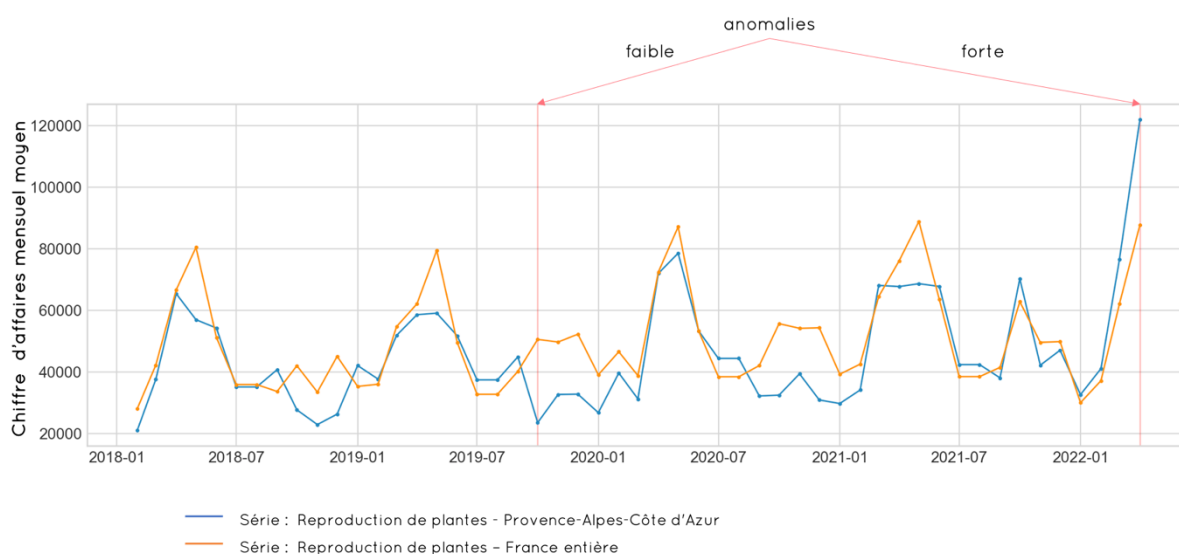
Ce graphique démontre bien l'existence d'une relation entre le CA mensuel moyen pour la France entière et le CA mensuel moyen pour la région PACA pour les activités de reproduction de plantes. Sur la base de ce modèle de régression linéaire, il est possible d'expliquer une grande partie mais pas l'intégralité des évolutions de la série locale (dans le graphique ci-dessus, les points sont proches de la droite mais ne sont pas exactement dessus).

La deuxième étape consiste à analyser les variations locales ne pouvant être expliquées par les variations nationales et à identifier les potentielles anomalies. Le graphique ci-dessous présente la part inexpliquée des variations locales en fonction du temps.



Sur l'ensemble de ce graphique, deux points, en octobre 2019 et avril 2022 sortent de la plage de fluctuation « normale » (définie ci-après). Ainsi, l'activité liée à la reproduction de plantes dans la France entière explique l'évolution de ce même secteur dans la région PACA sauf en octobre 2019 et en avril 2022. L'indicateur donne donc un signal respectivement faible et fort pour ces deux mois.

Le graphique ci-après présente ces deux points sur les courbes locales et nationales de CA moyen, permettant de comprendre le mécanisme de détection d'anomalie : pour ces deux points, les écarts entre les courbes nationales et locales sont particulièrement importants.



Comme pour les autres indicateurs test, cet indicateur renvoie trois signaux possibles. Un mois est considéré comme singulier si la part inexplicée du chiffre d'affaires s'écarte de plus de 3 fois de l'intervalle interquartile pour un signal fort et de 1,5 fois pour un signal faible. Le tableau suivant présente les seuils retenus pour cet indicateur.

Signal	Valeur numérique	Chiffre d'affaires non expliqué par la série nationale
Pas de signal	0	Déviations inférieures à 1,5 intervalle interquartile
Signal faible	1	Déviations comprises entre 1,5 et 3 intervalles interquartiles
Signal fort	2	Déviations supérieures à 3 intervalles interquartiles

### 3.8 Analyse de la corrélation avec des secteurs proches

Les secteurs d'activité ont été regroupés en fonction de leur corrélation passée<sup>3</sup>. Par exemple, sur la NAF de niveau 2, les secteurs d'activité suivant sont considérés comme proches :

- Construction de maisons individuelles
- Ingénierie, études techniques
- Services d'aménagement paysager
- Travaux d'installation d'eau et de gaz en tous locaux
- Travaux d'installation d'équipements thermiques et de climatisation
- Travaux d'installation électrique dans tous locaux
- Travaux de maçonnerie générale et gros œuvre de bâtiment
- Travaux de menuiserie bois et PVC
- Travaux de menuiserie métallique et serrurerie
- Travaux de peinture et vitrerie

Pour les nomenclatures de niveaux 3 à 5, plusieurs tables de regroupements ont été réalisées afin de prendre en compte la disponibilité des séries.

<sup>3</sup> Un regroupement selon la corrélation passée a été privilégié à un regroupement sur la base de la nomenclature NAF. En effet, les corrélations passées font apparaître des liens entre les secteurs d'activités qui sont pourtant éloignés dans la nomenclature NAF (par exemple dans l'analyse de la NAF 732 P.1 les secteurs *Commerce de voitures et de véhicules automobiles légers* et *d'Enseignement de la conduite* sont dans le même groupe de secteurs d'activités mais sont pourtant éloignés dans la nomenclature.

Les corrélations sont très marquées par les restrictions et impacts économiques liés à la crise du Covid. Par exemple, certains secteurs d'activité ont eu le même calendrier de fermeture et sont alors corrélés. Les tables de corrélation traduisent notamment ces fermetures et leurs impacts économiques. Ces tables seront mises à jour une fois l'intégralité des effets économiques de la pandémie passés en excluant éventuellement cette période.

Ces groupes de secteurs permettent de créer, pour chacun, une série synthétique qui doit expliquer la série analysée. L'indicateur transcrit les évolutions spécifiques au mois M et au mois M-12<sup>4</sup> n'étant pas expliqué par la série synthétique.

Cet indicateur est construit en trois étapes :

1. Une série synthétique est créée sur la base des corrélations passées observées
2. La relation entre la série revue et la série synthétique est analysée,
3. Les variations ne pouvant pas être expliquée par les variations de la série synthétique sont revues. Si une variation pour le mois M ou le moins M-12<sup>5</sup> est anormale, l'indicateur test retourne un signal de niveau faible ou fort.

Les étapes 2 et 3 ont un fonctionnement proche de celui de l'indicateur décrit précédemment.

L'exemple de la série liée au commerce de détail de livres en magasin spécialisé (NAF 732) en France permet de comprendre le fonctionnement de cet indicateur. L'analyse des corrélations passées nous permet d'identifier un groupe composé des secteurs suivants :

- Activités des marchands de biens immobiliers
- Boulangerie et boulangerie-pâtisserie
- Charcuterie
- Commerce de détail d'autres équipements du foyer
- Commerce de détail de biens d'occasion en magasin
- Commerce de détail de matériels audio et vidéo en magasin spécialisé
- Commerce de détail de matériels de télécommunication en magasin spécialisé
- Fabrication d'articles de joaillerie et bijouterie
- Transformation du thé et du café

La première étape consiste à créer un indicateur synthétique sur la base des indicateurs de ces secteurs.

Le graphique ci-dessous présente ces deux séries. Il apparaît que la série synthétique<sup>6</sup> suit relativement bien la série du Commerce de détail de livres en magasin spécialisé (NAF 732) (France entière).

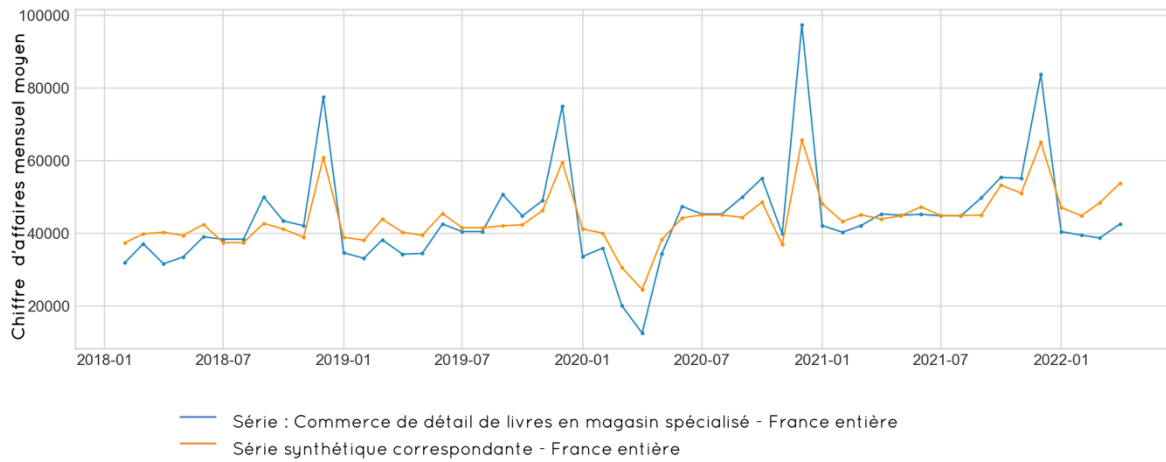
---

<sup>4</sup> Ou le trimestre T et le trimestre T-4

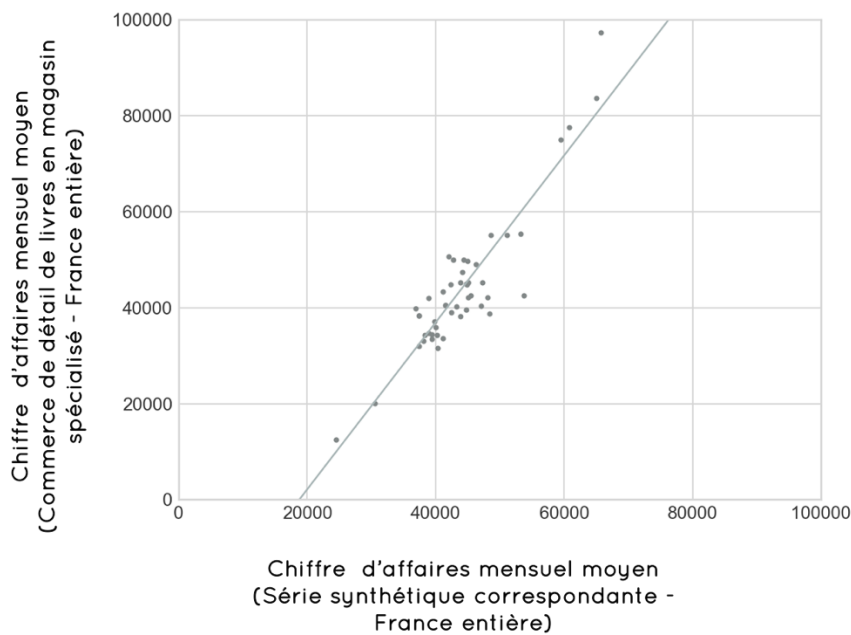
<sup>5</sup> Ou le trimestre T et le trimestre T-4

<sup>6</sup> La série synthétique est composée des séries suivantes :

- Activités des marchands de biens immobiliers
- Boulangerie et boulangerie-pâtisserie
- Charcuterie
- Commerce de détail d'autres équipements du foyer
- Commerce de détail de biens d'occasion en magasin
- Commerce de détail de livres en magasin spécialisé
- Commerce de détail de matériels audio et vidéo en magasin spécialisé
- Commerce de détail de matériels de télécommunication en magasin spécialisé
- Fabrication d'articles de joaillerie et bijouterie
- Transformation du thé et du café



Dans un deuxième temps, la relation entre la série revue et la série synthétique est déterminée sur la base d'une régression linéaire. Cette analyse est présentée dans le graphique suivant pour l'exemple des librairies.



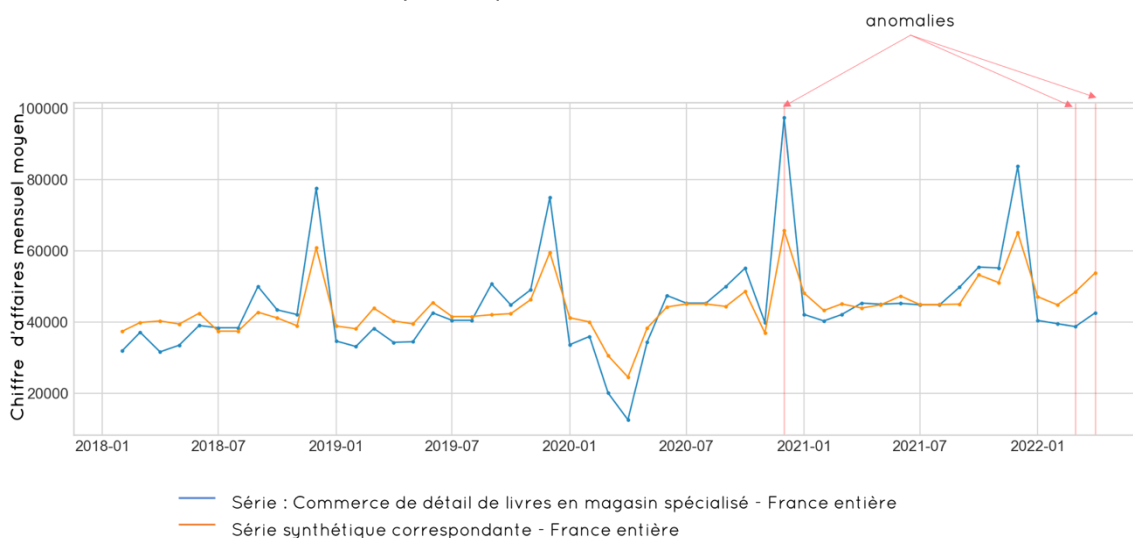
Après une analyse de corrélation, la part résiduelle (c'est-à-dire la partie de la série analysée non expliquée par la série synthétique) est estimée. Cette part résiduelle est revue. Cette part résiduelle est présentée dans le graphique ci-dessous pour l'exemple des librairies.





S'il y a une déviation importante de cette part résiduelle sur un mois, alors l'indicateur test renvoie un signal. L'ICA de cette série des librairies sera identifié comme potentiellement anormal pour cet indicateur pour les mois de décembre 2020, mars et avril 2022 et les trois mois respectifs de l'année suivante (décembre 2021, mars et avril 2023).

Ces anomalies potentielles sont également visibles sur le graphique présentant l'évolution de la série du commerce livre et de la série synthétique.



Tout comme l'indicateur précédent, cet indicateur test renvoie également trois signaux possibles sur la base de l'intervalle interquartile observé sur la partie résiduelle.

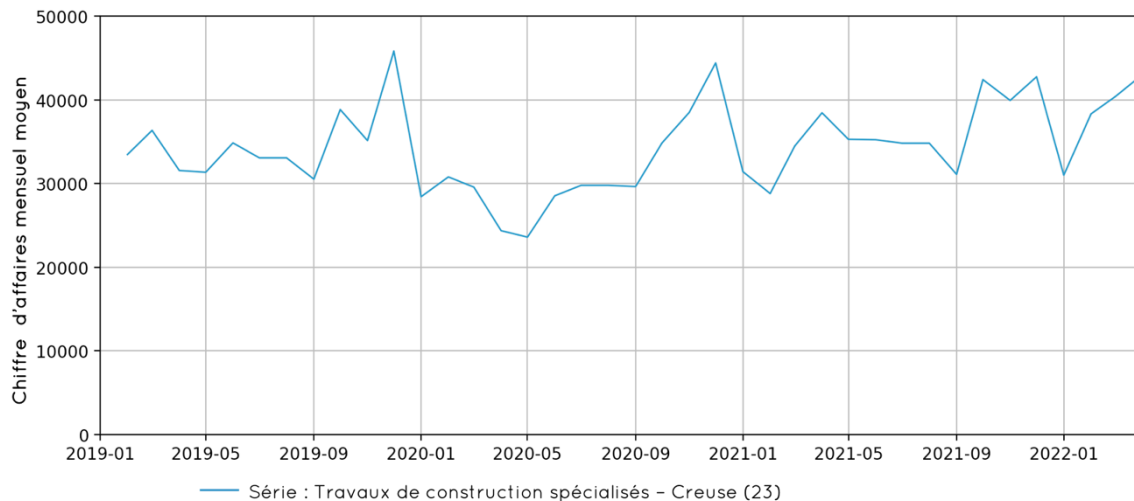
Signal	Valeur numérique	Chiffre d'affaires non expliqué par la série synthétique
Pas de signal	0	Déviations inférieures à 1,5 intervalle interquartile
Signal faible	1	Déviations comprises entre 1,5 et 3 intervalles interquartiles
Signal fort	2	Déviations supérieures à 3 intervalles interquartiles

### 3.9 Analyse de la saisonnalité et d'une variation spécifique

Cet indicateur est construit en deux étapes :

1. Le chiffre d'affaires est désaisonnalisé (i.e. les variations saisonnières sont supprimées),
2. Les variations n'étant pas liées aux saisons sont analysées. Si une variation pour le mois M ou le moins M-12<sup>7</sup> est anormale, l'indicateur test retourne un signal de niveau faible ou fort.

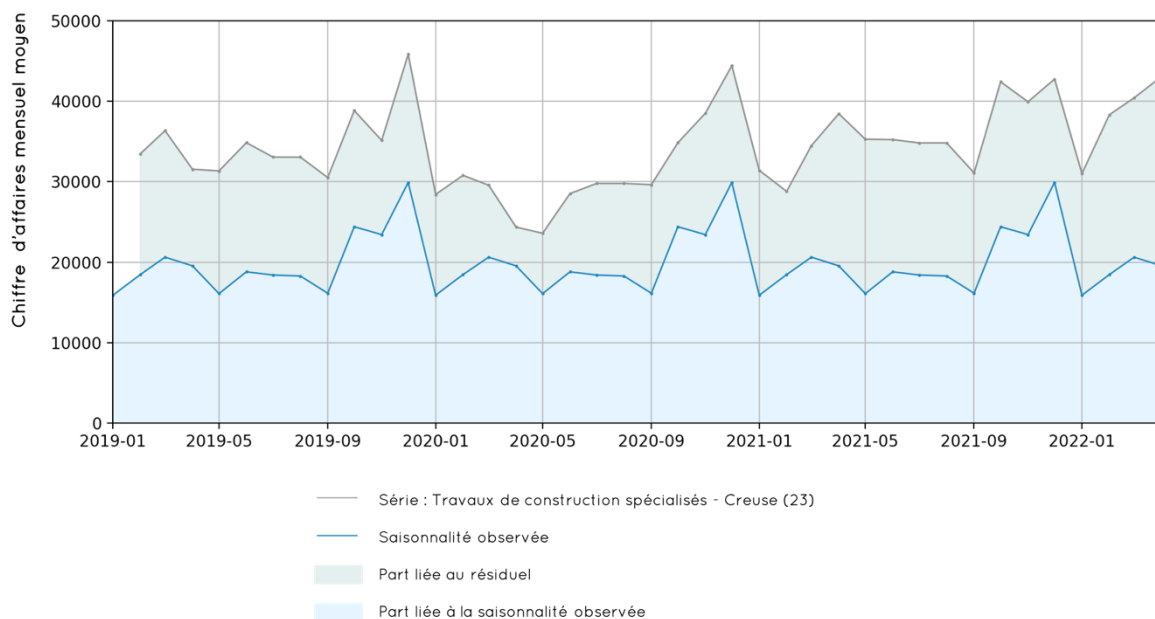
Le graphique suivant présente l'évolution du chiffre d'affaires des travaux de construction spécialisés dans la Creuse. Des pics en décembre sont observés, sans être singuliers car répétitifs.



Cette série est divisée en deux composantes :

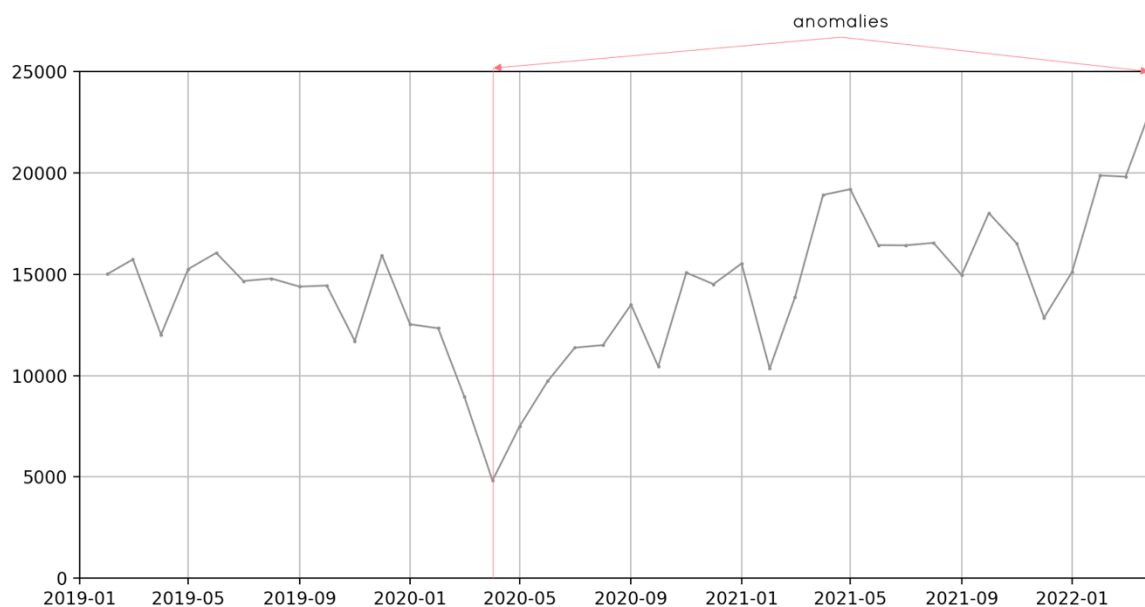
- Une composante saisonnière et
- Une composante non expliquée (appelée le résiduel) qui correspond à la série désaisonnalisée.

**Cette décomposition est illustrée dans le graphique ci-dessous.**



<sup>7</sup> Ou le trimestre T et le trimestre T-4

Le graphique suivant présente le résiduel (c'est-à-dire la série désaisonnalisée).



— Part liée au résiduel (i.e. série désaisonnalisée) : Travaux de construction spécialisés - Creuse

Les variations de cette série désaisonnalisée sont analysées. Comme pour les deux indicateurs précédents, cet indicateur renvoie trois signaux possibles sur la base de la déviation par rapport à des valeurs cibles liées à l'intervalle interquartile

Signal	Valeur numérique	Déviations du chiffre d'affaires non expliquée par la saisonnalité
Pas de signal	0	Déviations inférieures à 1,5 intervalles interquartiles
Signal faible	1	Déviations comprises entre 1,5 et 3 intervalles interquartiles
Signal fort	2	Déviations supérieures à 3 intervalles interquartiles

Pour cet exemple, la série désaisonnalisée ne laisse plus apparaître de pic pour les mois de décembre. Les pics observés pour les mois de décembre ne sont pas des anomalies. À l'inverse, il apparaît que les mois d'avril 2020 et d'avril 2022 présentent des anomalies : le chiffre d'affaires pour le mois d'avril 2020 (resp. 2022) est fortement inférieur (resp. supérieur) à ceux des autres mois d'avril bien que le mois d'avril ne soit pas impacté par une saisonnalité particulière. Ainsi, pour le calcul de l'ICA les mois d'avril 2020, 2021, 2022 et 2023 auront un signal fort pour cet indicateur.