



A more human resource.®

**LE RAPPORT NATIONAL SUR
L'EMPLOI EN FRANCE D'ADP
METHODOLOGIE**



A more human resource.®

Introduction

Ce document décrit le modèle du Rapport National sur l'Emploi en France d'ADP qui est publié tous les mois sur l'emploi dans le secteur privé hors agriculture. Alors qu'un rapport officiel sur l'emploi en France est publié par l'INSEE tous les trimestres, les estimations de l'évolution de l'emploi en France sont générées et publiées mensuellement par ADP sur la base des données transactionnelles de paie d'ADP en France et avant la publication des chiffres du chômage mensuels.

Le modèle a été redéveloppé pour plusieurs raisons. La principale est due aux changements dans la méthodologie de l'INSEE sur l'estimation de l'emploi salarié, aussi bien dans les alertes que les résultats trimestriels, mais pour prendre en compte la plus large mesure permise par la couverture croissante d'ADP pour tous les principaux secteurs et enfin pour intégrer dans le modèle de meilleurs outils de benchmarking et de nettoyage des données.

Dans ce modèle, les données d'ADP englobent tous les 16 principaux secteurs inclus dans les rapports trimestriels de l'INSEE. C'est dû notamment par les changements méthodologiques introduits par l'INSEE en 2017. Les données utilisées pour la modélisation sont un panel de données contenant en moyenne 70 000 établissements de différentes tailles et de différents secteurs. Pour chaque société, le nombre de salariés actifs est enregistré à un jour spécifique du mois qui est défini comme étant la date d'édition de la dernière fiche de paie. La période couverte débute en 2009 néanmoins l'échantillon pour le modèle commence en 2013 pour une meilleure inférence statistique.

Mise en place du modèle

Dans le modèle actuel et la mise en place de la procédure, plusieurs choses ont changé. Premièrement, le processus de nettoyage des données a été modifié pour que l'algorithme de paires appariées soit toujours respecté. Par ailleurs, après une inspection approfondie des données de paie d'ADP, les valeurs critiques et limites utilisées dans les changements mensuels du nombre d'emplois dans les données brutes ont été ajustées pour identifier les aberrations. De plus, considérant la part croissante du nombre d'entreprises du secteur agricole ayant des salariés, le secteur a été inclus parmi les secteurs modélisés, essentiellement à cause de potentiels effets indirects sur les autres secteurs. De même les techniques d'interpolation pour les chiffres trimestriels de l'INSEE ont été améliorées et les biais de fin du mois réduits. Comme indiqué précédemment, l'ensemble des données disponibles des chiffres officiels de l'INSEE a été révisé dans le nouveau modèle de reporting standard.

En suivant une technique révisée de nettoyage des données, l'échantillon a été réduit jusqu'à 2013 (jusqu'à la dernière observation disponible). De nombreux nouveaux facteurs de valeurs ont été testés pour les différents secteurs et ajoutés au modèle. Après ajustement et redéfinition des contraintes variables des différents VARX, le modèle a été divisé en deux parties. La première partie à la forme d'un vecteur autorégressif standard pour des variables sélectionnées, comme dans la mise en



A more human resource.®

place du modèle original. La seconde partie du modèle introduit des équations OLS individuelles incluant des termes autorégressifs pour certains secteurs. Ces secteurs incluent le commerce, les services aux entreprises, l'agriculture, les services publics, les services domestiques et l'industrie. Ces deux modèles doivent être vus comme des modèles concurrents, avec des interférences statistiques et économiques mesurées à chaque mise à jour pour déterminer quel modèle produit les résultats et prévisions les plus satisfaisants.

Description des données et Analyse

Dans cette section nous analysons la technique de nettoyage des données. Comme mentionné précédemment, les données d'ADP sont générées mensuellement et le panel se compose d'environ 70 000 établissements de tailles différentes et issues de différents secteurs d'activité. Les noms et codes des secteurs d'activités figurent dans le tableau ci-dessous.

Codes et noms des secteurs

Code	Description	Catégories agrégées
AZ	Agriculture, sylviculture et pêche	Agriculture, sylviculture et pêche
DE	Energie, eau et déchets	Energie, eau et déchets
C1	Alimentation et boissons	Industrie
C2	Cokéfaction et raffinage	
C3	Fabrication de machines et d'équipements	
C4	Fabrication d'équipements de transport	
C5	Autres biens manufacturés	
FZ	Construction	Construction
GZ	Commerce	Commerce
HR	Transports	Transports
IZ	Hébergement et restauration	Hébergement et restauration
JZ	Information et communication	Information et Communication
KZ	Services financiers	Services financiers
LZ	Services immobiliers	Services immobiliers
MN	Services aux entreprises	Services aux entreprises
OQ	Services publics	Services publics
RU	Services domestiques	Services domestiques



A more human resource.®

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 utilisée pour le modèle commence en 2013. Les données étant non-compensées, i.e. ne contiennent pas le même nombre d'observations à chaque date et/ou toutes les caractéristiques ne sont pas observables dans toutes les données à toutes les périodes, on utilise l'interpolation pour gérer les observations manquantes. Cette approche est utilisée pour maintenir les paires dans l'algorithme en place et de compenser les écarts mensuels ou bimestriels dans les données ou compenser les variations mensuelles de taille de l'échantillon.

Après avoir pris en compte les éléments manquants, les aberrations sont supprimées pour lisser les évolutions mensuelles. Après inspection des variations des données mensuelles, le critère pour supprimer les aberrations est personnalisé pour les différentes tailles d'établissement pour se rapprocher de la réalité et minimiser les erreurs de mesure.

Modèle VARX

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 Sims 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 $E(u_t u_t^T) = \Sigma$. 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.



A more human resource.®

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.

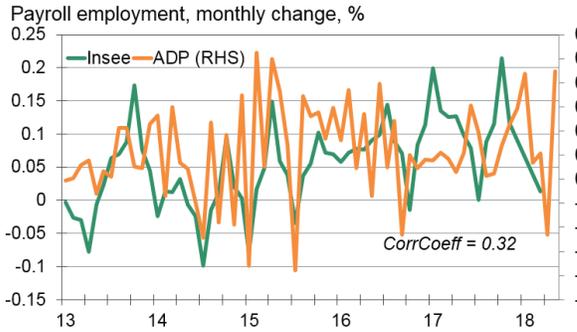
Resultats

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. Il est visible que les résultats dans la plupart des cas sont satisfaisants et représentent une amélioration par rapport au précédent modèle. Les représentations dans les tableaux ci-dessous se réfèrent aux comportements de l'échantillon.

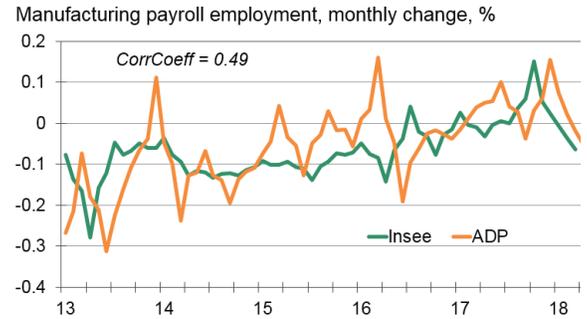


A more human resource.®

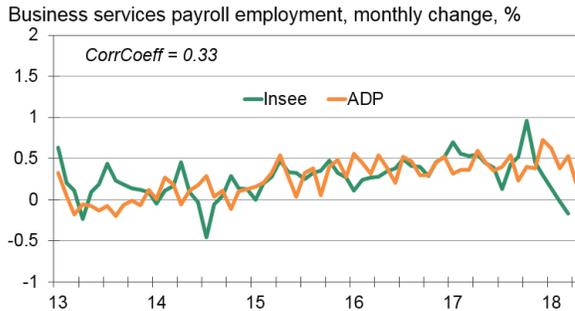
Performances au sein de l'échantillon des chiffres d'ADP comparés aux chiffres officiels de l'INSEE



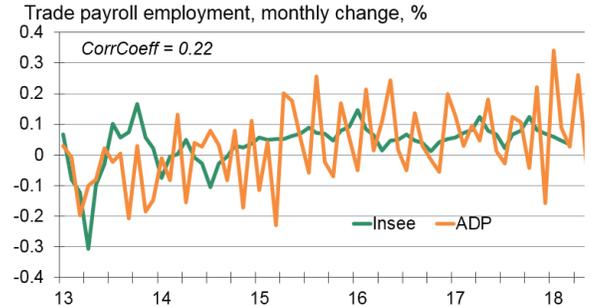
Sources: ADP, Insee, Moody's Analytics



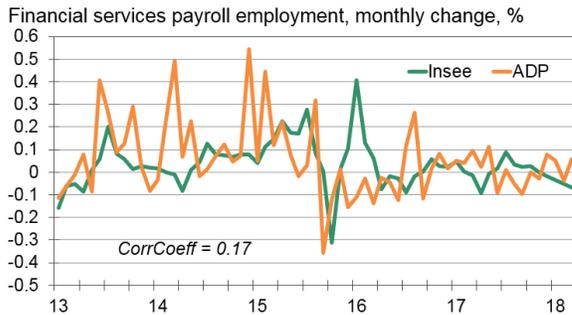
Sources: ADP, Insee, Moody's Analytics



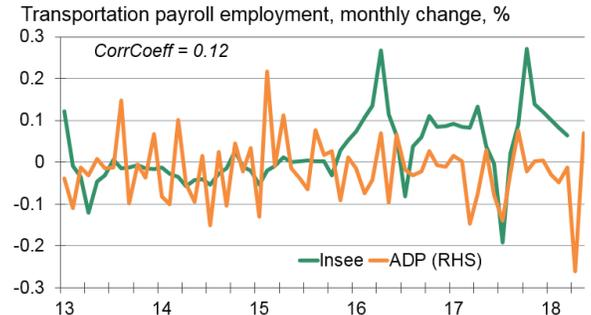
Sources: ADP, Insee, Moody's Analytics



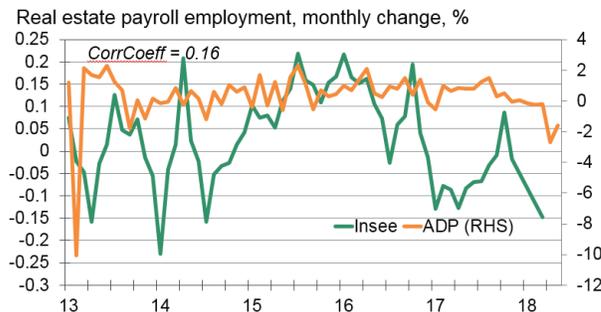
Sources: ADP, Insee, Moody's Analytics



Sources: ADP, Insee, Moody's Analytics



Sources: ADP, Insee, Moody's Analytics



Sources: ADP, Insee, Moody's Analytics