

# Le Rapport National sur l'Emploi en France d'ADP

## Méthodologie détaillée :

En étroite collaboration avec Moody's Analytics, Inc. et avec son équipe expérimentée d'experts du marché du travail, *ADP Research Institute*® a mis au point le Rapport National sur l'Emploi en France®. Ce rapport a pour but d'estimer les chiffres de l'emploi en France, en complément de ceux publiés par l'Institut National de la Statistique et des Études Économiques (INSEE). Tandis que les données sur l'emploi salarié non-agricole sont publiées chaque trimestre par l'INSEE, les projections d'ADP concernant les évolutions de l'emploi en France seront générées sur une base mensuelle. Les données d'ADP tiennent compte des 16 secteurs d'activité figurant dans le rapport trimestriel de l'INSEE.

## Analyse des données :

Les données d'ADP sont générées mensuellement et le panel se compose d'environ 75 000 établissements de tailles différentes et issues de différents secteurs d'activité. Les noms et les codes des secteurs d'activité figurent dans le tableau ci-dessous.

*Tableau 1 : Descriptions des secteurs d'activité :*

<b>Code</b>	<b>Description</b>
DE	Energie, eau et déchets
C1	Produits alimentaires
C2	Coke et raffinage pétrolier
C3	Machines et biens d'équipement
C4	Matériel de transport
C5	Autres biens industriels
FZ	Construction
GZ	Commerce
HZ	Transport
IZ	Services d'hébergement et de restauration
JZ	Information et communication
KZ	Services financiers
LZ	Services immobiliers
MN	Services aux entreprises
OQ	Services non-marchand
RU	Services aux ménages

Pour chaque établissement, le nombre de salariés actifs est notifié chaque mois à un jour précis qui correspond au jour de l'élaboration des bulletins de paie. La période s'étend de mai 2009 à juillet 2015. Les données étant non-compensées, on utilise l'interpolation pour gérer les observations manquantes. Une fois ces dernières traitées, on retire les valeurs aberrantes afin de lisser les changements de pourcentages mensuels. Les critères servant à supprimer les valeurs aberrantes sont définis selon la taille des entreprises afin de refléter au mieux la réalité et de minimiser les éventuelles erreurs.

Dans la dernière étape de la modélisation, nous nous penchons sur les plus grandes fluctuations d'échantillons se produisant certains mois. Il est nécessaire de procéder à un lissage plus approfondi. En effet, même lorsque le taux de croissance au niveau de l'entreprise est cohérent après la suppression des valeurs aberrantes, nous pouvons observer de grandes fluctuations dans le nombre total de salariés actifs certains mois et dans certains secteurs d'activité. Ces fluctuations peuvent prendre la forme de forts pics ou de fortes chutes dans les chiffres de l'emploi totaux du secteur pendant un ou plusieurs mois, ou d'un changement en faveur d'un niveau durablement plus élevé. Ces fluctuations sont dues à des modifications plus importantes qu'à l'accoutumée dans le nombre d'établissements appartenant à un secteur donné. Afin de lisser les chutes importantes dans les chiffres de l'emploi au niveau d'un secteur d'activité, nous appliquons une procédure de moyenne mobile.

Après être passé par les étapes de traitement des données mentionnées ci-dessus, nous cumulons le nombre de salariés actifs par paire (mois M/mois M-1) pour chaque secteur et chaque mois. Nous obtenons ainsi des séries de données mensuelles pouvant être utilisées pour des estimations. Le changement dans le pourcentage mensuel est calculé pour chaque mois et pour chaque secteur d'activité. Des ajustements saisonniers sont ensuite effectués dans E-views à l'aide du module d'ajustement saisonnier X13. Le tableau n°2 présente les statistiques synthétiques de chaque secteur permettant d'obtenir les changements de pourcentages mensuels ajustés selon la saison. C'est pour cela que nous limitons notre échantillon à la période 2012-2015.

### **Comparaison entre les estimations en matière d'emploi d'ADP et les estimations de l'INSEE**

Les rapports de l'INSEE présentent chaque trimestre les niveaux d'emploi, les variations absolues d'un trimestre à l'autre et les changements de pourcentage d'un trimestre à l'autre pour les secteurs d'activités figurant dans le tableau n°1. Les niveaux d'emploi trimestriels correspondent au nombre de personnes employées le dernier jour du trimestre, quelle que soit la durée de leur travail en France métropolitaine. Les données trimestrielles se basent sur un sondage trimestriel réalisé par le Ministère du Travail (sondage ACEMO), sur les statistiques trimestrielles de l'URSSAF (organisme chargé de collecter les contributions sociales des employeurs), d'EPURE ainsi que sur les statistiques trimestrielles de Pôle Emploi pour les données sur le travail temporaire. L'estimation rapide de la situation trimestrielle de l'emploi, qui se fonde uniquement sur la première source et qui ne prend pas en compte les changements survenus dans les entreprises de moins de 10 salariés et dans les nouvelles entreprises, est publiée 45 jours après la fin de chaque trimestre et concerne uniquement les secteurs de la production de marchandises, de la

construction et des services. La première estimation utilisant les données d'EPURE et fournissant une répartition détaillée par secteur d'activité est publiée 70 jours après la fin d'un trimestre. Les données sont ensuite révisées 160 jours après la fin du trimestre.

Selon les données d'ADP et de l'INSEE, le commerce et les services aux entreprises semblent représenter la majorité des emplois. Comparativement aux données de l'INSEE, les données ADP font état d'une part plus grande attribuée aux secteurs de l'industrie, de l'information et de la communication, du commerce et des services financiers.

### Régression et résultats :

Nous utilisons la méthode vectorielle autorégressive pour modéliser et prévoir les changements mensuels dans la situation de l'emploi. Les modèles vectoriels autorégressifs (VAR) sont largement utilisés pour les analyses de séries chronologiques multivariées. Les modèles VAR typiques traitent les variables comme étant endogènes suite à la critique de Sim en 1980 sur l'hypothèse du caractère exogène des modèles macro-économiques. Les modèles VAR peuvent inclure des restrictions, parmi lesquelles figure le caractère exogène de certaines variables. Ils peuvent également être modifiés afin d'inclure des termes déterministes et des variables exogènes. Les modèles VAR sont utilisés pour modéliser et prévoir le comportement dynamique de séries chronologiques économiques et financières. Ils peuvent également servir dans le cadre de l'inférence structurelle et de l'analyse politique. Le modèle VAR basique (p) est défini par l'équation suivante :

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$$

Dans cette équation,  $A_1$  correspond à des matrices de coefficients et  $u_t$  correspond à un processus de bruit blanc dimensionnel K dans lequel  $E(u_t) = 0$  and  $E(u_t u_t^T) = \Sigma u$ . Le modèle VAR (p) est un processus stationnaire avec moyenne invariable dans le temps et structure à variance et covariance. En plus des décalages de variables endogènes, le modèle VAR peut également intégrer des variables exogènes explicatives à droite. Dans ce cas, le modèle VAR est représenté par l'équation suivante :

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + B_1 X_{t-1} + \dots + B_p X_{t-p} + u_t$$

dans laquelle  $X_t$  représente le vecteur de variables exogènes.

Les coefficients du processus VAR (p) sont estimés en soumettant chacune des équations à OLS. L'estimateur OLS est identique à l'estimateur GLS s'il n'y a pas de restriction dans les paramètres. Dans le cas d'un processus gaussien distribué normalement  $y_t \sim N(0, \Sigma u)$ , l'estimateur OLS est identique à l'estimateur ML. Les procédures habituelles d'inférence statistique peuvent être appliquées si le processus est stable. S'il y a des variables intégrées selon lesquelles  $y_t \sim I(1)$ , le processus est alors instable et les variables peuvent être co-intégrées. La méthode OLS/ML peut néanmoins être utilisée pour l'estimation des paramètres du modèle mais les statistiques-t et les statistiques-F habituelles peuvent engendrer des conclusions trompeuses lorsqu'elles sont utilisées pour la vérification des hypothèses. Le processus VAR peut également être représenté sous forme structurelle si des restrictions sont imposées sur les

paramètres du modèle, c'est-à-dire des éléments de la matrice A, de la matrice B ou des deux matrices dans l'équation ci-dessus.

Dans notre cas, l'utilisation du modèle VAR nous permet d'intégrer l'effet indirect et l'effet de rétroaction sur l'emploi au sein des différents secteurs d'activité et de mesurer le mouvement commun de l'emploi dans les différents secteurs grâce à des indicateurs conjoncturels tournés vers le passé et vers le futur. Nous commençons par l'estimation d'un VAR sans restriction avec un décalage dans lequel les variables endogènes sont représentées par les changements mensuels dans 11 secteurs d'activité. Nous choisissons d'intégrer les données sur l'emploi des secteurs C1, C2, C3, C4 et C5 au sein d'un seul secteur que nous appelons industrie. Les autres secteurs inclus dans notre modèle correspondent aux secteurs de l'INSEE listés dans le tableau n°1. Pour les variables exogènes explicatives, on utilise des indices de confiance des entreprises pour l'industrie et les services publiés chaque mois par l'INSEE ainsi que les changements de pourcentages dans les niveaux d'emploi pour chaque secteur publiés trimestriellement par l'INSEE. Afin d'utiliser dans notre modèle les données sur l'emploi de l'INSEE qui sont exprimées à une fréquence trimestrielle, nous convertissons ces données dans une fréquence mensuelle à l'aide de l'outil de conversion de moyenne à correspondance quadratique de E-views. Le modèle VAR initial sans restriction utilisé dans notre estimation peut être exprimé de la façon suivante :

$$\begin{aligned}
 PC_{tm,t} &= \text{const} + a_1 PC_{tm,t-1} + a_1 PC_{tm,t-1} + a_2 PC_{de,t-1} + a_3 PC_{fz,t-1} + a_4 PC_{gz,t-1} + a_5 PC_{hz,t-1} + a_6 PC_{iz,t-1} \dots \\
 &+ a_{10} PC_{mn,t-1} + a_{11} PC_{oq,t-1} + a_{12} PC_{ru,t-1} + a_{13} BC_{mf,t} + a_{14} BC_{serv,t} + a_{15} PC_{tminsee,t-2} \\
 &\quad + a_{16} PC_{deinsee,t-2} + a_{16} PC_{fzinsee,t-2} + a_{17} PC_{gzinsee,t-2} + a_{18} PC_{hzinsee,t-2} u_t \\
 &\quad + \dots + a_{18} PC_{izinsee,t-2} + \dots a_{22} PC_{mminsee,t-2} + a_{23} PC_{oqinsee,t-2} + a_{24} PC_{ruinsee,t-2} \\
 &\quad + u_t
 \end{aligned}$$

$$\begin{aligned}
 PC_{de,t} &= \text{const} + \\
 a_1 PC_{tm,t-1} &+ a_1 PC_{tm,t-1} + a_2 PC_{de,t-1} + a_3 PC_{fz,t-1} + a_4 PC_{gz,t-1} + a_5 PC_{hz,t-1} + a_6 PC_{iz,t-1} \dots \\
 &+ a_{10} PC_{mn,t-1} + a_{11} PC_{oq,t-1} + a_{12} PC_{ru,t-1} + a_{13} BC_{mf,t} + a_{14} BC_{serv,t} + a_{15} PC_{tminsee,t-2} \\
 &\quad + a_{16} PC_{deinsee,t-2} + a_{16} PC_{fzinsee,t-2} + a_{17} PC_{gzinsee,t-2} + a_{18} PC_{hzinsee,t-2} u_t \\
 &\quad + \dots + a_{18} PC_{izinsee,t-2} + \dots a_{22} PC_{mminsee,t-2} + a_{23} PC_{oqinsee,t-2} + a_{24} PC_{ruinsee,t-2} \\
 &\quad + u_t
 \end{aligned}$$

$$\begin{aligned}
 PC_{fz,t} &= \text{const} + a_1 PC_{tm,t-1} + a_1 PC_{tm,t-1} + a_2 PC_{de,t-1} + a_3 PC_{fz,t-1} + a_4 PC_{gz,t-1} + a_5 PC_{hz,t-1} + a_6 PC_{iz,t-1} \dots \\
 &+ a_{10} PC_{mn,t-1} + a_{11} PC_{oq,t-1} + a_{12} PC_{ru,t-1} + a_{13} BC_{mf,t} + a_{14} BC_{serv,t} + a_{15} PC_{tminsee,t-2} \\
 &\quad + a_{16} PC_{deinsee,t-2} + a_{16} PC_{fzinsee,t-2} + a_{17} PC_{gzinsee,t-2} + a_{18} PC_{hzinsee,t-2} u_t \\
 &\quad + \dots + a_{18} PC_{izinsee,t-2} + \dots a_{22} PC_{mminsee,t-2} + a_{23} PC_{oqinsee,t-2} + a_{24} PC_{ruinsee,t-2} \\
 &\quad + u_t
 \end{aligned}$$

.....

$$\begin{aligned}
& PC_{ru,t} \\
& = \text{const} + a_1 PC_{tm,t-1} + a_1 PC_{tm,t-1} + a_2 PC_{de,t-1} + a_3 PC_{fz,t-1} + a_4 PC_{gz,t-1} + a_5 PC_{hz,t-1} + a_6 PC_{iz,t-1} \dots \\
& + a_{10} PC_{mn,t-1} + a_{11} PC_{oq,t-1} + a_{12} PC_{ru,t-1} + a_{13} BC_{mf,t} + a_{14} BC_{serv,t} + a_{15} PC_{tminsee,t-2} \\
& \quad + a_{16} PC_{deinsee,t-2} + a_{16} PC_{fzinsee,t-2} + a_{17} PC_{gzinsee,t-2} + a_{18} PC_{hzinsee,t-2} u_t \\
& \quad + \dots + a_{18} PC_{izinsee,t-2} + \dots a_{22} PC_{mninsee,t-2} + a_{23} PC_{oqinsee,t-2} + a_{24} PC_{oqinsee,t-2} \\
& \quad + u_t
\end{aligned}$$

Dans le modèle ci-dessus, PC signifie changement de pourcentage. Les séries pour lesquelles il est mentionné INSEE dans le sous-index sont des changements de pourcentage dans les données INSEE sur l'emploi pour chaque secteur d'activité. Enfin, BC signifie confiance des entreprises, MF signifie industrie et SERV est utilisé pour désigner les services.

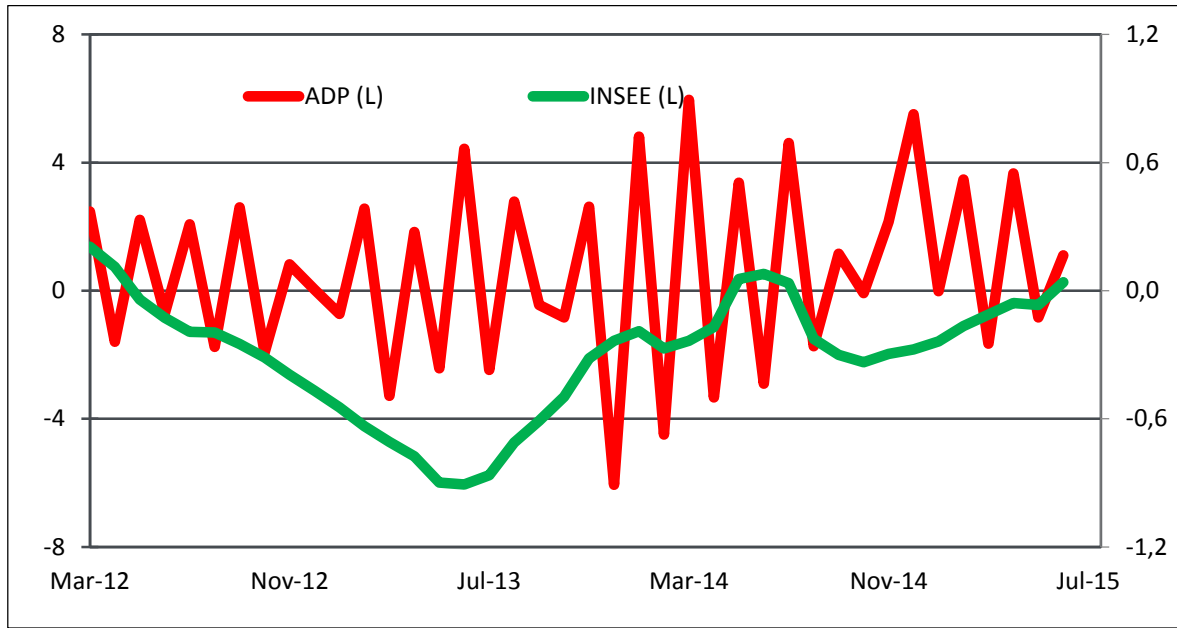
Après l'estimation initiale, nous limitons certaines estimations de coefficient à zéro. Plusieurs raisons nous obligent à imposer des restrictions. Tout d'abord, ces dernières nous forcent à atténuer les effets des indices de confiance des entreprises et des données sur l'emploi de l'INSEE pour un même secteur sur les changements dans l'emploi estimés dans l'hypothèse où ces effets vont dans des directions opposées. Par ailleurs, ces restrictions sont imposées afin de mesurer l'effet des facteurs les plus importants : il s'agit des variations retardées des pourcentages dans les données INSEE de l'emploi pour le secteur correspondant et l'indice de confiance des entreprises sur l'avenir. Ces facteurs sont statistiquement plus importants et plus prononcés si les estimations de leur coefficient sont négligeables dans les spécifications moins parcimonieuses et optimales. Enfin, les variables statistiquement négligeables sont abandonnées pour améliorer l'adéquation générale du modèle.

Une fois l'estimation du modèle réalisée, on réalise des tests résiduels et de stabilité. Une version F des statistiques Wald est utilisée pour tester la signification conjointe des coefficients du modèle car il est préférable d'avoir un échantillon de petite taille. Nous pouvons constater que les indices de confiance des entreprises ont une grande importance statistique et constituent des signes positifs. Les résultats des tests confirment que l'augmentation de l'emploi selon l'INSEE pour un secteur donné pendant deux périodes a un effet grandement positif et statistiquement important sur les pourcentages d'emploi ADP du même secteur. Les régressions montrent également que les effets indirects et les effets d'interaction parmi les secteurs d'activité sont relativement considérables. Ces effets peuvent être considérés comme des estimations de coefficient statistiquement importantes pour la majorité des changements concernant l'emploi au niveau des secteurs d'activité compris dans les données INSEE. Enfin, les statistiques du chi carré et F indiquent que les variables dans la régression ont une importance statistique conjointe.

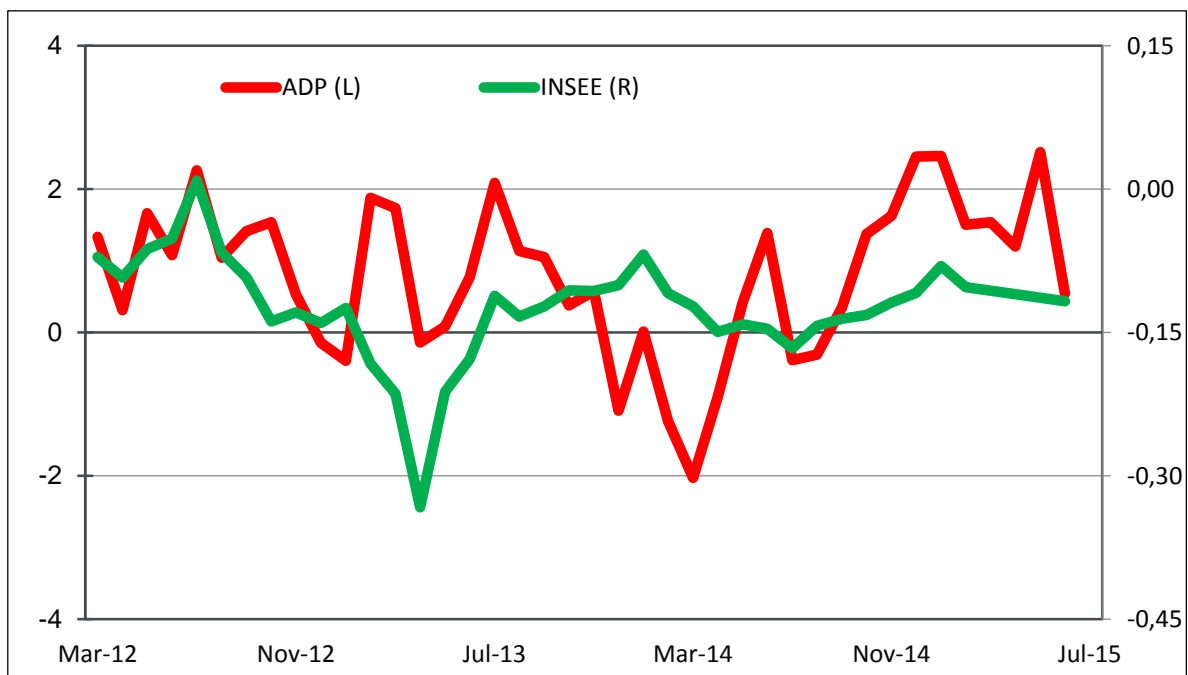
### Résultats :

Dans les graphiques ci-dessous, nous plaçons les valeurs ajustées des changements mensuels de pourcentages dans les données ADP ainsi que les variations de pourcentages des données INSEE.

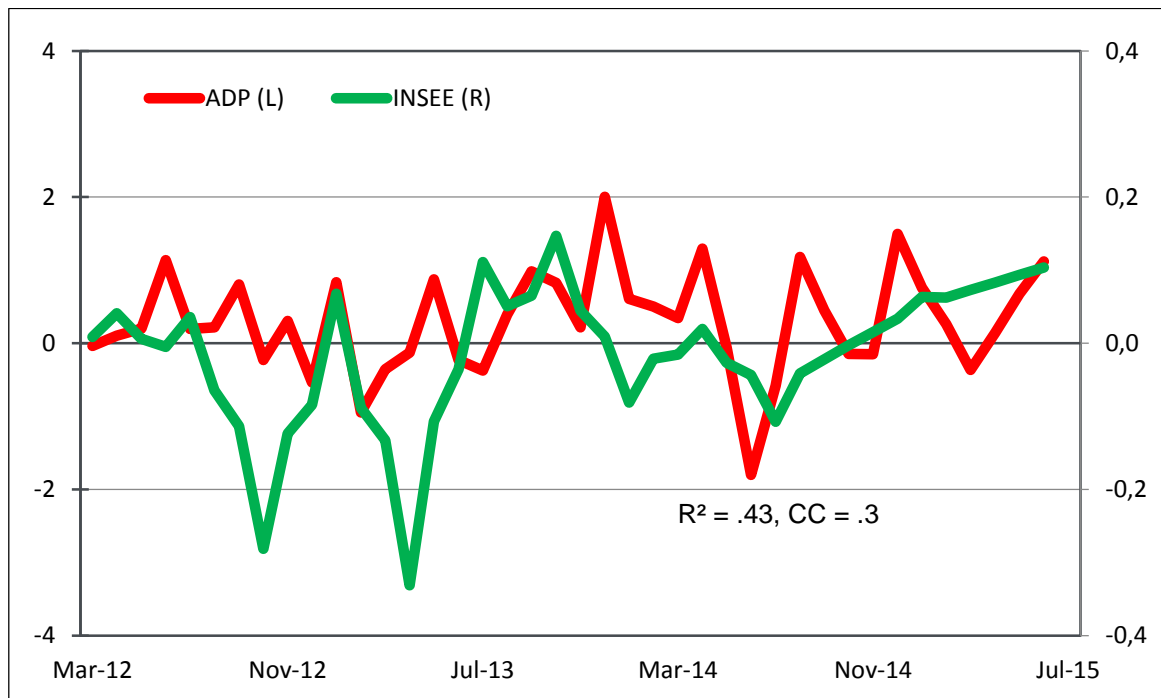
**Graphique n°1. Emploi total, % de changement mensuel, SA**



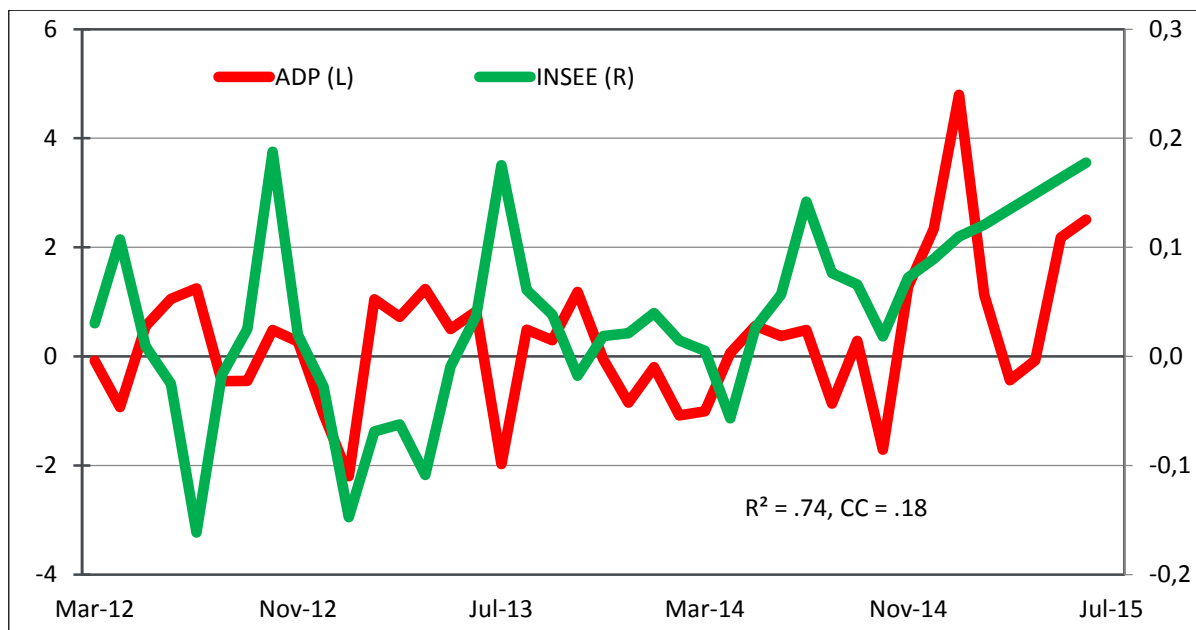
**Graphique n°2. Total industrie, % de changement mensuel, SA**



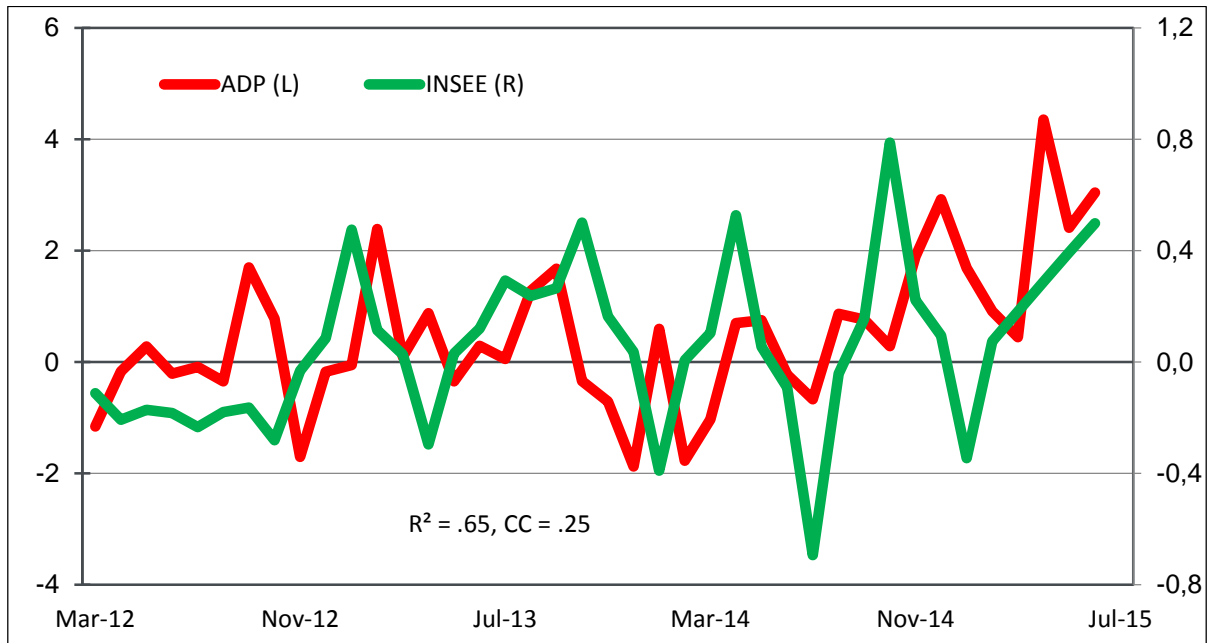
Graphique n°3. Commerce, % de changement mensuel, SA



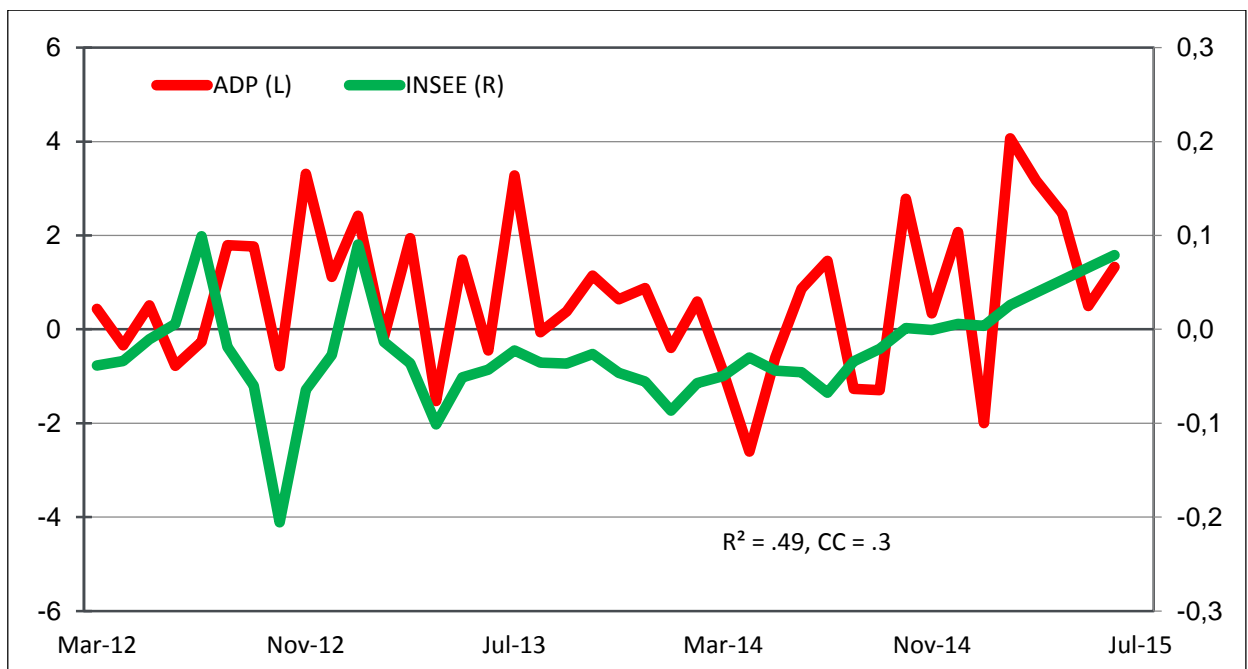
Graphique n°4. Services financiers, % de changement mensuel, SA



Graphique n°5. Services aux entreprises, % de changement mensuel, SA



Graphique n°6. Transport, % de changement mensuel, SA





Graphique n°7. Immobilier, % de changement mensuel, SA

