



Prévision de court terme de la croissance du PIB français à l'aide de modèles à facteurs dynamiques

Marie BESSEC
Catherine DOZ

PRÉVISION DE COURT TERME DE LA CROISSANCE DU PIB FRANÇAIS À L'AIDE DE MODÈLES À FACTEURS DYNAMIQUES

Marie BESSEC*
Catherine DOZ*

Ce document de travail n'engage que ses auteurs. L'objet de sa diffusion est de stimuler le débat et d'appeler commentaires et critiques.

* **Marie BESSEC** était en poste à la Direction Générale du Trésor au Ministère de l'Économie, des Finances et de l'Industrie (France).

* **Catherine DOZ** est en poste à Paris School of Economics, Université Paris1 Panthéon-Sorbonne, (France)

marie.bessec@banque-france.fr
catherine.doz@univ-paris1.fr

TABLE DES MATIÈRES

Résumé / Abstract	3
Introduction	4
1 Les méthodes de prévision utilisées dans les organismes français et européens	4
1.1 Les organismes français	5
1.2 Les organismes internationaux	6
2 Les modèles à facteurs et leur utilisation en prévision	9
2.1 Les modèles à facteurs	9
2.2 Estimation d'un modèle à facteurs dynamiques	12
2.3 Utilisation en prévision	14
2.4 Le choix de la spécification du modèle	15
3 Utilisation des modèles à facteurs dynamiques pour la prévision du PIB français	16
3.1 Les données	16
3.2 Analyse en échantillon et à information complète	17
3.3 Analyse hors échantillon et en pseudo temps réel	19
Conclusion	24
Bibliographie	28
Annexe : Modèle espace-état et filtre de Kalman	34

RÉSUMÉ

Les modèles à facteurs sont de plus en plus utilisés pour la prévision de court terme du PIB par les banques centrales et les grands organismes internationaux. Ces modèles permettent en effet de prendre en compte l'information apportée par un grand nombre de variables, en résumant cette information sous la forme d'un petit nombre de variables latentes appelées facteurs. Dans leur forme dynamique, ces modèles permettent à la fois de prendre en compte la dépendance temporelle des facteurs et de spécifier les variables observées en fonction des valeurs contemporaines et retardées des facteurs. Les techniques d'estimation associées peuvent être adaptées en cas de valeurs manquantes des indicateurs, de sorte qu'il n'est pas nécessaire de recourir à des modèles auxiliaires pour traiter les problèmes posés par les différences dans les délais de publication des variables utilisées.

Dans cette étude, nous examinons les performances de ces modèles pour la prévision du taux de croissance du PIB français sur des horizons courts. Nous utilisons une base constituée d'une centaine de variables parmi lesquelles des variables d'enquêtes, des indicateurs réels, des variables monétaires et financières et des indicateurs sur l'environnement international. L'évaluation hors échantillon et en pseudo temps réel des modèles étudiés montre généralement la supériorité des prévisions issues des modèles à facteurs sur celles issues des processus *benchmark* habituels. Les prévisions restent néanmoins fragiles avant le début du trimestre, lorsque l'on dispose de peu d'information. Dans ce cadre, on montre également que l'utilisation de variables financières et internationales permet d'améliorer les prévisions au début des campagnes de prévisions.

ABSTRACT

In recent years, factor models have received increasing interest from central banks and international organizations to forecast macroeconomic variables. These models make it possible to exploit the information provided by a large number of variables. In their dynamic form, factor models take the time dependence of the factors into account and allow for a relationship between the forecast variable and the contemporary and lagged factors. Moreover, the estimation procedure can be adapted to handle the missing values at the end of the sample due to publication lags.

We examine the performance of these models in forecasting the French GDP growth rate over short horizons. The factors are extracted from a large data set including surveys balances, real, financial and international variables. A pseudo real time evaluation over the last decade exhibits a gain relative to the usual benchmarks. However, forecasts remain inaccurate before the beginning of the quarter. We also show that the use of international and financial variables can improve forecasts at the longest horizons.

Introduction

Les modèles à facteurs connaissent depuis plusieurs années un développement important dans la littérature (voir Stock et Watson (2010) ou Bai et Ng (2008b) pour une revue sur ces modèles) et sont de plus en plus utilisés pour la prévision de court terme du PIB par les banques centrales et d'autres organismes internationaux. Ces modèles sont utilisés sous leur forme statique (par exemple à la Fed sous l'impulsion des travaux de Stock et Watson, 1999, 2002a et 2002b) ou sous leur forme dynamique (à la BCE suite aux travaux de Doz, Giannone et Reichlin, (2006, 2007) et de Giannone, Reichlin et Small (2008), à la Banque d'Italie avec l'indicateur Eurocoin développé par Altissimo et al. (2001, 2007), etc). Ils sont encore relativement peu utilisés dans les organismes français.

Les modèles à facteurs présentent plusieurs avantages par rapport aux outils classiques. Ils permettent tout d'abord de prendre en compte l'information apportée par un grand nombre de variables, celle-ci étant résumée sous la forme d'un petit nombre de facteurs qui feront ensuite office de variables explicatives dans un modèle de régression classique. Par ailleurs, il est possible d'adapter ces modèles en cas d'observations manquantes en fin de période. Il s'agit là d'une propriété importante pour le conjoncturiste, puisqu'il est confronté au problème posé par les délais de publication des indicateurs conjoncturels (assez courts pour les soldes d'enquêtes ou les variables financières, plus longs pour les variables réelles comme l'indice de production industrielle ou la consommation manufacturière) : lorsque nous utilisons un modèle à facteurs, il n'est pas nécessaire de développer des modèles auxiliaires pour prévoir les observations manquantes ou d'utiliser des modèles différents suivant le mois du trimestre où nous effectuons la prévision, *i.e.* suivant l'ensemble d'information qui est à la disposition du conjoncturiste.

Nous proposons ici une application de ces techniques à la prévision du taux de croissance du PIB français au trimestre précédent¹, courant et suivant. Nous utilisons une base de données constituée d'une centaine de variables parmi lesquelles des variables d'enquête, des indicateurs réels, des variables monétaires et financières et des indicateurs sur l'environnement international. Une évaluation hors échantillon montre que la qualité de prévisions fournies par les modèles à facteurs est satisfaisante, même si les prévisions restent fragiles lorsque l'horizon de prévision est éloigné. Des pistes d'amélioration suggérées en conclusion de l'étude pourront être explorées pour gagner en précision.

L'étude est organisée comme suit. La première section présente une revue des méthodes utilisées pour la prévision de court terme du PIB dans un certain nombre d'institutions françaises et internationales et décrit en particulier l'utilisation qui est faite des modèles à facteurs dans certaines d'entre elles. La deuxième section présente les modèles à facteurs dans leur forme statique puis dynamique, ainsi que les méthodes d'estimation et de prévision associées à ces modèles. La troisième section présente les données utilisées dans l'étude et examine les performances prévisionnelles des modèles à facteurs en échantillon et hors échantillon.

1. Les méthodes de prévision utilisées dans les organismes français et européens

Nous proposons ici un survol des méthodes employées pour la prévision à court terme du PIB dans les principales institutions françaises et dans un certain nombre d'organismes internationaux (une synthèse en est faite dans le tableau 3 en annexe). Nous constaterons en particulier que plusieurs organismes ont développé des outils fondés sur l'utilisation des modèles à facteurs, qui sont utilisés parallèlement à d'autres approches et que plusieurs méthodes d'estimation de ces modèles sont utilisées.

¹ Rappelons que, pendant la première moitié d'un trimestre donné, le chiffre du PIB du trimestre précédent n'est pas encore publié et qu'il fait donc encore l'objet d'une « prévision ».

1.1 Les organismes français

Les services conjoncturels de l'Insee, la Direction Générale du Trésor et la Banque de France, utilisent deux approches complémentaires pour prévoir le taux de croissance du PIB français sur de courts horizons : l'utilisation de maquettes macro-sectorielles et des outils d'étalonnages du PIB construits à partir de données d'enquêtes.

Les maquettes macro-sectorielles sont utilisées en deux temps. Au Trésor, les prévisions de certaines composantes de l'offre (production de l'industrie manufacturière, de l'industrie agro-alimentaire, production d'énergie) et de la demande (consommation des ménages, consommation des Administrations publiques (APU), investissement dans la construction) sont élaborées. Ces prévisions sont fondées sur des étalonnages simples (ou modèles *bridge*) avec comme variables explicatives des soldes d'enquêtes (principalement de l'Insee), substitués à des indicateurs quantitatifs réels au fur et à mesure de leur disponibilité dans le trimestre. Pour la plupart, ces étalonnages ont été construits de façon à reproduire la méthodologie des comptes trimestriels ou en utilisant l'algorithme GETS². Dans un deuxième temps, les prévisions macro-sectorielles d'offre et de demande sont mises en cohérence dans un cadre comptable à plusieurs secteurs d'activité³.

À la Banque de France, le modèle OPTIM⁴ décompose l'activité du côté de l'offre pour la prévision du trimestre en cours et suivant⁵ : il s'agit donc de prévoir séparément la production des principaux secteurs d'activité (biens agro-alimentaires, biens manufacturés, énergie, construction et services marchands), puis d'en déduire une prévision du taux de croissance du PIB en pondérant les prévisions sectorielles par des poids estimés par régression linéaire sur une période assez courte. Des étalonnages simples sont utilisés pour prévoir les composantes. Ces étalonnages sont les mêmes quelle que soit la position dans le trimestre et les valeurs manquantes sont prévues par des modèles autorégressifs et moyenne mobile (ARMA) ; deux équations alternatives sont néanmoins considérées, avec ou sans l'Indice de la production industrielle (l'IPI) du sous secteur concerné (sauf pour la prévision de la production de services où nous utilisons l'IPI de la branche manufacturière). Les variables explicatives utilisées sont des soldes d'enquêtes de la Banque de France, de l'Insee et de la Commission européenne et des indicateurs quantitatifs réels. Les équations ont été construites avec l'algorithme GETS utilisé en deux temps : l'algorithme est d'abord appliqué par source (par exemple, aux soldes de l'enquête dans les services, puis aux soldes de l'enquête dans l'industrie, etc), puis à l'ensemble des variables ainsi présélectionnées⁶.

En parallèle à l'utilisation de ces maquettes, des étalonnages du PIB construits à partir de données d'enquêtes permettent de contrôler le diagnostic global. À l'Insee, Erkel-Rousse et Minodier (2009) ont développé des étalonnages du taux de croissance trimestriel du PIB par mois et par horizon de prévision suivant Dubois et Michaux (2006) afin de prendre en compte l'évolution de l'ensemble d'information au cours du trimestre ; les variables explicatives sont sélectionnées parmi les soldes d'opinion des enquêtes de l'Insee dans l'industrie et dans les services avec l'algorithme GETS⁷. La Direction Générale du Trésor (DG Trésor) utilise également des étalonnages du taux de croissance du PIB, avec deux

² GEneral To Specific. La version originelle a été proposée par Hoover et Perez (1999) et étendue par Hendry et Krolzig (1999, 2001).

³ Les résultats obtenus par l'Insee sont publiés une fois par trimestre (en mars, juin, octobre et décembre) dans la note (ou le point) de conjoncture de l'Insee. Bercy ne communique pas sur ses prévisions infra-annuelles.

⁴ Développée par Irac et Sédillot en 2002, la version actuelle du modèle est présentée par Barhoumi et al. (2008).

⁵ Une décomposition est également effectuée côté demande (prévision de la consommation des ménages, de la consommation des APU, de l'investissement, des exportations et des importations) à titre illustratif mais n'est pas exploitée pour la prévision du PIB.

⁶ Dans la version originelle, Irac et Sédillot recouraient à des ACP plutôt qu'à l'algorithme GETS.

⁷ Plus récemment, Minodier (2010) a adapté ces étalonnages pour prévoir les premiers résultats plutôt que les résultats définitifs des comptes trimestriels. L'auteur montre théoriquement et empiriquement pour la France qu'il peut être préférable d'utiliser comme variable expliquée la série de premiers résultats du PIB (*i.e.* la série constituée des premières publications du PIB par les comptes trimestriels, 45 jours après la fin du trimestre) plutôt que la série définitive (*i.e.* la chronique disponible à l'instant où l'on prévoit) s'il s'agit de prévoir les premiers résultats du taux de croissance du PIB.

types d'étalonnages. Les premiers (Bessec, 2010) sont fondés exclusivement sur des soldes d'enquêtes de l'Insee (industrie, services et bâtiment) : les étalonnages sont également construits au mois le mois et par horizon de prévision ; la sélection des variables explicatives est effectuée en utilisant une variante par blocs de la méthode GETS (Hendry et Krolzig, 2005). Les seconds mettent plus simplement en relation le taux de croissance du PIB au PMI composite.

La Banque de France utilise également des étalonnages reliant le PIB à des données d'enquêtes (le modèle ISMA)⁸ pour le trimestre en cours et pour le trimestre suivant⁹. Cependant ces étalonnages font appel aux modèles à facteurs : l'information apportée par l'enquête de la Banque de France dans l'industrie et les services est résumée en utilisant une analyse en composantes principales (ACP) et une méthode originale de sélection des facteurs, basée sur l'algorithme GETS est employée. De façon plus précise, l'utilisation du modèle à facteurs comporte plusieurs étapes : dans un premier temps, une ACP est effectuée sur les soldes mensuels de chaque secteur ou sous-secteur ; on retient les premiers facteurs de chaque ACP ; nous trimestrialisons ensuite les facteurs retenus¹⁰ ; enfin, l'algorithme GETS est utilisé pour sélectionner ces facteurs préalablement sélectionnés (et leurs retards) comme variables explicatives du taux de croissance du PIB contemporain (pour la prévision du trimestre courant) ou futur (pour la prévision du trimestre suivant).

Plus récemment, deux études ont été menées à la Banque de France sur l'utilisation des modèles à facteurs statiques et dynamiques pour la prévision de court terme de l'activité. De ces travaux est issu un modèle à facteurs utilisé pour la prévision du PIB en zone euro pour le trimestre en cours ou le trimestre précédent. La première étude conduite avec d'autres banques centrales européennes (Barhoumi et al., 2009) porte sur la prévision du taux de croissance du PIB de la zone euro et de dix autres pays la composant, dont la France, aux trimestres courant, suivant et précédent, en tenant compte des délais de parution des indicateurs ; la deuxième (Barhoumi, Darné et Ferrara, 2010) se concentre sur la prévision du PIB français au trimestre suivant et lorsque les indicateurs sont renseignés aux trois mois du trimestre à prévoir. Les variables explicatives utilisées sont des variables d'enquêtes, des variables réelles, financières et de prix. Les deux articles comparent quatre méthodes d'estimation des modèles à facteurs : la méthode de Stock et Watson (1999, 2002a et 2002b), les deux méthodes de Doz, Giannone et Reichlin (2006, 2007) avec estimation en deux étapes (voir section suivante) ou par quasi maximum de vraisemblance, la méthode de Forni et al. (2004, 2005). Elles concluent toutes deux à la supériorité des modèles à facteurs sur les méthodes de référence classiques (combinaison d'étalonnages bivariés, de VAR bivariés ou simple processus AR). Selon la première étude, les modèles estimés par filtre de Kalman semblent donner de meilleurs résultats mais la deuxième étude indique que les résultats ne sont pas significativement différents des méthodes les plus simples. Elle suggère par ailleurs qu'il n'est pas nécessaire de recourir à une base de données très détaillée et que l'utilisation d'un nombre de facteurs, supérieur à celui préconisé par les critères de Bai et Ng (2002), améliore la qualité des prévisions.

1.2 Les organismes internationaux

a) La Banque Centrale Européenne (BCE)

Sur la période récente, la BCE dispose de deux approches concurrentes pour la prévision de court terme du taux de croissance de la zone euro (trimestres précédent, courant et

⁸ L'indicateur synthétique mensuel d'activité (ISMA) a été introduit en 1997. Des modifications ont été apportées depuis. Voir Darné et Brunhes-Lesage (2007) pour une présentation de l'indicateur sous sa forme actuelle.

⁹ Seules les prévisions du PIB au trimestre courant sont publiées en même temps que l'enquête mensuelle de conjoncture dans l'industrie et les services de la Banque de France.

¹⁰ À ce niveau, plusieurs méthodes sont considérées : la moyenne des trois derniers mois connus, la moyenne des seuls mois connus du trimestre ou la moyenne des trois mois du trimestre après avoir prolongé les mois manquants grâce à un processus autorégressif.

suivant)¹¹. Ces deux approches sont utilisées deux fois par mois, en milieu de mois après la publication des indicateurs réels tels l'IPPI et en fin de mois à la sortie des enquêtes et des données financières.

La première approche repose sur la combinaison de prévisions tirés d'une dizaine d'étalonnages classiques (Rünstler et Sédillot, 2003, Diron, 2006) : plusieurs étalonnages simples¹² reliant le taux de croissance du PIB à un petit nombre d'indicateurs mensuels (indicateurs réels, financiers, soldes d'enquêtes) sont estimés tous les mois. Les indicateurs mensuels sont prévus sur la base de simples modèles autorégressifs¹³ sur les mois manquants, puis sont trimestrialisés par moyenne pour alimenter les étalonnages. La prévision finale du taux de croissance du PIB est obtenue par simple moyenne des prévisions des différents étalonnages. Rünstler et Sédillot utilisent seulement comme variables explicatives des indicateurs réels et des soldes d'enquête, tandis que Diron considère également des variables financières.

La deuxième approche repose sur des modèles à facteurs dynamiques mis en place à la BCE (et à la Fed) suivant la méthode présentée par Giannone, Reichlin et Small (2008). Alors que la première approche mobilise relativement peu d'indicateurs mensuels - jusqu'à 15 dans Diron (2006) -, l'ensemble d'information intègre ici 85 indicateurs mensuels, réels, financiers et issus d'enquêtes de conjoncture. Nous utilisons un filtre de Kalman pour calculer les observations manquantes des facteurs dues aux mois manquants des indicateurs mensuels. Le modèle à facteurs est estimé par la méthode d'estimation en deux étapes (ACP puis filtre de Kalman) proposée par Doz, Giannone et Reichlin (2007). Dans ce cadre, Bańbura et Rünstler (2011) proposent une mesure de la contribution des variables aux prévisions et l'appliquent à la prévision de court terme du PIB de la zone euro¹⁴.

Angelini, Camba-Méndez, Giannone, Rünstler et Reichlin (2008) comparent les deux approches et concluent à la supériorité des modèles à facteurs pour la prévision du taux de croissance du PIB de la zone euro à court terme. Angelini, Bańbura et Rünstler (2008) réalisent une étude similaire mais s'intéressent en plus de la prévision du PIB à celle de ses composantes du côté de l'offre et de la demande. Dans les modèles à facteurs, les composantes sont d'abord prévues séparément puis simultanément en introduisant une contrainte (la relation comptable entre le PIB et ses composantes) dans la représentation espace état. Cette étude donne également des résultats favorables aux modèles à facteurs pour la prévision du PIB. Dans une étude portant sur la prévision du taux de croissance de l'activité de la zone euro et de 6 pays européens dont la France, pour un horizon de prévision allant de 1 à 12 trimestres, Caggiano, Kapetanios et Labhard (2009) montrent empiriquement qu'il y a un gain à filtrer préalablement les variables dont sont extraits les facteurs selon les critères proposés par Boivin et Ng (2006) et à combiner les prévisions obtenues de plusieurs modèles à facteurs.

b) La Commission européenne

La Commission européenne publie des prévisions macro-économiques annuelles et infra-annuelles deux fois par an (généralement en mai et en novembre) pour les pays membres de l'Union européenne, la zone euro et l'Union européenne dans son ensemble. Ces prévisions concernent l'année en cours et l'année suivante et sont le résultat d'un processus itératif : elles ne reposent pas sur un modèle économétrique centralisé mais sur l'analyse effectuée par des équipes en charge d'un pays, chacune d'entre elles reposant à des degrés

¹¹ Par ailleurs, des prévisions à moyen terme de l'activité sur l'année en cours et suivante sont construites par la BCE à partir des prévisions nationales du réseau de Banques Centrales nationales et font l'objet d'un processus itératif entre les banques centrales nationales et la BCE.

¹² Diron (2006) utilise 8 étalonnages (avec 15 variables explicatives au total) quand Rünstler et Sédillot (2003) en considèrent 7 (avec 8 variables explicatives). Voir le tableau en annexe pour plus de détails.

¹³ Des spécifications plus élaborées sont utilisées par Rünstler et Sédillot (modèle BVAR notamment).

¹⁴ Ils montrent sur la base de cette statistique que les variables d'enquêtes et financières contribuent davantage à la prévision du PIB de la zone euro que les indicateurs réels une fois les délais de publication des variables pris en compte.

variables sur des méthodes statistiques et des jugements d'experts. La cohérence de ces prévisions, notamment en termes de flux commerciaux, est ensuite vérifiée. Les agrégats sur la zone euro et l'Union européenne ne sont pas prévus directement mais obtenus par agrégation des prévisions des pays. Depuis 2006, la Commission publie également entre ces deux campagnes de prévision une actualisation des prévisions du PIB et de l'inflation pour sept des principaux membres de la zone euro et pour l'année en cours seulement. Ces prévisions dites intérimaires sont en grande partie basées sur des étalonnages classiques.

En parallèle, la Commission européenne disposait jusqu'à fin 2007 pour la prévision de court terme du PIB de la zone euro d'un modèle à facteurs dynamiques SLID (Grenouilleau, 2004 et 2006)¹⁵. Comme dans Stock et Watson (2002a et 2002b), les facteurs étaient déterminés par ACP sur des variables d'enquêtes, réelles et financières trimestrialisées au niveau des secteurs et des pays (les variables mensuelles sont trimestrialisées par moyennes mobiles sur 3 mois en prenant en compte les délais de parution des variables). En comptant les retards, la base comportait 2000 séries. La méthode utilisée se distinguait de la littérature à plusieurs niveaux. En particulier, un filtrage préalable des variables était effectué : n'étaient retenues que les variables fortement corrélées au PIB (le seuil étant déterminé de façon endogène)¹⁶. Par ailleurs, les prévisions étaient calculées de façon itérative. Dans un premier temps, la base dont étaient extraits les facteurs excluait le PIB courant. On obtenait une première estimation de la valeur courante du PIB¹⁷. La série coïncidente de PIB, avec pour dernière valeur l'estimation de première étape, était ensuite incluse dans la base. Cela permettait de produire une nouvelle estimation de la valeur courante du PIB. On remplaçait par cette valeur la dernière observation de la série coïncidente de la base et on ré-estimait une nouvelle valeur courante du PIB. Ces calculs étaient répétés jusqu'à ce que les estimations du PIB entre deux itérations ne diffèrent plus significativement.

c) L'Organisation de coopération et de développement économiques (OCDE)

L'OCDE publie deux fois par an, en juin et en décembre, des prévisions annuelles et infra-annuelles sur un horizon de 18 mois à deux ans dans les perspectives économiques de l'OCDE. Entre chaque publication principale, elle publie également une évaluation intérimaire (généralement en mars et en septembre). Des projections d'un ensemble de variables macro-économiques sont réalisées dans ce cadre pour chacun des 30 pays membres et certains pays hors OCDE. Les prévisions reposent également sur une combinaison de méthodes statistiques et de jugements d'experts. Un modèle développé par Pain et al. (2005) permet de vérifier la cohérence internationale des projections.

Par ailleurs, dans l'évaluation de la situation de court terme, une attention particulière est portée à des étalonnages construits par Sédillot et Pain (2005) pour six grands pays de l'OCDE dont la France et la zone euro. Des étalonnages classiques relient le taux de croissance du PIB à des indicateurs réels, financiers et à des soldes d'enquêtes. Des modèles auxiliaires sont utilisés pour prévoir les mois manquants des indicateurs mensuels (ensuite trimestrialisés). La sélection des variables explicatives de l'étalonnage du PIB et des modèles auxiliaires est réalisée suivant une procédure semi-automatique¹⁸. Dans le cas de la France, les variables retenues sont deux variables d'enquête de l'Insee, la tendance récente de la production et les perspectives personnelles de production dans l'industrie manufacturière et deux variables réelles : l'IPI et la consommation des ménages en produits

¹⁵ Auparavant, la Commission utilisait des étalonnages classiques avec des soldes d'enquêtes, des variables réelles et des variables financières sur la zone euro et sur les États-Unis (Grasmann et Keereman, 2001). La publication des résultats du SLID a été interrompue fin 2007.

¹⁶ Voir aussi Boivin et Ng (2006, 2008) pour une discussion sur la pré-sélection des variables avant l'extraction des facteurs et comme indiqué précédemment, Caggiano, Kapetanios et Labhard (2009) pour une illustration empirique du gain pour la prévision du taux de croissance du PIB français, européen et de plusieurs grands pays.

¹⁷ Pour simplifier, on décrit simplement la prévision du trimestre courant.

¹⁸ Les indicateurs mensuels sont d'abord classés suivant le R² de la relation liant le PIB à ses retards et à l'indicateur et ses retards. Les plus performants (et leurs retards) sont ensuite intégrés dans un modèle liant le PIB à ses retards (jusqu'à 4 retards pour l'endogène et les explicatives) et la meilleure combinaison de variables explicatives parmi toutes les combinaisons possibles est sélectionnée sur la base d'un critère d'information.

manufacturés (pour la France, l'introduction des variables réelles est utile pour la prévision du trimestre courant lorsque les indicateurs sont connus sur les deux premiers mois du trimestre au moins, mais s'avère inutile pour la prévision du trimestre suivant).

Il apparaît donc que les modèles à facteurs sont maintenant beaucoup utilisés dans les institutions européennes, mais qu'ils sont encore relativement peu utilisés pour la prévision de court terme du taux de croissance du PIB français. En appliquant ces modèles à la prévision du taux de croissance du PIB français, notre étude se situe donc dans la lignée des études antérieurement menées par Barhoumi et al. (2009) ou par Barhoumi, Darné et Ferrara (2010) mais, contrairement à ce qui est fait dans ces deux articles, nous ne chercherons pas à comparer les prévisions issues de méthodes d'estimation différentes du modèle à facteurs sous-jacent et nous ne retiendrons que la méthode d'estimation en deux étapes proposée par Doz, Giannone et Reichlin (2007). En revanche, nous étudierons l'impact sur les résultats de diverses méthodes de sélection du modèle à facteurs sous-jacent, et de diverses méthodes de construction de la prévision.

2. Les modèles à facteurs et leur utilisation en prévision

Cette section présente de façon succincte les modèles à facteurs, tout d'abord dans leur forme *statique*, puis dans leur extension *dynamique*. Les différentes méthodes d'estimation de ces modèles sont ensuite présentées, puis on explicite les différentes méthodes pouvant être utilisées pour construire une prévision fondée sur l'estimation préalable d'un modèle à facteurs.

2.1. Les modèles à facteurs

a) Le modèle à facteurs statique

Les modèles à facteurs ont pour but de fournir une représentation parcimonieuse de l'information apportée par un grand nombre de variables lorsque ces variables sont corrélées. Dans ces modèles, nous supposons que les variables observées peuvent être décrites en fonction d'un petit nombre de variables latentes inobservables, appelées facteurs ou facteurs communs (*common factors*), et que ces facteurs communs latents sont la source des corrélations entre les variables observées. Dans le cadre statique, il existe deux types de modèles à facteurs : les modèles à facteurs exacts, dans lesquels les facteurs expliquent toute la corrélation entre les variables, et les modèles à facteurs approchés, adaptés au cadre où le nombre de variables observées tend vers l'infini et dans lesquels les facteurs expliquent la plus grande partie des corrélations entre les variables (la partie résiduelle étant négligeable).

Plus formellement, en notant n le nombre de variables étudiées, T le nombre d'observations dont on dispose pour chaque variable et x_{it} l'observation de la variable i à l'instant t , le modèle exact à q facteurs s'écrit comme suit :

$$x_{it} = \mu_i + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + \dots + \lambda_{iq}f_{qt} + e_{it} , \quad (1)$$

pour $i = 1, \dots, n$, $t = 1, \dots, T$, $q < n$, soit sous forme matricielle :

$$x_t = \mu + \Lambda f_t + e_t , \quad t = 1, \dots, T \quad (2)$$

avec $x_t = (x_{1t}, \dots, x_{nt})'$ et $e_t = (e_{1t}, \dots, e_{nt})'$ des vecteurs de dimension n , $f_t = (f_{1t}, \dots, f_{qt})'$ un vecteur de dimension q , Λ une matrice de dimension (n, q) et avec les hypothèses suivantes : $E(e_t) = 0$, $E(f_t) = 0$, $E(e_t e_t') = D = \text{diag}(d_1, \dots, d_n)$, $E(f_t f_t') = I_q$, $E(f_t e_t') = 0 \quad \forall (t, \tau)$, $E(f_t f_\tau') = 0 \quad \forall (t, \tau)$, $t \neq \tau$, $E(e_t e_\tau') = 0 \quad \forall (t, \tau)$, $t \neq \tau$.

Dans la suite, on se ramènera au cas où $\mu=0$ et on travaillera avec des variables préalablement centrées.

Lorsque q est très petit devant n , le modèle permet effectivement d'obtenir une représentation parcimonieuse des covariances entre les x_{it} , comme on le verra ci-dessous.

Dans ce modèle statique, les q facteurs communs ne sont pas auto-corrélés. On peut en outre supposer, sans perte de généralité, qu'ils ne sont pas corrélés entre eux, et qu'ils sont de variance unitaire. Le terme e_{it} appelé composante spécifique ou idiosyncratique (*idiosyncratic component*) représente la part de la variable x_{it} qui n'est pas expliquée par les facteurs communs. Comme les e_{it} sont deux à deux non corrélés, toute la corrélation entre les variables observées passe par les facteurs.

Les poids ou *factor loadings* λ_{ij} mesurent la covariance entre la variable observée i et le facteur commun j . La variance de chaque variable peut alors se réécrire comme :

$V(x_{it}) = \sum_{j=1}^q \lambda_{ij}^2 + d_i$. Le terme λ_{ij}^2 représente la part de la variance de x_{it} expliquée par le facteur j et le terme $\sum_{j=1}^q \lambda_{ij}^2$ la part totale de la variance (*communality*) captée par les q

facteurs. En outre, la matrice de variance-covariance du vecteur des variables observées s'écrit : $V(x_t) = \Lambda\Lambda' + D$ et, comme D est diagonale, les covariances entre les variables observées s'expriment explicitement en fonction des *factor loadings*. Ainsi, la matrice de variance-covariance de x_t s'exprime en fonction des $n(q+1)$ paramètres de Λ et D au lieu de dépendre de $n(n+1)/2$ paramètres si on ne suppose pas l'existence d'un modèle à facteurs. Par ailleurs, il faut noter que le modèle est invariant par changement d'échelle, si bien qu'il est équivalent de décomposer la matrice de variance-covariance de x_t ou sa matrice de corrélation.

Dans le modèle statique approché, on ne suppose plus que les termes idiosyncratiques sont deux à deux non corrélés. Nous supposons simplement que, dans la corrélation entre les variables observées, la part qui est due à la corrélation entre les termes idiosyncratiques est négligeable devant la part due aux facteurs communs. Si l'on continue à noter $E(e_t e_t') = D$ (avec ici une matrice D non diagonale), on fait l'hypothèse que lorsque le nombre n des variables observées tend vers l'infini, la matrice D reste bornée alors que la matrice $\Lambda\Lambda'$ est non bornée. Ainsi, comme $V(x_t) = \Lambda\Lambda' + D$, on peut considérer que la part de la corrélation entre variables qui n'est pas expliquée par les facteurs est négligeable.

b) Le modèle à facteurs dynamiques

Les modèles à facteurs dynamiques ont pour objet une description parcimonieuse de la dynamique commune aux variables observées (ou des co-mouvements des variables observées). Ces modèles généralisent les modèles statiques (exacts ou approchés) de deux façons : d'une part, les facteurs communs sont autocorrélés (en général leur dynamique est modélisée sous une forme VAR ou éventuellement VARMA), et d'autre part les variables observées peuvent être affectées par les valeurs contemporaines des facteurs, mais aussi par leurs valeurs retardées. Dans les deux cas, le modèle peut se ramener, moyennant des changements de notations appropriés, à une forme qui est proche de celle des modèles à facteurs statiques.

Considérons tout d'abord le cadre des modèles à facteurs dynamiques exacts. Dans ce cadre, si l'on suppose que la dynamique des facteurs est correctement représentée par un modèle VAR(p), et si l'on note toujours $x_t = (x_{1t}, \dots, x_{nt})'$ le vecteur des variables observées,

on peut définir une première classe de modèles dans lequel les facteurs n'interviennent que par l'intermédiaire de leur valeur contemporaine. Ces modèles ont la forme suivante :

$$x_t = \Lambda_0 f_t + e_t \quad (3)$$

$$f_t = \sum_{i=1}^p A_i^0 f_{t-i} + \varepsilon_t \quad (4)$$

avec (ε_t) un bruit blanc, (e_t) un processus (qui peut être un bruit blanc ou avoir une dynamique) dont les composantes sont deux à deux non corrélées ($E(e_{it}e_{jt}) = 0 \forall t, \tau, i, j, i \neq j$), et sont non corrélées aux facteurs.

Si l'on note $F_t = (f_t', \dots, f_{t-p+1}')$, on montre facilement que ce type de modèle peut être aussi écrit sous la forme suivante :

$$x_t = \Lambda F_t + e_t \quad (5)$$

$$F_t = A F_{t-1} + \underbrace{B \varepsilon_t}_{\xi_t} \quad (6)$$

avec Λ, A des matrices qui s'expriment de façon simple en fonction des matrices $\Lambda_0, A_1, \dots, A_p$ et $B = (I_q, 0 \dots 0)'$. Le vecteur $F_t = (f_t', \dots, f_{t-p+1}')$ est ici un vecteur de taille $r = pq$, et on dit que l'on a un modèle dans lequel le nombre de facteurs statiques est r (les facteurs statiques sont les composantes de F_t), alors que le nombre de facteurs dynamiques est q (les facteurs dynamiques sont les composantes de f_t).

Le même type de formulation peut aussi être obtenu lorsque le facteur intervient non seulement de manière contemporaine mais aussi avec ses retards, c'est-à-dire dans le cadre de modèle de la forme :

$$x_t = \Lambda_0 f_t + \dots + \Lambda_s f_{t-s} + e_t \quad (7)$$

$$f_t = \sum_{i=1}^p A_i^0 f_{t-i} + \varepsilon_t \quad (8)$$

Ainsi, la forme générale d'un modèle exact à facteurs dynamiques est donnée par :

$$\begin{cases} x_t = \Lambda F_t + e_t & \text{avec } V(e_t) = \Phi = \text{diag}(\varphi_1, \dots, \varphi_n) \\ F_t = A F_{t-1} + \underbrace{B \varepsilon_t}_{\xi_t} & \text{avec } V(\xi_t) = \Sigma_\xi \end{cases} \quad \begin{matrix} (9a) \\ (9b) \end{matrix}$$

où (ε_t) est un bruit blanc, (e_t) un processus (qui peut être un bruit blanc ou avoir une dynamique) dont les composantes sont deux à deux non corrélées ($E(e_{it}e_{jt}) = 0, \forall t, \tau, i, j, i \neq j$) et sont non corrélées aux facteurs et B une matrice (r, q) de rang q avec $q \leq r$. On dit alors que r est le nombre de facteurs statiques du modèle et que q est le nombre de facteurs dynamiques.

Comme dans le cas statique, on étend le champ d'application de ces modèles en introduisant le modèle à facteurs dynamiques approché, lorsque le nombre n des variables observables tend vers l'infini. Dans ce type de modèle, on autorise les composantes du vecteur e_t à être corrélées entre elles, mais on suppose que la part de la dynamique des observables qui est liée aux composantes idiosyncratiques est négligeable devant la part liée aux facteurs.

Le système d'équations (9) a une forme qui relève de la notion générale des modèles *espace état* ou modèles *état mesure*. L'équation (9a) appelée *équation de mesure* décrit la relation entre la variable observée dite *variable de mesure*, ici x_t , et la *variable d'état*

inobservée, ici F_t . La deuxième équation (9b) appelée *équation d'état* décrit comment les variables latentes sont générées à partir de leurs retards et d'innovations (une présentation de la représentation espace état et des algorithmes d'estimation est donnée en annexe).

2.2 Estimation d'un modèle à facteurs dynamiques

Cette étude se situe dans le cadre des modèles à facteurs dynamiques approchés, qui est le cadre communément retenu pour l'analyse de données macroéconomiques. Diverses méthodes d'estimation de ces modèles ont été proposées dans la littérature (voir Bai et Ng (2008b) ou Stock et Watson (2010) pour un survol complet de ces méthodes).

La méthode la plus couramment employée est celle de l'Analyse en Composantes Principales (ACP), initialement proposée par Stock et Watson (2002a). Cette méthode est utilisée dans le cadre d'un modèle à facteurs statiques (ou d'un modèle à facteurs dynamiques ramené à une forme statique suivant la démarche explicitée ci-dessus). Sous les hypothèses usuellement retenues dans la spécification du modèle à facteurs approchés, on montre que l'ACP¹⁹ permet d'obtenir des estimateurs convergents des paramètres du modèle et une approximation des facteurs qui converge vers leur vraie valeur lorsque le nombre n des séries étudiées et le nombre T des observations tendent vers l'infini.

Cependant, d'autres méthodes d'estimation ont été proposées pour permettre la prise en compte de la dynamique des facteurs. Forni *et al.* (2000) ont proposé une méthode d'estimation fondée sur l'analyse de la densité spectrale des observations. Doz, Giannone et Reichlin (2006, 2007) ont proposé, d'une part, une méthode d'estimation par pseudo-maximum de vraisemblance et, d'autre part, une méthode d'estimation en deux étapes basée sur le filtre de Kalman.

Cette méthode d'estimation en deux étapes est assez simple à mettre en œuvre et présente en outre l'avantage de s'adapter facilement au cas de valeurs manquantes qui est, comme nous l'avons dit précédemment, un des problèmes importants auquel le conjoncturiste doit faire face. Elle a été, par exemple, utilisée par Giannone, Reichlin et Small (2008) pour effectuer des prévisions du PIB des États-Unis puis de la zone euro ou par Angelina *et al.* (2008) et Bańbura et Rünstler (2011) pour calculer une prévision de court terme du PIB de la zone euro.

Il importe ici de souligner que la mise en œuvre d'une ACP impose de disposer d'un échantillon cylindré de données et que ceci est une contrainte très pénalisante lorsqu'on s'intéresse à la prévision conjoncturelle. En effet, si l'on tronque l'échantillon à la dernière date où toutes les données sont disponibles, on se prive d'une partie de l'information dont on dispose.

La méthode en deux étapes proposée par Doz, Giannone et Reichlin (2007) permet au contraire de calculer la meilleure approximation de la valeur des facteurs à chaque date en tenant compte de toute l'information disponible. En effet, sous l'hypothèse de normalité des perturbations, on sait que le filtre et le lisseur de Kalman permettent d'obtenir, pour une valeur donnée des paramètres, l'approximation optimale des variables latentes en fonction de l'ensemble de l'information disponible sur les variables observables. Il nous semble donc qu'elle est particulièrement adaptée à la problématique de la prévision conjoncturelle et c'est donc cette méthode que nous retiendrons ici.

Si l'on note $\theta = (\Lambda, A_1, \dots, A_p, \Phi, \Sigma_\varepsilon, B)$ les paramètres du modèle, cette méthode comprend les deux étapes suivantes :

1. La matrice Λ de l'équation (9a) est d'abord estimée par ACP sur le sous-échantillon de données cylindrées, *i.e.* sur la période commune où toutes les variables x_{it} sont renseignées. On calcule aussi une estimation préliminaire des facteurs à chaque date

¹⁹ En pratique l'ACP est menée sur la matrice de corrélation empirique des observations, puisque le modèle est invariant par changement d'échelle.

du sous-échantillon. Cette estimation préliminaire des facteurs est utilisée pour obtenir une estimation des paramètres de l'équation (9b) en utilisant des techniques de régression standard. La matrice B est estimée par ACP sur les résidus estimés $\hat{\xi}_t$.

2. Les facteurs sont ensuite estimés sur toute la période d'estimation, *i.e.* en incluant les instants du temps où certaines variables présentent des observations manquantes. Ceci peut être fait grâce aux algorithmes de filtrage et de lissage de Kalman appliqués à la représentation espace état (9).

À l'itération t , suivant que la variable i est disponible ou non, on pose :

$$E(e_{it}^2) = \begin{cases} \varphi_i & \text{si } x_{it} \text{ est disponible} \\ +\infty & \text{sinon} \end{cases}$$

dans la matrice R de l'algorithme (cf. annexe) avec φ_i la variance de e_{it} estimée sur l'échantillon cylindré. Autrement dit, si la variable i n'est pas renseignée à l'instant t , on remplace le terme diagonal à la ligne i par un terme infini dans la matrice R de l'algorithme de filtrage. La matrice étant ensuite inversée, ceci équivaut à donner un poids nul à la variable i à l'itération t . On en déduit une nouvelle estimation des facteurs sur toute la période : par construction, il s'agit de l'approximation optimale des facteurs en fonction de l'ensemble de l'information disponible.

2.3 Utilisation en prévision

Lorsque, comme c'est le cas ici, la variable à prévoir est une variable trimestrielle (notée y_t), alors que les données utilisées, et donc aussi les facteurs, sont à valeurs mensuelles, la prévision repose tout d'abord sur une trimestrialisation des facteurs. Si l'on note $f_{i,t}^Q$ la valeurs trimestrialisée du $i^{\text{ème}}$ facteur à la date t issue de l'estimation du modèle à facteurs, la prévision repose sur une régression de y_t (ici le taux de croissance trimestriel du PIB) sur les $f_{i,t}^Q$.

Deux approches sont alors généralement utilisées et, dans le cadre de cette étude, nous utiliserons les deux approches afin de comparer leurs performances en prévision. La première approche est adaptée au cas des modèles à facteurs statiques, mais peut évidemment aussi être utilisée dans le cadre des modèles à facteurs dynamiques (dont on a vu qu'ils admettent une représentation statique). Elle consiste à estimer par MCO le modèle :

$$y_{t+h} = \sum_{i=1}^r \delta_i f_{i,t}^Q + \varepsilon_{t+h} \quad t = 1, \dots, T-h \quad (10a)$$

et à calculer ensuite la prévision de y_{T+h} à la date T en utilisant la formule suivante :

$$\hat{y}_{T+h|T} = \sum_{i=1}^r \hat{\delta}_i f_{i,T}^Q \quad (10b)$$

Dans le cadre de la présente étude, lorsqu'on utilisera cette approche, on le fera en pratique en utilisant deux équations, l'une utilisée pour la prévision du PIB au trimestre courant et au trimestre précédent, et l'autre utilisée pour la prévision au trimestre suivant. Plus précisément : pour la prévision du trimestre courant et du trimestre précédent, on estime l'équation (10a) avec $h = 0$, puis on calcule la prévision en utilisant l'équation (10b) pour $h = 0$ et $h = -1$; pour la prévision du trimestre suivant, on utilise les équations (10a) et (10b) avec $h = 1$. Ce type d'approche qui a été introduit, dans un cadre statique, par Stock et Watson (2002a, 2002b) est utilisé par exemple par Boivin et Ng (2006), Barhoumi et al. (2009) ou Barhoumi, Darné et Ferrara (2010).

La deuxième approche est spécifiquement liée au cadre dynamique et utilise l'estimation de la dynamique des facteurs qui est obtenue lorsqu'on estime le modèle à facteurs. En effet, si les facteurs vérifient un modèle de la forme $f_t = \sum_{i=1}^p A_i^0 f_{t-i} + \varepsilon_t$, il est possible d'obtenir de façon récursive une prévision $f_{T+h|T}$ de f_{T+h} à la date T en utilisant les valeurs estimées des matrices A_i^0 et des facteurs. On peut ensuite trimestrialiser les prévisions obtenues et déterminer une prévision $f_{i,T+h|T}^Q$ du facteur trimestrialisé. On estime alors par les moindres carrés ordinaires l'équation reliant le taux de croissance du PIB aux facteurs trimestrialisés qui lui sont contemporains $f_{i,t}^Q$:

$$y_t = \sum_{i=1}^r \delta_i f_{i,t}^Q + \varepsilon_t \quad t = 1, \dots, T \quad (11a)$$

(cette équation coïncide avec l'équation (10a) lorsque cette dernière est estimée avec $h = 0$), et la prévision de y_{T+h} à la date T est obtenue en utilisant la formule suivante :

$$\hat{y}_{T+h|T} = \sum_{i=1}^r \hat{\delta}_i f_{i,T+h|T}^Q \quad (11b)$$

Ce type d'approche est utilisé par Giannone, Reichlin et Small (2008), par Angelini, Bańbura et Rünstler (2008), ou par Bańbura et Rünstler (2011).

Dans cette étude, nous proposons en outre de comparer les prévisions obtenues en utilisant ces deux approches à celles qui sont obtenues en utilisant une troisième méthode, qui est en quelque sorte intermédiaire entre les deux précédentes. Dans cette nouvelle approche, nous utilisons les équations (10a) et (10b), mais en modifiant la démarche lorsque la date T à laquelle la prévision est faite correspond au premier ou au deuxième mois d'un trimestre. Plus précisément, nous utilisons l'équation (10a) comme précédemment, mais nous utilisons la représentation VAR pour obtenir une prévision des facteurs sur l'ensemble des mois du trimestre concerné, et nous calculons la prévision en appliquant l'équation (10b) au facteur trimestrialisé associé.

2.4 Le choix de la spécification du modèle

Jusqu'à présent, nous avons supposé le nombre de facteurs et l'ordre du VAR sur les facteurs connus, ce qui n'est bien sûr pas le cas en pratique. Dans notre application, nous avons utilisé deux méthodes alternatives pour choisir la spécification du modèle.

a) Les critères d'information

Nous utilisons d'abord les critères d'information classiques AIC ou BIC pour le choix de l'ordre p du VAR :

$$AIC(p) = \ln[\det(\hat{\Sigma}_\varepsilon)] + \frac{2N^2 p}{T} \quad \text{et} \quad BIC(p) = \ln[\det(\hat{\Sigma}_\varepsilon)] + \frac{N^2 p \ln(T)}{T}$$

où T est le nombre d'observations, N le nombre de variables du système, $\hat{\Sigma}_\varepsilon$ la matrice de variance covariance des résidus estimés du VAR. On choisit l'ordre $p \in \{1, \dots, p_{\max}\}$ qui minimise le critère d'information.

Bai et Ng (2002, 2007) ont par ailleurs proposé des critères pour le choix du nombre de facteurs. Dans leur article de 2002, ils proposent une première série de critères adaptés au cas des modèles à facteurs statiques :

$$PC_1(r) = V(r, \hat{F}_r) + r \frac{N+T}{NT} \ln \left(\frac{NT}{N+T} \right) \quad IC_1(r) = \ln(V(r, \hat{F}_r)) + r \frac{N+T}{NT} \ln \left(\frac{NT}{N+T} \right)$$

$$PC_2(r) = V(r, \hat{F}_r) + r \frac{N+T}{NT} \ln C_{NT}^2 \quad IC_2(r) = \ln(V(r, \hat{F}_r)) + r \frac{N+T}{NT} \ln C_{NT}^2$$

$$PC_3(r) = V(r, \hat{F}_r) + r \frac{\ln C_{NT}^2}{C_{NT}^2} \quad IC_3(r) = \ln(V(r, \hat{F}_r)) + r \frac{\ln C_{NT}^2}{C_{NT}^2}$$

avec $C_{NT}^2 = \min(N, T)$ et $V(k, \hat{F}_k) = N^{-1} \sum_{i=1}^N \hat{\sigma}_i^2$ où $\hat{\sigma}_i^2 = \hat{e}_i' \hat{e}_i / T$. Le nombre de facteurs à retenir, lorsqu'on applique un de ces critères, est le nombre $r \in \{1, \dots, r_{\max}\}$ qui minimise le critère concerné.

Dans leur article de 2007, les mêmes auteurs ont proposé une deuxième série de critères pour déterminer le nombre de facteurs dynamiques q . Soit Σ_ε la matrice de variance des résidus du VAR estimé sur les r facteurs statiques, et soient $c_1 \geq c_2 \geq \dots \geq c_r \geq 0$ les

valeurs propres ordonnées de la matrice Σ_ε . On définit les quantités $D_{1,k} = \left(\frac{c_{k+1}^2}{\sum_{j=1}^r c_j^2} \right)^{1/2}$ et

$D_{2,k} = \left(\frac{\sum_{j=k+1}^r c_j^2}{\sum_{j=1}^r c_j^2} \right)^{1/2}$, et on leur associe les deux règles de décision suivantes :

$$\kappa_1 = \{k : D_{1,k} < m / \min[N^{0,5-\delta}, T^{0,5-\delta}]\} \text{ et } \kappa_2 = \{k : D_{2,k} < m / \min[N^{0,5-\delta}, T^{0,5-\delta}]\}$$

On choisit $q_1 = \min\{q \in \kappa_1\}$ ou $q_2 = \min\{q \in \kappa_2\}$. Les auteurs utilisent $\delta = 0,1$ et montrent au travers de simulations que l'on obtient des résultats corrects pour $m = 1^{20}$.

En pratique, ces différents critères sont utilisés en trois temps. Nous utilisons d'abord l'un des six critères PC ou IC (Bai et Ng, 2002) pour déterminer le nombre de facteurs optimal $r^* \in \{1, \dots, r_{\max}\}$ dans un cadre statique. Nous estimons ensuite un VAR sur ces r^* facteurs et nous choisissons l'ordre $p^* \in \{1, \dots, p_{\max}\}$ du VAR de façon à minimiser le critère AIC ou BIC usuel. Enfin, nous appliquons les critères de Bai et Ng (2007) sur la matrice de variance covariance ou de corrélation des résidus du VAR(p^*) pour obtenir le nombre de facteurs dynamiques optimal q^* .

²⁰ Ils proposent également une variante fondée sur la matrice de corrélation des résidus du VAR. Sur la base de simulations, les auteurs recommandent d'utiliser dans ce cas $m=1,25$ pour le premier critère et $m=2,25$ pour le second.

b) Le critère RMSE

Plusieurs études montrent qu'en pratique, l'utilisation des critères de Bai et Ng peut conduire à choisir un nombre trop restreint de facteurs, ce qui dégrade la qualité des prévisions (voir par exemple Barhoumi, Darné et Ferrara (2010), pour une application sur la prévision du PIB français et Schumacher (2007) pour une illustration sur le PIB allemand). Une explication possible est que le choix de la spécification du modèle à facteurs est fait de façon totalement indépendante de la variable à prévoir.

Aussi, comme Schumacher (2007), nous proposons, comme alternative aux critères d'information et dans une optique de comparaison des résultats obtenus, de choisir le nombre de facteurs qui minimise le critère RMSE dans la régression du PIB sur les facteurs :

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \quad \text{avec} \quad \hat{y}_t = \sum_{i=1}^r \hat{\delta}_i f_{i,t}^0 \quad t = 1, \dots, T$$

avec \hat{y}_t le PIB estimé tiré de l'équation (10a) ou (11a) suivant la méthode utilisée. L'ordre p du processus VAR sur les facteurs est également choisi en fonction du critère RMSE.

3 Utilisation des modèles à facteurs dynamiques pour la prévision du PIB français

3.1 Les données

La base de données utilisée est constituée de 93 variables (voir tableau 4). Comme la plupart des travaux portant sur la prévision du PIB à partir des modèles à facteurs (voir tableau 3), nous avons retenu quatre groupes de variables :

- des *soldes d'enquêtes* : les principaux soldes des enquêtes de l'Insee rentrant dans la construction des indicateurs synthétiques (ou climat des affaires) dans l'industrie, les services, le bâtiment, le commerce de détail et l'enquête auprès des ménages ;
- des *variables réelles* : la consommation des ménages en produits manufacturés et ses composantes, les immatriculations de véhicules neufs, les mises en chantier et les permis de construire, l'indice de production industrielle et ses composantes et des variables sur le marché du travail ;
- des *variables nominales* – monétaires et financières : les taux d'intérêts, la pente des taux, plusieurs indices boursiers, un indice de volatilité du marché, les agrégats monétaires et des indices de prix ;
- des *indicateurs de l'environnement international* : les taux de change de l'euro par rapport aux grandes devises et des indicateurs sur l'économie allemande et américaine.

Beaucoup de conjoncturistes fondent leurs prévisions sur des variables d'enquête et, au fur et à mesure de leur disponibilité, sur des variables réelles, en particulier l'indice de production industrielle, la consommation des ménages en produits manufacturés, les mises en chantier et permis de construire et les données douanières du commerce extérieur (voir section 1). Les variables nominales et internationales considérées ici sont en revanche moins utilisées dans les étalonnages. Nous verrons dans nos simulations qu'il y a un gain à mobiliser ces variables, en particulier pour la prévision du trimestre suivant.

Les estimations présentées ici sont réalisées sur les séries publiées début juin 2010. Faute de séries disponibles, nous n'étudions pas l'impact des révisions des données sur les résultats. Les variables sont corrigées des variations saisonnières²¹ et sont, pour la plupart, de fréquence mensuelle. Les séries trimestrielles ont été mensualisées en utilisant un *spline*

²¹ Les séries de construction neuve (mises en chantier et permis de construire) et les agrégats monétaires fournis bruts par le MEDMM et la Banque de France ont été désaisonnalisées en appliquant la méthode X12-ARIMA aux séries prises en logarithme.

cubique²². Certaines variables financières publiées à une fréquence journalière ou hebdomadaire sont mensualisées en prenant la dernière observation du mois. À noter enfin les délais de publication différents des séries : certaines variables sont disponibles au cours du mois qu'elles renseignent (les soldes d'enquêtes et les variables financières) mais les variables réelles sont connues avec un ou deux mois de retard.

Une analyse de la stationnarité des variables a été réalisée. Nous avons appliqué plusieurs tests de racine unitaire aux séries (tests ADF, Phillips Perron et KPSS) à l'ensemble des séries (prises en logarithme pour les séries réelles et financières, sauf les taux d'intérêt et les séries sur le marché du travail). Afin de contrôler l'effet de la dernière récession conduisant en fin de période à un saut en niveau sévère dans la plupart des séries, les tests ont été réalisés sur la période 1980-2007. Les résultats sont résumés dans le tableau 5. À peu d'exceptions près, les soldes d'enquête sont stationnaires, les variables réelles et financières intégrées d'ordre un (sauf la pente des taux d'intérêt stationnaire selon la plupart des tests). Suivant ces résultats, les variables non stationnaires ont été différenciées de façon à obtenir des variables stationnaires, puis toutes les variables ont été centrées et réduites.

3.2 Analyse en échantillon et à information complète

L'analyse est d'abord réalisée à ensemble d'information complet, c'est-à-dire sans prendre en compte les délais de publication des indicateurs. Autrement dit, les résultats décrits ici sont ceux qui seraient obtenus quelques jours avant la publication des comptes trimestriels, lorsque l'on dispose notamment de l'indice de production industrielle sur les trois mois du trimestre.

Tableau 1 : Résultats d'estimation sur l'échantillon complet

Trimestre prévu	T-1 et T	T+1
Équation estimée	Équation (10a) avec h= 0	Équation (10a) avec h=1
C	0,41 [14,13]	0,41 [13,00]
F1	0,10 [13,43]	0,06 [7,48]
F2	0,11 [7,75]	0,14 [8,83]
F3	-0,06 [-2,13]	-
F5	-	-0,22 [-5,27]
F9	-0,36 [-2,83]	-
F10	-0,20 [-2,52]	0,40 [5,15]
R ²	0,78	0,74
R ² bar	0,77	0,73
LM(4)	0,13	0,39
Hétérosc.	0,93	0,27
JB	0,87	0,64

Les t-statistiques sont données entre crochets sous les coefficients estimés. Les dernières lignes reportent les probabilités critiques du test de Breusch Godfrey d'autocorrélation d'ordre 4, du test d'hétéroscédasticité de Breusch Pagan Godfrey et du test de normalité de Jarque Bera.

Source : calculs des auteurs.

Nous reportons d'abord les résultats d'estimation des modèles sur la période 1990Q1-2009Q3 et en utilisant les 93 variables. Sur l'échantillon total et pour un nombre de facteurs dynamiques q maximal de 5 et des retards s et p maximaux de 2 et 3 respectivement, les critères de Bai et Ng (2002, 2007) conduisent à retenir $q=5$ et $s=1$, et le critère BIC un ordre

²² Il s'agit d'une technique d'interpolation des données fondée sur un polynôme du troisième degré.

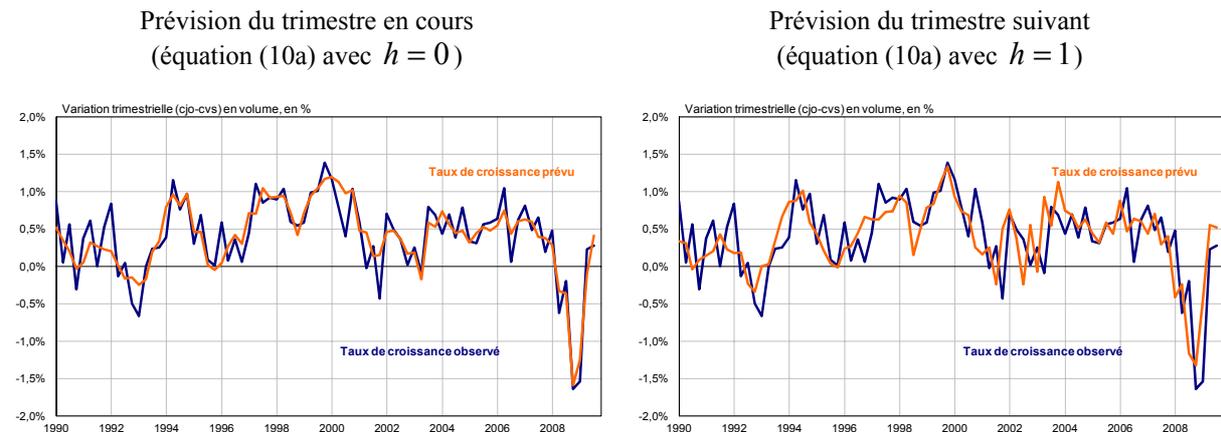
$p=1$ pour le VAR sur les facteurs (on utilise le critère IC_2 en statique et les critères κ_1 et κ_2 fondés sur la matrice de covariance en dynamique)²³.

Nous estimons donc tout d'abord le modèle associé à cette spécification, en utilisant la méthode en deux étapes (ACP puis filtre et lisseur de Kalman) proposée par Doz, Giannone et Reichlin (2007). Nous estimons ensuite l'équation (10a) dans les deux spécifications qui sont utilisées à l'étape de prévision, c'est-à-dire la spécification associée à $h = 0$ (utilisée pour la prévision du trimestre courant et du trimestre précédent) et la spécification associée à $h = 1$ (utilisée pour la prévision du trimestre suivant).

Nous présentons dans le Tableau 1 les résultats de ces estimations. Après élimination des facteurs non significatifs, les deux équations comportent cinq variables explicatives, presque toutes identiques. L'ajustement semble de très bonne qualité (coefficients de détermination respectivement de 78 % et 74 %). Ce résultat doit être relativisé dans le cas où $h = 0$ puisque l'ajustement est réalisé en supposant que l'ensemble des indicateurs est connu sur les trois mois du trimestre à estimer (à titre de comparaison, un processus autorégressif complété de l'indice de production industrielle connu sur l'ensemble du trimestre conduit à un R^2 très élevé aussi, de l'ordre de 0,65), mais il est encourageant dans le cas où $h = 1$.

Les graphiques 1 retracent le taux de croissance du PIB observé et ajusté pour ces deux équations. Les deux modèles reproduisent assez fidèlement les fluctuations du PIB même si les creux de 1993 et 2001 ne sont pas bien captés. En revanche, l'ajustement est de bonne qualité pendant la dernière récession en 2008-2009.

Graphiques 1 : taux de croissance du PIB observé et ajusté



Source : Insee, calculs des auteurs.

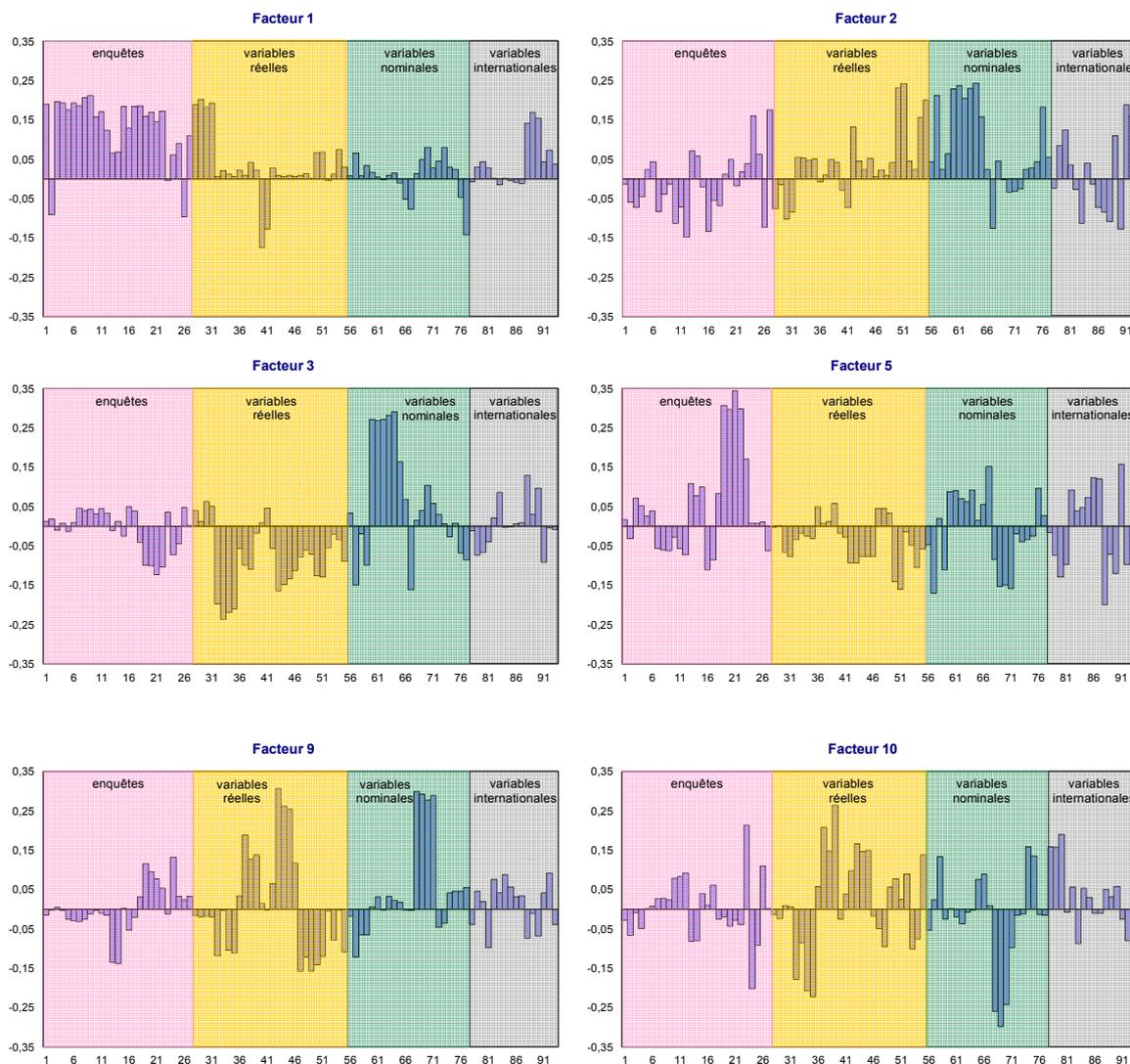
Les graphiques 2 présentent le poids de chacune des 93 variables (centrées et réduites) dans la construction des six facteurs retenus dans les deux modèles.

- le premier axe reflète principalement l'information apportée sur l'activité par les variables d'enquête (dans l'industrie, les services, le bâtiment et le commerce de détail) ;
- les deuxième et troisième facteurs sont davantage de nature financière, au vu du poids important des indices boursiers dans leur construction ;
- le cinquième facteur pourrait retracer les comportements de consommation compte tenu de la part importante des soldes de l'enquête auprès des ménages, du montant des prêts et des taux d'intérêt ;

²³ Le critère RMSE conduit à retenir une représentation moins parcimonieuse avec $q=5$, $s=2$ et $p=2$. Nous étudierons dans la suite les différences entre ces deux spécifications dans les résultats hors échantillon.

- le neuvième facteur dépend fortement des indices de production industrielle, des taux d'intérêt et des permis de construire ;
- le dixième facteur est principalement lié aux indicateurs sur la construction (les mises en chantier, les permis de construire, les taux d'intérêt notamment sur les prêts immobilier et le solde sur l'opportunité de faire des achats importants dans l'enquête auprès des ménages).

Graphiques 2 : construction des facteurs



Source : calculs des auteurs.

3.3 Analyse hors échantillon et en pseudo temps réel

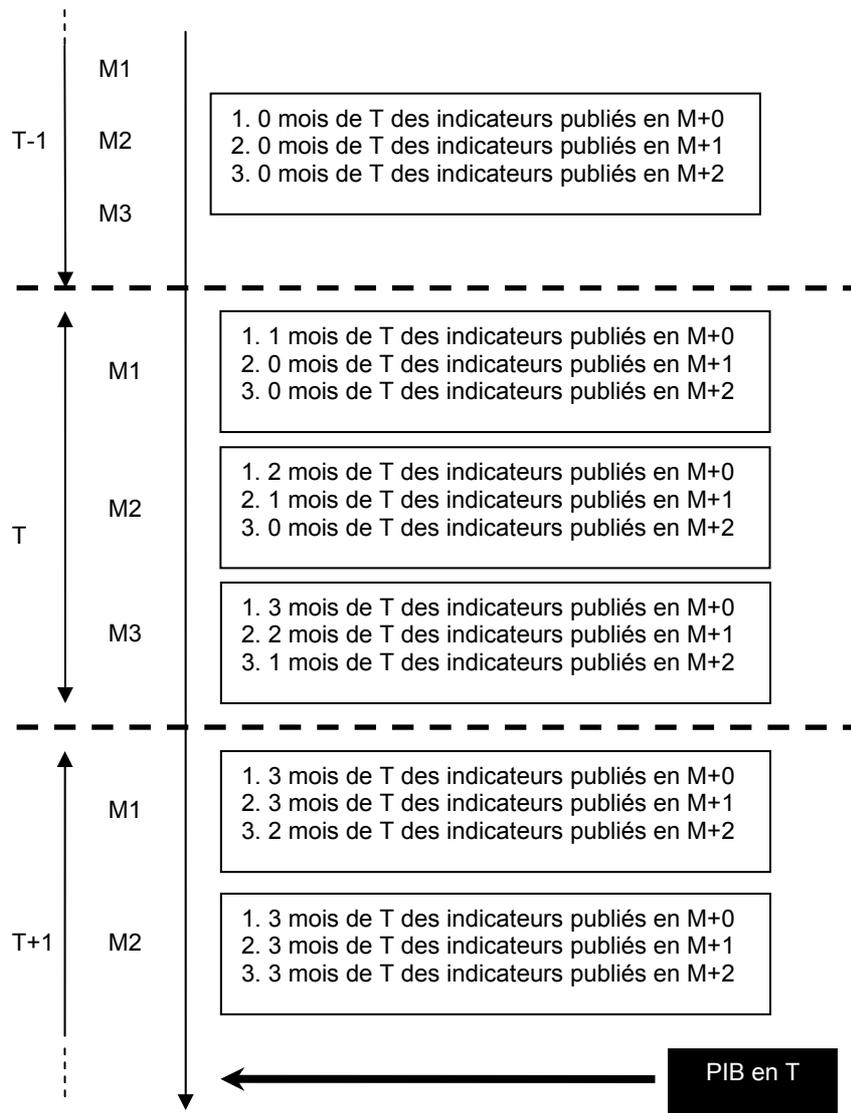
a) Description du cadre expérimental

Nous évaluons maintenant la qualité des prévisions en nous plaçant dans des conditions réelles de prévision en faisant une évaluation des prévisions hors échantillon, c'est-à-dire hors de la période d'estimation et en tenant compte du délai de publication des variables.

Compte tenu des différents délais de parution des variables, nous évaluerons les modèles de prévision en huit instants du temps en partant du premier mois du trimestre précédent $T-1$ (pas d'information sur le trimestre T) au mois 2 du trimestre suivant $T+1$ où l'on dispose des

observations des indicateurs à tous les mois du trimestre T . Le diagramme qui suit présente la chronologie des exercices de prévisions du PIB du trimestre T .

Graphique 3 : calendrier des prévisions et évolution de l'ensemble d'information



Note : les « indicateurs publiés en M+0 » désignent les indicateurs mensuels publiés au cours du mois (par exemple, le jugement des industriels sur leur activité en janvier est connu à la fin du mois de janvier), les « indicateurs publiés en M+1 » les indicateurs mensuels publiés avec 1 mois de retard (par exemple, la consommation manufacturière au mois de janvier est connue fin février) et les « indicateurs publiés en M+2 » les indicateurs connus avec 2 mois de retard (par exemple, l'IPI du mois de janvier est publié début mars).

De façon classique, l'analyse hors échantillon est réalisée comme suit. On retire de l'échantillon de 80 observations les 39 dernières, on estime les modèles sur la période restante, du premier trimestre de 1990 au quatrième trimestre de 1999, puis on construit les prévisions aux différents horizons et les erreurs de prévision associées. On augmente ensuite l'échantillon d'une observation, on ré-estime les modèles, on calcule les nouvelles prévisions et les erreurs de prévision correspondantes. On augmente ensuite de nouveau l'échantillon d'une observation, on réplique les calculs précédents et ainsi de suite jusqu'à ce que l'on ait parcouru les 39 dernières observations.

Il est important de procéder de la sorte car les résultats peuvent différer sensiblement en échantillon et hors échantillon. Il faut donc privilégier l'évaluation hors échantillon qui correspond aux conditions réelles d'exercice du prévisionniste. L'évaluation diffère

cependant des conditions réelles sur deux plans. En premier lieu, l'exercice n'est pas totalement hors échantillon car le choix de la spécification (le choix de l'ordre p du VAR et du nombre de facteurs) a été réalisé sur l'échantillon total. Comme mentionné précédemment, il s'agit par ailleurs d'un exercice en *pseudo temps réel* : nous tenons compte des délais de publications des séries et nous tronquons chaque série de façon à n'utiliser que les valeurs de la série qui auraient été disponibles à la date où la prévision est calculée, mais nous ne prenons pas en compte les révisions des séries dans le temps²⁴ (or les indicateurs quantitatifs réels comme le PIB, l'IPI ou la consommation des ménages en produits manufacturés sont fortement révisés).

À l'issue de cette expérience, on dispose d'une série d'erreurs de prévision pour chacun des huit horizons de prévision auxquelles sont appliqués les critères d'évaluation classiques : les critères RMSFE et MAFE qui mesurent la dispersion des erreurs de prévision et leur amplitude moyenne et le test de Pesaran et Timmermann qui permet d'évaluer l'aptitude des modèles à prévoir le sens de variation de la série (soit ici à prévoir les accélérations ou décélération de l'activité). On s'assurera à ce niveau que la proportion de sens de variation correctement prédits excède 50 % (on fait alors mieux qu'en tirant à pile ou face !).

Nous avons réalisé cet exercice pour trois modèles à facteurs : le modèle *MF1* représente le modèle à facteurs obtenus sur les 93 variables. Le modèle *MF2* est construit à partir des indicateurs habituellement considérés par les conjoncturistes, *i.e.* les soldes d'enquête et les variables réelles. La comparaison des deux modèles permet de mesurer l'apport de l'ajout des blocs nominal et international pour la prévision de court terme. Nous considérons aussi un troisième modèle, qui utilise toutes les variables sauf les variables réelles qui présentent les délais de publication les plus longs, et sont donc susceptibles de dégrader les prévisions les plus précoces. En effet, lorsque l'horizon de prévision est éloigné, l'indisponibilité d'un grand nombre de séries impose de tronquer davantage l'échantillon pour réaliser l'estimation de première étape (par ACP), puisqu'elle est faite sur un échantillon cylindré. Lorsqu'on ajoute les variables réelles, les facteurs et leur dynamique sont donc estimés avec plus de données, mais sur une période moins longue.

Pour évaluer la sensibilité des résultats au choix du nombre de facteurs et de retards, nous avons par ailleurs défini les spécifications de deux façons alternatives, sur la base des critères d'information ou sur la base de la qualité d'ajustement (critère RMSE) pour un nombre de facteurs dynamiques maximal de 5 et des retards s et p maximaux de 2 et 3. Ces critères ont été appliqués pour les trois ensembles de variables sur l'échantillon total.

Par ailleurs, comme nous l'avons indiqué dans la section 2.3, pour chacune des spécifications considérées, nous comparons les résultats obtenus en utilisant les trois méthodes de prévision suivantes :

- méthode 1 : utilisation des équations (11a) et (11b), avec utilisation du modèle VAR pour la prévision des valeurs futures des facteurs ;
- méthode 2 : utilisation des équations (10a)²⁵ et (10b), sans prolongement des facteurs sur le trimestre en cours ;
- méthode 3 : utilisation des équations (10a) et (10b) avec utilisation du modèle VAR pour prolonger les facteurs sur le trimestre en cours.

Par construction, les méthodes 1 et 3 coïncident pour la prévision du trimestre en cours et du trimestre précédent (c'est-à-dire pour un nombre de mois avant la publication allant de - 4 à 0 dans le schéma de la figure 3), puisque l'équation (10a) coïncide avec l'équation (11a)

²⁴ Il n'est pas possible de mener l'exercice en utilisant les valeurs des séries qui étaient réellement disponibles à la date où la prévision est calculée, puisque généralement, seules les valeurs révisées sont disponibles.

²⁵ Rappelons que l'équation (10a) est estimée pour deux valeurs de h : pour prévoir le trimestre en cours ou le trimestre précédent, on estime l'équation avec $h=0$; pour prévoir le trimestre suivant, on estime l'équation avec $h=1$.

lorsque $h = 0$. De même, les méthodes 2 et 3 coïncident à chaque fin de trimestre, puisqu'il n'y a pas lieu, dans ce cas, de prolonger les facteurs : ceci se produit lorsque le nombre de mois avant publication est égal à -5 ou à -2 . Enfin, les trois méthodes coïncident lorsque le trimestre à prévoir est écoulé (c'est-à-dire pour un nombre de mois avant publication allant de -2 à 0).

Comme dans la plupart des études analogues, nous comparons en outre les prévisions issues des modèles à facteurs avec celles tirées de deux spécifications simples de référence (*benchmark*) : une *marche aléatoire avec une constante* (la prévision du taux de croissance du PIB est alors obtenue comme la moyenne du processus) et un *processus autorégressif d'ordre 2* (pour cette spécification autorégressive, le choix de retenir deux retards a été fait au vu du corrélogramme, du corrélogramme partiel et des tests de diagnostic usuels sur l'ensemble de la période). Les prévisions à un horizon de prévision supérieur à un sont obtenues par itération de l'équation obtenue.

b) Résultats

Les résultats de l'évaluation des prévisions du premier trimestre 2000 au troisième trimestre 2009 sont présentés dans les tableaux 2. Pour chacun des modèles étudiés, on évalue les erreurs de prévision en calculant la racine carrée de l'erreur quadratique moyenne (RMFSE, tableau 2a), l'erreur absolue moyenne (MAFE, tableau 2b) et en présentant les résultats du test de signe (tableau 2c).

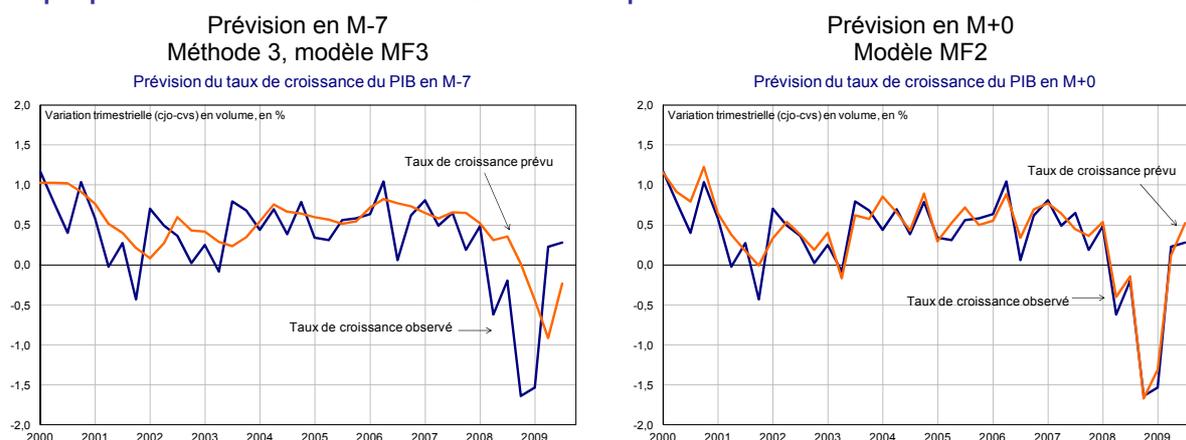
Les résultats sont présentés pour les trois méthodes de prévision répertoriées précédemment, en distinguant les horizons de prévision de façon à mettre en évidence ceux où ces méthodes sont effectivement différentes et ceux où elles coïncident par construction.

Pour chacune des méthodes de prévision proposées, on étudie les prévisions obtenues à partir des modèles à facteurs (MF1, MF2, MF3) associés aux trois sous-ensembles de données présentés ci-dessus. En outre, deux spécifications sont retenues pour chaque modèle à facteurs : la spécification issue de l'application des critères d'information, et la spécification issue du critère fondé sur le RMSE.

La comparaison au processus autorégressif ou à la marche aléatoire est favorable aux modèles à facteurs, en particulier aux trimestres T et $T+1$. En termes de RMSFE, le gain est de 22 % par rapport au processus autorégressif lors de la première prévision (ligne -7 du tableau) pour la meilleure spécification de modèle à facteurs (MF3 spécifié à partir des critères d'information) mais de façon générale, la qualité des prévisions n'est pas très bonne pendant le trimestre $T-1$. Le gain grandit à mesure que l'on se rapproche de la publication des comptes trimestriels pour se situer quelques jours avant la publication des comptes trimestriels dans une fourchette comprise entre 38 % pour le modèle à facteurs le moins performant et 60 % pour la meilleure spécification.

La qualité des prévisions tirées des modèles à facteurs s'améliore de façon quasi continue à mesure que l'on s'approche de la date de publication du PIB et donc que l'ensemble d'information sur le trimestre à prévoir grandit. L'erreur absolue moyenne, égale à 0,37 % dans la meilleure spécification sept mois avant la publication du PIB, n'est plus que de 0,15 % quelques jours avant. Ce résultat est illustré par les graphiques 4 qui représente les prévisions obtenues de manière récursive au premier mois du trimestre $T-1$ (graphique de gauche) et les prévisions quelques jours avant la parution des comptes trimestriels (graphique de droite) pour la spécification la plus performante à chacun de ces horizons.

Graphiques 4 : taux de croissance du PIB observé et prévu



Source : Insee, calculs des auteurs.

Concernant la méthode de prévision on constate que, pour les prévisions qui sont faites pendant les trois mois du trimestre $T-1$ et pour chacun des modèles étudiés, la méthode 3 (utilisant l'équation (10a) avec $h=1$ et le prolongement des facteurs sur l'ensemble du trimestre $T-1$ par utilisation du VAR) est la plus performante. Le résultat reste vrai pendant les deux premiers mois du trimestre T (l'équation (10a) est alors spécifiée avec $h=0$ donc les méthodes 1 et 3 coïncident). Notons que cette supériorité de la méthode 3 est valable aussi bien en termes de RMFSE qu'en termes de MAFE. Ce résultat montre que la spécification dynamique des facteurs, et donc la possibilité d'effectuer une prévision de ces facteurs, a un apport réel en termes de prévision. Cependant, parmi les méthodes de prévision que nous avons testées, la méthode 3, qui relie directement la valeur à prévoir y_T aux valeurs des facteurs pendant le trimestre $T-1$ est plus performante que la méthode 1 qui relie y_T aux valeurs contemporaines des facteurs et impose donc de faire une prévision des facteurs à un horizon plus long²⁶.

Concernant les critères de choix de modèles, les résultats sont systématiquement favorables aux critères d'information (sauf lors des deux mois précédant la publication du PIB, où les modèles sélectionnés par minimisation du RMSE donnent des résultats équivalents ou meilleurs). Or les critères d'information conduisent généralement à retenir des spécifications plus parcimonieuses²⁷. Nos résultats ne vont donc pas dans le même sens que ceux de Barhouni, Darné et Ferrara (2010), mais notre étude n'est pas directement comparable à la leur puisque nous prenons en compte l'évolution de l'ensemble d'information dans le trimestre et la période d'évaluation est différente. En particulier, pour les horizons de prévision les plus éloignés, les spécifications parcimonieuses donnent de meilleurs résultats.

Enfin, s'agissant de la sensibilité des résultats au choix des variables, on constate tout d'abord que la prise en compte des variables nominales (monétaires et financières) et des variables d'environnement international améliore la qualité des prévisions en début de période. En effet, le modèle MF3 (qui utilise toutes les variables sauf les variables réelles) est celui qui conduit aux meilleurs résultats, pendant toute la durée du trimestre $T-1$ et pendant les deux premiers mois du trimestre T . Ce résultat (qui se retrouve d'ailleurs dans le fait que les résultats du modèle MF1 sont meilleurs que ceux du modèle MF2 pour les mêmes horizons de prévision) semble d'autant plus intéressant que ces variables sont souvent écartées des outils de prévision de court terme. En revanche, la supériorité du

²⁶ Lorsqu'on se trouve au début du trimestre $T-1$, il faut prévoir les facteurs sur un horizon de 6 mois, et par conséquent la variance de l'erreur de prévision obtenue avec le modèle VAR devient importante.

²⁷ Nous retenons $q=5$, $s=p=1$ contre $q=5$, $s=2$, $p=1$ pour MF1, $q=4$, $s=p=1$ contre $q=5$, $s=2$, $p=3$ pour MF3 mais $q=4$, $s=2$, $p=1$ pour MF2 suivant les deux méthodes ; pour $h=1$, lorsque l'on utilise deux équations, le critère RMSFE conduit par ailleurs à retenir $q=5$, $s=p=2$ pour MF1, $q=5$, $s=2$ et $p=3$ pour MF2 et MF3.

modèle MF3 sur le modèle MF1 pour ces horizons semble indiquer que l'inclusion des variables réelles dégrade la qualité des prévisions les plus précoces fournies par les modèles à facteurs. Lorsqu'on se rapproche de la date de publication du PIB, le meilleur modèle est le modèle MF2, c'est-à-dire celui qui est estimé sur les variables d'enquête et les variables réelles.

On valide ainsi les pratiques de conjoncturistes consistant à écarter les indicateurs réels comme la production industrielle en début de trimestre (elle n'est souvent prise en compte dans les outils de prévision que lorsque les deux premiers mois du trimestre sont publiés, soit compte tenu des délais de publications de l'indicateur, seulement au début du premier mois du trimestre suivant). En résumé, ce travail d'analyse hors échantillon suggère de construire en amont les facteurs en utilisant toute la base sauf les données réelles au début de la campagne de prévision, puis en utilisant les données réelles et soldes d'enquêtes lorsque l'on se rapproche de la date de publication du PIB.

Conclusion

Dans cette étude, nous examinons les performances d'un outil de prévision du taux de croissance du PIB français sur des horizons courts, fondé sur les modèles à facteurs dynamiques. Ces modèles permettent de prendre en compte l'information apportée par un grand nombre de variables, celle-ci étant résumée sous la forme d'un petit nombre de facteurs. Dans leur forme dynamique, ils autorisent une dépendance temporelle des facteurs et une dépendance des variables observées à des valeurs contemporaines et retardées des facteurs. Les techniques d'estimation associées peuvent être adaptées en cas de valeurs manquantes des indicateurs, de sorte qu'il n'est pas nécessaire de recourir à des modèles auxiliaires.

L'évaluation hors échantillon et en pseudo temps réel de ces techniques montre généralement la supériorité des prévisions issues des modèles à facteurs sur celles issues des processus de référence habituels. Les prévisions restent néanmoins fragiles avant le début du trimestre, lorsque l'on dispose de peu d'information. Dans ce cadre, on montre également que l'utilisation de variables financières et internationales permet d'améliorer les prévisions au début des campagnes de prévisions.

Diverses voies d'amélioration de ces résultats pourraient être explorées. L'utilisation de différents jeux de variables initiales semble conduire à des performances différentes : il pourrait donc être intéressant d'utiliser en amont de la construction des facteurs les techniques de sélection des variables présentées par Boivin et Ng (2006) ou plus récemment par Bai et Ng (2008a). L'emploi de ces techniques par Caggiano et al. (2009) ou Schumacher (2010) met en effet en évidence un gain pour la prévision du PIB, et l'application faite par Charpin (2009), sur données françaises, de la méthode proposée par Bai et Ng (2008a) conduit à des résultats encourageants. En outre, l'introduction de non-linéarités dans la spécification a été encore assez peu explorée dans le cadre des modèles à facteurs, et pourrait aussi être une piste importante d'amélioration des résultats.

Tableau 2a : Évaluation des prévisions - RMFSE

**Prévisions faites au trimestre précédent
Méthode 1 : prévisions des facteurs avec le VAR**

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,58	0,59	0,56	0,58	0,60	0,60	0,51	0,51
-6 (T-1)	0,58	0,56	0,55	0,57	0,59	0,59	0,47	0,50
-5 (T-1)	0,58	0,56	0,50	0,53	0,55	0,55	0,41	0,41

Méthode 2 : deux équations de prévisions sans prolonger les facteurs sur T

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,58	0,59	0,77	1,06	0,78	0,75	0,56	0,69
-6 (T-1)	0,58	0,56	0,49	0,62	0,55	0,57	0,46	0,51
-5 (T-1)	0,58	0,56	0,41	0,43	0,54	0,51	0,39	0,41

Méthode 3 : deux équations de prévisions et prolongement des facteurs sur T

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,58	0,59	0,54	0,58	0,56	0,51	0,46	0,51
-6 (T-1)	0,58	0,56	0,52	0,55	0,54	0,54	0,44	0,50
-5 (T-1)	0,58	0,56	0,41	0,43	0,54	0,51	0,39	0,41

N.B. Les méthodes 2 et 3 coïncident en -5

Prévisions faites pendant le trimestre à prévoir, aux horizons où les méthodes 1 et 3 coïncident

Méthodes 1 et 3

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-4 (T)	0,58	0,56	0,47	0,51	0,50	0,50	0,38	0,38
-3 (T)	0,58	0,47	0,40	0,47	0,41	0,41	0,34	0,37

Méthode 2

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-4 (T)	0,58	0,56	0,68	0,80	0,73	0,73	0,41	0,47
-3 (T)	0,58	0,47	0,42	0,50	0,45	0,45	0,35	0,41

Prévisions faites pendant le trimestre à prévoir ou le trimestre suivant, aux horizons où les 3 méthodes coïncident

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-2 (T)	0,58	0,47	0,39	0,46	0,34	0,34	0,37	0,41
-1 (T+1)	0,58	0,47	0,24	0,24	0,19	0,19	0,29	0,23
0 (T+1)	0,58	0,47	0,22	0,21	0,19	0,19	0,29	0,23

MF1 : modèle à facteurs sur les 93 variables

MF2 : modèle à facteurs sur les soldes d'enquête et les variables réelles

MF3 : modèle à facteurs sur les soldes d'enquête, les variables nominales et internationales

Tableau 2b : Évaluation des prévisions - MAFE

**Prévisions faites au trimestre précédent
Méthode 1 : prévisions des facteurs avec le VAR**

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,40	0,40	0,39	0,40	0,44	0,44	0,37	0,37
-6 (T-1)	0,40	0,40	0,38	0,41	0,44	0,44	0,34	0,35
-5 (T-1)	0,40	0,40	0,36	0,38	0,40	0,40	0,31	0,32

Méthode 2 : deux équations de prévisions sans prolonger les facteurs sur T

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,40	0,40	0,56	0,74	0,57	0,58	0,39	0,50
-6 (T-1)	0,40	0,40	0,35	0,44	0,40	0,45	0,31	0,35
-5 (T-1)	0,40	0,40	0,31	0,33	0,39	0,37	0,29	0,32

Méthode 3 : deux équations de prévisions et prolongement des facteurs sur T

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,40	0,40	0,39	0,45	0,41	0,39	0,32	0,36
-6 (T-1)	0,40	0,40	0,36	0,40	0,40	0,40	0,30	0,34
-5 (T-1)	0,40	0,40	0,31	0,33	0,39	0,37	0,29	0,32

N.B. Les méthodes 2 et 3 coïncident en -5.

Prévisions faites pendant le trimestre à prévoir, aux horizons où les méthodes 1 et 3 coïncident

Méthodes 1 et 3

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-4 (T)	0,40	0,40	0,34	0,38	0,37	0,37	0,29	0,29
-3 (T)	0,40	0,35	0,30	0,36	0,30	0,30	0,26	0,28

Méthode 2

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-4 (T)	0,40	0,40	0,45	0,57	0,52	0,52	0,33	0,33
-3 (T)	0,40	0,35	0,32	0,36	0,33	0,33	0,28	0,31

Prévisions faites pendant le trimestre à prévoir ou le trimestre suivant, aux horizons où les 3 méthodes coïncident

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-2 (T)	0,40	0,35	0,29	0,33	0,28	0,28	0,28	0,30
-1 (T+1)	0,40	0,35	0,18	0,19	0,15	0,15	0,22	0,18
0 (T+1)	0,40	0,35	0,17	0,17	0,15	0,15	0,22	0,19

MF1 : modèle à facteurs sur les 93 variables

MF2 : modèle à facteurs sur les soldes d'enquête et les variables réelles

MF3 : modèle à facteurs sur les soldes d'enquête, les variables nominales et internationales

Tableau 2c : Évaluation des prévisions - Tests de signe

Méthode 1 : prévisions des facteurs avec le VAR

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,50 (0,50)	0,50 (0,50)	0,58 (0,16)	0,55 (0,26)	0,47 (0,63)	0,47 (0,63)	0,58 (0,16)	0,63 (0,05)
-6 (T-1)	0,50 (0,50)	0,58 (0,16)	0,66 (0,02)	0,61 (0,09)	0,50 (0,50)	0,50 (0,50)	0,66 (0,02)	0,63 (0,05)
-5 (T-1)	0,50 (0,50)	0,58 (0,16)	0,55 (0,26)	0,61 (0,09)	0,58 (0,16)	0,58 (0,16)	0,68 (0,01)	0,63 (0,05)

Méthode 2 : deux équations de prévisions sans prolonger les facteurs sur T

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,50 (0,50)	0,50 (0,50)	0,66 (0,02)	0,53 (0,37)	0,68 (0,01)	0,53 (0,37)	0,79 (0,00)	0,66 (0,02)
-6 (T-1)	0,50 (0,50)	0,58 (0,16)	0,68 (0,01)	0,61 (0,09)	0,60 (0,09)	0,47 (0,63)	0,76 (0,00)	0,68 (0,01)
-5 (T-1)	0,50 (0,50)	0,58 (0,16)	0,63 (0,05)	0,61 (0,09)	0,50 (0,50)	0,47 (0,63)	0,71 (0,00)	0,66 (0,02)

Méthode 3 : deux équations de prévisions et prolongement des facteurs sur T

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-7 (T-1)	0,50 (0,50)	0,50 (0,50)	0,58 (0,16)	0,53 (0,37)	0,58 (0,16)	0,53 (0,37)	0,71 (0,00)	0,53 (0,37)
-6 (T-1)	0,50 (0,50)	0,58 (0,16)	0,68 (0,01)	0,55 (0,26)	0,53 (0,37)	0,37 (0,95)	0,71 (0,00)	0,58 (0,16)
-5 (T-1)	0,50 (0,50)	0,58 (0,16)	0,63 (0,05)	0,61 (0,09)	0,50 (0,50)	0,47 (0,63)	0,71 (0,00)	0,66 (0,02)

N.B. Les méthodes 2 et 3 coïncident en -5

Prévisions faites pendant le trimestre à prévoir, aux horizons où les méthodes 1 et 3 coïncident

Méthodes 1 et 3

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-4 (T)	0,50 (0,50)	0,58 (0,16)	0,55 (0,26)	0,50 (0,50)	0,47 (0,63)	0,47 (0,63)	0,61 (0,09)	0,63 (0,05)
-3 (T)	0,50 (0,50)	0,32 (0,99)	0,58 (0,16)	0,66 (0,02)	0,53 (0,37)	0,53 (0,37)	0,58 (0,16)	0,61 (0,09)

Méthode 2

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-4 (T)	0,50 (0,50)	0,58 (0,16)	0,47 (0,63)	0,47 (0,63)	0,53 (0,37)	0,53 (0,37)	0,53 (0,37)	0,61 (0,09)
-3 (T)	0,50 (0,50)	0,32 (0,99)	0,58 (0,16)	0,71 (0,00)	0,61 (0,09)	0,61 (0,09)	0,61 (0,09)	0,50 (0,50)

Prévisions faites pendant le trimestre à prévoir ou le trimestre suivant, aux horizons où les 3 méthodes coïncident

Mois avant publication du PIB	Marche aléatoire	AR(2)	MF1 crit d'info	MF1 RMSE	MF2 crit d'info	MF2 RMSE	MF3 crit info	MF3 RMSE
-2 (T)	0,50 (0,50)	0,32 (0,99)	0,53 (0,37)	0,61 (0,09)	0,63 (0,05)	0,63 (0,05)	0,58 (0,16)	0,58 (0,16)
-1 (T+1)	0,50 (0,50)	0,32 (0,99)	0,79 (0,00)	0,71 (0,00)	0,71 (0,00)	0,71 (0,00)	0,66 (0,02)	0,74 (0,00)
0 (T+1)	0,50 (0,50)	0,33 (0,99)	0,82 (0,00)	0,79 (0,00)	0,82 (0,00)	0,82 (0,00)	0,63 (0,05)	0,71 (0,00)

MF1 : modèle à facteurs sur les 93 variables

MF2 : modèle à facteurs sur les soldes d'enquête et les variables réelles

MF3 : modèle à facteurs sur les soldes d'enquête, les variables nominales et internationales

BIBLIOGRAPHIE

- Altissimo F., Bassanetti A., Cristadoro R., Forni M., Hallin M., Lippi M., Reichlin L., Veronese G. (2001).** “A real time coincident indicator for the euro area business cycle”, *CEPR Discussion Paper Series* No. 3108.
- Altissimo F., Cristadoro R., Forni M., Lippi M., Veronese G. (2007).** “New Eurocoin: Tracking economic growth in real time”, *Banca d'Italia* No. 631.
- Angelini E., Camba-Méndez G., Giannone D., Rünstler G., Reichlin L. (2008).** “Short-term forecasts of Euro Area GDP growth”, Working Paper Series 949, ECB.
- Bai J., Ng S. (2002).** “Determining the number of factors in approximate factor models”, *Econometrica*, 70, pp. 191-221.
- Bai J., Ng S. (2007).** “Determining the number of primitive shocks in factor models”, *Journal on Business and Economic Statistics*, 25, pp. 52-60.
- Bai, J., Ng, S. (2008a).** Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146, pp. 304–317.
- Bai, J., Ng, S. (2008b).** Large Dimensional Factor Analysis. Foundations and Trends in Econometrics, vol 3 n°2, pp. 89–163.
- Bañbura M., Rünstler G. (2011).** “A look into the factor model black box - publication lags and the role of hard and soft data in forecasting GDP”, *International Journal of forecasting* 27(2), pp. 333-346.
- Barhoumi K., Brunhes-Lesage V., Darné O., Ferrara L., Pluyaud B., Rouvreau B. (2008).** “Monthly forecasting of French GDP : A revised version of the OPTIM model”, *Note d'Etudes et de Recherche de la Banque de France*, No. 222, 2008.
- Barhoumi K., Darné O., Ferrara L. (2010).** “Are disaggregate data useful for factor analysis in forecasting French GDP?”, *Journal of Forecasting*, 29(1-2), pp. 132–144.
- Barhoumi K., Rünstler G., Cristadoro R., Reijer A. D., Jakaitiene A., Jelonek P., Rua A., Ruth K., Benk S., Nieuwenhuyze C. V. (2009).** “Short-term forecasting of GDP using large monthly data sets: a pseudo real-time forecast evaluation exercise”, *Journal of forecasting*, 28(7), pp. 595-611.
- Bessec M. (2010).** “Étalonnages du taux de croissance du PIB français sur la base d'enquête de conjoncture”, *Économie et Prévision*, 193, 77-99.
- Boivin J., Ng S. (2006).** “Are more data always better for factor analysis”, *Journal of Econometrics*, 132, pp. 169-194.
- Caggiano G., Kapetanios G., Labhard V. (2009).** “Are more data always better for factor analysis? Results for the Euro Area, the six largest euro area countries and the UK”, *Working Paper Series* 1051, European Central Bank.
- Charpin F.**, Estimation précoce de la croissance. De la régression LARS au modèle à facteurs, *Revue de l'OFCE* 2009/1, N° 108, p. 31-48.
- Darné O., Brunhes-Lesage V. (2007).** “L'indicateur synthétique mensuel d'activité (ISMA) : une révision”, *Note d'Etudes et de Recherche de la Banque de France*, NER - E#171.
- Diron M. (2006).** “Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data”, *Working Paper Series* 622, European Central Bank.
- Doz C., Giannone D., Reichlin L. (2006).** “A quasi maximum likelihood approach for large approximate dynamic factor models”, *CEPR Discussion Paper*, No. 5724. À paraître dans *Review of Economics and Statistics*.
- Doz C., Giannone D., Reichlin L. (2007).** “A two-step estimator for large approximate dynamic factor models based on Kalman filtering”, *CEPR Discussion Paper*, No. 6043. À paraître dans *Journal of Econometrics*.
- Dubois E., Michaux E. (2006).** “Étalonnage à l'aide d'enquêtes de conjoncture : De nouveaux résultats”, *Économie et Prévision*, 172, pp. 11-28.

- Erkel-Rousse H., Minodier C. (2009).** “Do Business Tendency Surveys in Industry and Services Help in Forecasting GDP Growth? A Real-Time Analysis on French Data”, Direction des Études et Synthèses Économiques, Insee.
- Forni M., Hallin M., Lippi M., Reichlin L. (2004).** “The generalized dynamic factor model: consistency and rates”, *Journal of Econometrics*, 119, pp. 231-255.
- Forni M., Hallin M., Lippi M., Reichlin L. (2005).** “The generalized dynamic factor model: one-sided estimation and forecasting”, *Journal of the American Statistical Association*, 100, pp. 830-840.
- Giannone D., Reichlin L., Small D. (2008).** “Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases”, *Journal of Monetary Economics*, Vol. 55, No. 4, May 2008, pp. 665-676.
- Grasmann P., Keereman F. (2001).** “An indicator-based short term forecast for quarterly GDP in euro area”, *Economic Paper* No. 154. Commission européenne.
- Grenouilleau D. (2004).** “A sorted leading indicator dynamic (SLID) factor model for short-run euro-area GDP forecasting”, *Economic Paper* No. 219. Commission européenne.
- Grenouilleau D. (2006).** “The Stacked Leading Indicators Dynamic Factor Model: A Sensitivity Analysis of Forecast Accuracy using Bootstrapping”, *Economic Paper* No. 249. Commission Européenne.
- Hendry D.F., Krolzig H-M. (1999).** “Improving on data mining reconsidered by K.D”, Hoover and S.J. Perez. *Econometric Journal*, 2, pp. 41-58.
- Hendry D.F., Krolzig H-M. (2001).** “Computer automation of general-to-specific model selection procedures”, *Journal of Economic Dynamic and Control*, 25(6-7), pp. 831-866.
- Hendry D.F., Krolzig H-M. (2005).** “The properties of automatic GETS modelling”, *The Economic Journal*, 115 (March), C32-C61.
- Hoover K.D., Perez S.J. (1999).** “Data mining reconsidered: a general to specification approach to specification search”, *Econometric Journal*, 2, pp. 167-191.
- Irac D., Sédillot F. (2002).** “Short run assessment of French economic activity using OPTIM”, Note d’Études et de Recherche de la Banque de France, N°088.
- Minodier C. (2010).** “Avantages comparés des séries des premières valeurs publiées et des séries des valeurs révisées. Un exercice de prévision en temps réel de la croissance trimestrielle du PIB en France”, Direction des Études et Synthèses Économiques, Insee.
- Pain N., Mouragane A., Sédillot F., Le Fouler L. (2005).** “The new international trade model”, OECD Economic Department *Working Paper*, No.440.
- Rünstler G., Sédillot F. (2003).** “Short term estimates of Euro Area real GDP by means of monthly data”, *Working Paper Series* 276, European Central Bank.
- Schumacher C. (2007).** “Forecasting German GDP using alternative factor models based on large datasets”, *Journal of Forecasting*, 26(4), pp. 271-302.
- Schumacher C. (2010).** “Factor forecasting using international targeted predictors: The case of German GDP”, *Economic Letters*, 2010, 107, pp. 95-98.
- Sédillot F. et Pain N. (2005).** “Indicator models of real GDP growth in the major OECD economies”, *OECD Economic Studies*, 40, pp. 167-217.
- Stock J., Watson M. (1999).** “Forecasting inflation”, *Journal of Monetary Economics*, 44, pp. 293-335.
- Stock J., Watson M. (2002a).** “Forecasting Using Principal Components from a Large Number of Predictors”, *Journal of the American Statistical Association*, 97, pp. 1167-1179.
- Stock J., Watson M. (2002b).** “Macroeconomic forecasting using diffusion indexes”, *Journal of Business and Economic Statistics*, 20, pp. 147-162.
- Stock J., Watson M. (2010).** “Dynamic Factor Models”, Prepared for the Oxford Handbook of Economic Forecasting, M.P. Clements et D.F. Hendry (eds), *Oxford University Press*.

Tableau 3 : Les différentes approches de prévision du PIB français et européen

Organisme	Pays	Horizon	Modèle	Variables explicatives	Réduction de l'ensemble d'information	Références
DG Trésor	FR	h=-1,0,1	Maquette macro-sectorielle « le Garde Fou » alimentée par des étalonnages classiques	- Variables d'enquête : industrie, services, commerce de détail et de gros, bâtiment de l'Insee, commerce de détail de la Banque de France,	GETS, critères d'information ou choix a priori suivant la méthodologie des comptes trimestriels	
				- Variables réelles : IPI, conso manuf., immatriculations, mises en chantier, permis de construire, conso d'électricité, livraison de combustibles et carburants, statistiques douanières du commerce extérieur, chiffres d'affaires du commerce, température, données CNAM sur les dépenses de santé, ...		
				- Variable financière : taux de change €/§		
				Variables d'enquête de l'Insee : industrie, services et bâtiment		
Insee	FR	h=-1,0,1 h=0	Étalonnage du PIB	Indice PMI composite	GETS par blocs	Bessec (2010)
				Variables d'enquêtes	Choix a priori	
				Variables réelles	GETS	Erkel-Rousse et Midodier (2009)
				Variables d'enquête de l'Insee (industrie et services)	GETS	
Banque de France	FR	h=0,1	Maquette macro-sectorielle « OPTIM »	- Variables d'enquête de l'Insee (industrie, services, construction, commerce de détail et ménages), la Banque de France (industrie, services et commerce de détail) et de la Commission Européenne	GETS	Irac et Sédillot (2002) Barhoumi et al. (2008)
				- Variables réelles : IPI, conso manuf., prix, conso d'électricité, mises en chantier, statistiques douanières du commerce extérieur		
				- Variable financière : taux de change €/§		
				Soldes d'enquêtes de la Banque de France dans l'industrie et les principaux secteurs du champ manufacturier et les services		
Banque de France	FR	h=0,1	Étalonnage du PIB « ISMA »	- Variables d'enquête : industrie, services, commerce détail, ménages	ACP puis GETS	Darné et Brunhes-Lesage (2007)
				- Variables réelles : IPI, conso manuf., immatriculations de véhicules neufs, ventes de véhicules industriels, mises en chantier, importations et exportations,		
				- Variables financières : indice boursier français, taux d'intérêt court et long et taux d'intérêt des prêts immobiliers et prix du pétrole et prix à la consommation		
OCDE	FR, BE, DE, IT, NL, PT, LT, HU, PL et ZE	h=-1,0,1	Modèle à facteurs sur le PIB	Variables d'enquêtes, réelles et financières, prix	- Modèle à facteurs dynamiques estimé suivant 4 méthodes alternatives (SW, DGR, FHLR) - Choix du nombre de facteurs : Bai et Ng (2002, 2007) versus choix fixe a priori	Barhoumi, Darné, Ferrara (2010)
OCDE	FR, ZE, US, DE, IT, UK, JP	h=0,1	Étalonnage du PIB	- Variables d'enquête : industrie et ménages (enquêtes Insee)	- Modèle à facteurs dynamiques estimé suivant 3 méthodes alternatives (SW, DGR, FHLR) - Choix du nombre de facteurs : critère RMSE	Barhoumi et al. (2009)
				- Variables réelles : IPI, IPI construction, consommation de biens manufacturés, ventes de détail, immatriculations de véhicules neufs, exports et imports mensuels, taux de chômage		
				- Variables financières : TI court, long et taux de change réel		
				Pré-sélection d'indicateurs sur la base du R ² de modèle bivarié		
				Sélection de la meilleure combinaison d'indicateurs avec le critère de Schwarz		Sédillot et Pain (2005)

Commission Européenne	ZE	h=0,1,2	Modèle à facteurs sur le PIB « SLID model »	<ul style="list-style-type: none"> - Variables d'enquête : industrie, services, construction, commerce de détail et ménages (source : Commission Européenne), PMI - Variables réelles : IPI, ventes de détail et du commerce de gros, données mensuelles du commerce extérieur en valeur, mises en chantier, immatriculations, variables sur le marché du travail et créations d'entreprises, défaillances d'entreprises - Variables financières : prix, agrégats monétaires, indices boursiers, taux de change nominaux, pente des taux, indice de volatilité implicite du CAC40, DAX20, Eurostoxx50 et AEX <p>Les variables sont prises à un niveau sectoriel et par pays</p> <ul style="list-style-type: none"> - Variables d'enquête : industrie, commerce de détail, ménages et services - Variables réelles : IPI, IPI construction, ventes de détail et immatriculations <p>Des indices composites : l'indice TSEN de la Commission, l'EuroCOIN du CEPR et l'indicateur composite avancé de l'OCDE pour la zone euro</p> <p>+ Variables financières : <i>indice boursier, spread, taux de change effectif de la zone euro dans Diron</i></p>	Grenouilleau (2004, 2006)
	ZE	h=-1,0,1	Combinaison d'étalement sur le PIB	<ul style="list-style-type: none"> - Variables d'enquête : industrie, ménages, construction, commerce de détail et 2 soldes US sur les anticipations des industriels et ménages - Variables réelles : IPI et IPI sectoriels, exportations et importations intra et hors ZE, ventes de détail, immatriculations, emploi, indicateurs US (IPI, ventes de détail, chômage, emploi) - Variables financières : taux de change effectif nominal, taux de change effectif réel, taux de change bilatéraux, prix des matières premières, prix de l'or, prix du pétrole, indice boursier européen et US, taux d'intérêt européen et US 	Diron (2006), Rünstler et Sédillot (2003)
	ZE	h=-1,0,1	Modèle à facteurs sur le PIB	<ul style="list-style-type: none"> - Variables d'enquête : enquête de la commission européenne (business, construction, commerce de détail et ménages) - Variables réelles : IPI et IPI sectoriels, emploi et chômage, exportations et importations, ventes de détail et immatriculations - Variables financières : taux de change, taux d'intérêt et indices boursiers, agrégats monétaires, prix des matières premières - Variables internationales : indicateurs macro-économiques 	Barbura et Rünstler (2011)
	ZE	h=-1,0,1	Modèle à facteurs sur le PIB	<ul style="list-style-type: none"> - Variables d'enquête : enquête de la commission européenne (business, construction, commerce de détail et ménages) - Variables réelles : IPI et IPI sectoriels, emploi et chômage, exportations et importations, ventes de détail et immatriculations - Variables financières : taux de change, taux d'intérêt et indices boursiers, agrégats monétaires, prix des matières premières - Variables internationales : indicateurs macro-économiques 	Angelini, Barbura et Rünstler (2008)
BCE	FR et ZÉ, BE, DE, UK, IT, NL, ES	h=1, ..., 12	Modèle à facteurs sur le PIB	<ul style="list-style-type: none"> - Variables d'enquête : industrie, ménages, construction, commerce de détail et 2 soldes US sur les anticipations des industriels et ménages - Variables réelles : IPI et IPI sectoriels, exportations et importations intra et hors ZE, ventes de détail, immatriculations, emploi, indicateurs US (IPI, ventes de détail, chômage, emploi) - Variables financières : taux de change effectif nominal, taux de change effectif réel, taux de change bilatéraux, prix des matières premières, prix de l'or, prix du pétrole, indice boursier européen et US, taux d'intérêt européen et US 	Caggiano, Kapetanios et Labhard (2009)

Tableau 4 : Les données

Type	Secteur	Série	Fréquence	Début de la série	délais	Source
Enquête	Industrie	Production passée	M	avr-76	M+0	Insee
Enquête	Industrie	Stocks	M	avr-76	M+0	Insee
Enquête	Industrie	Carnets commandes globaux	M	avr-76	M+0	Insee
Enquête	Industrie	Carnets de commandes étrangers	M	avr-76	M+0	Insee
Enquête	Industrie	Perspectives personnelles de production	M	avr-76	M+0	Insee
Enquête	Industrie	Perspectives générales de production	M	avr-76	M+0	Insee
Enquête	Services	Activité passée	Q/M	88T1 à 00T2 puis juin 2000	M+0	Insee
Enquête	Services	Activité prévue	Q/M	88T1 à 00T2 puis juin 2000	M+0	Insee
Enquête	Services	Demande prévue	Q/M	88T1 à 00T2 puis juin 2000	M+0	Insee
Enquête	Bâtiment	Activité passée	Q/M	75T1 à 93T3 puis sept 1993	M+0	Insee
Enquête	Bâtiment	Activité prévue	Q/M	75T1 à 93T3 puis sept 1993	M+0	Insee
Enquête	Bâtiment	Effectifs passés	Q/M	75T2 à 93T3 puis sept 1993	M+0	Insee
Enquête	Bâtiment	Carnets de commandes	Q/M	75T1 à 93T3 puis sept 1993	M+0	Insee
Enquête	Bâtiment	TUC	Q/M	75T2 à 93T3 puis sept 1993	M+0	Insee
Enquête	Commerce de détail	Perspectives générales d'activité	M	janv-91	M+0	Insee
Enquête	Commerce de détail	Ventes passées	M	janv-91	M+0	Insee
Enquête	Commerce de détail	Commandes	M	janv-91	M+0	Insee
Enquête	Commerce de détail	Emploi prévu	M	janv-91	M+0	Insee
Enquête	Ménages	Situation financière perso – évolution passée	M	janv-70	M+0	Insee
Enquête	Ménages	Situation financière perso – évolution prévue	M	janv-70	M+0	Insee
Enquête	Ménages	Niveau de vie en France – évolution passée	M	janv-70	M+0	Insee
Enquête	Ménages	Niveau de vie en France - évolution attendue	M	janv-70	M+0	Insee
Enquête	Ménages	Opportunité de faire des achats importants	M	janv-70	M+0	Insee
Enquête	Industrie manuf.	Demande - évolution passée	Q	1976T2	Q+0	Insee
Enquête	Industrie manuf.	Demande - évolution prévue	Q	1976T2	Q+0	Insee
Enquête	Services	Résultats d'exploitation passés	Q	1988T1	Q+0	Insee
Enquête	Services	Résultats d'exploitation prévus	Q	1988T1	Q+0	Insee
Réel	Ménages	Immatriculations de véhicules particuliers neufs	M	janv-85	M+1	Insee
Réel	Ménages	Consommation de produits manufacturés	M	janv-85	M+1	Insee
Réel	Ménages	Conso de biens durables	M	janv-85	M+1	Insee
Réel	Ménages	Conso d'automobile	M	janv-85	M+1	Insee
Réel	Ménages	Conso de biens d'équipement logement	M	janv-85	M+1	Insee
Réel	Ménages	Conso de textile	M	janv-85	M+1	Insee
Réel	Ménages	Conso d'autres produits manufacturés	M	janv-85	M+1	Insee
Réel	Ménages	Vente de détail hors automobile	M	janv-95	M+2	Eurostat
Réel	Ménages	Taux de chômage	M	janv-83	M+2	Eurostat
Réel	Ménages	Taux de chômage des moins de 25 ans	M	janv-83	M+2	Eurostat
Réel	Ménages	Emplois vacants	M	janv-89	M+2	OCDE
Réel	Bâtiment	Mises en chantier total	M	janv-94	M+1	MEEDDM
Réel	Bâtiment	Mises en chantier collectif	M	janv-94	M+1	MEEDDM
Réel	Bâtiment	Mises en chantier individuel	M	janv-94	M+1	MEEDDM
Réel	Bâtiment	Mises en chantier résidentiel	M	janv-94	M+1	MEEDDM
Réel	Bâtiment	Permis construire total	M	janv-94	M+1	MEEDDM
Réel	Bâtiment	Permis construire collectif	M	janv-94	M+1	MEEDDM
Réel	Bâtiment	Permis construire individuel	M	janv-94	M+1	MEEDDM
Réel	Industrie	IPI – industrie totale (BE)	M	janv-90	M+2	Insee
Réel	Industrie	IPI - industrie manuf (CZ)	M	janv-90	M+2	Insee
Réel	Industrie	IPI - industries agricoles et alimentaires (C1)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – cokéfaction et raffinage (C2)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – équipements électriques (C3)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – automobile (CL1)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – transports hors auto (CL2)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – autres produits industriels (C5)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – industries extractives, énergie et eau (DE)	M	janv-90	M+2	Insee
Réel	Industrie	IPI – construction (F)	M	janv-90	M+2	Insee
Nominal	Bourse	CAC 40	M	juil-87	M+0	Nyse Euronext Paris
Nominal	Bourse	SP500	M	janv-80	M+0	Standard and Poors
Nominal	Bourse	FTSE	M	janv-84	M+0	FTSE
Nominal	Bourse	DAX	M	sept-59	M+0	Frankfurt SE
Nominal	Bourse	Eurostoxx50	M	déc-86	M+0	Financial Time
Nominal	Bourse	Nikkei	M	janv-81	M+0	OSE
Nominal	Bourse	Price Earning Ratio (US)	M	janv-54	M+1	Standard and Poors
Nominal	Bourse	Indice de volatilité Vix	M	janv-90	M+0	CBOE
Nominal	Monnaie	M1	M	janv-80	M+2	BdF
Nominal	Monnaie	M2	M	janv-80	M+2	BdF
Nominal	Monnaie	M3	M	janv-80	M+2	BdF
Nominal	Monnaie	Prêts (en valeur)	M	janv-80	M+2	BdF
Nominal	Taux d'intérêt	Taux d'intérêt immobilier	M	févr-78	M+1	BdF
Nominal	Taux d'intérêt	Taux d'intérêt à 3 mois	M	janv-70	M+1	FMI
Nominal	Taux d'intérêt	Taux d'intérêt à 1 an	M	janv-60	M+1	BdF
Nominal	Taux d'intérêt	Taux d'intérêt à 10 ans	M	janv-89	M+0	Daily press
Nominal	Taux d'intérêt	Pente des taux France	M	janv-90	M+0	Global insight
Nominal	Taux d'intérêt	Pente des taux US	M	janv-90	M+0	Global insight
Nominal	Prix	Or	M	janv-79	M+0	LBMA
Nominal	Prix	Pétrole	M	janv-80	M+1	FMI
Nominal	Prix	Matières premières	M	janv-80	M+1	FMI
Nominal	Prix	IPC	M	janv-90	M+1	Insee
International	Taux de change	Euro/dollar	M	janv-78	M+0	BCE
International	Taux de change	Euro/uk	M	déc-77	M+0	BCE
International	Taux de change	Euro/yen	M	déc-78	M+0	BCE
International	Taux de change	Euro/yuan	M	déc-78	M+0	BCE
International	Taux de change	Taux effectif de l'euro	M	janv-79	M+2	FMI
International	Taux de change	Taux effectif réel de l'euro	M	déc-79	M+2	FMI
International	Allemagne	Situation actuelle (IFO)	M	janv-91	M+0	IFO
International	Allemagne	Perspectives d'activité à 6 mois (IFO)	M	janv-91	M+0	IFO
International	Allemagne	Situation actuelle (ZEW)	M	déc-91	M+0	ZEW
International	Allemagne	Perspectives économiques (ZEW)	M	déc-91	M+0	ZEW
International	Allemagne	IPI manufacturé	M	janv-91	M+2	Destatis
International	US	Ventes de détail (en valeur)	M	janv-92	M+1	Census Bureau
International	US	IPI manufacturé	M	janv-72	M+1	Fed
International	US	Emploi	M	janv-48	M+1	BLS
International	US	Taux de chômage	M	janv-48	M+1	BLS
International	US	PMI manufacturé	M	janv-48	M+1	ISM

Tableau 5 : Résultats des tests de racine unitaire

Type	Secteur	Série	ADF	PP	KPSS	Transformation
			Conclusion	Conclusion	Conclusion	
Enquête	Industrie	Production passée	I(0)	I(0)	I(0)	-
Enquête	Industrie	Stocks	I(0)	I(0)	I(0)	-
Enquête	Industrie	Carnets commandes globaux	I(0)	I(0)	I(0)	-
Enquête	Industrie	Carnets de commandes étrangers	I(0)	I(0)	I(0)	-
Enquête	Industrie	Perspectives personnelles de production	I(0)	I(0)	I(0)	-
Enquête	Industrie	Perspectives générales de production	I(0)	I(0)	I(0)	-
Enquête	Services	Activité passée	I(0)	I(0)	I(0)	-
Enquête	Services	Activité prévue	I(0)	I(0)	I(0)	-
Enquête	Services	Demande prévue	I(0)	I(0)	I(0)	-
Enquête	Bâtiment	Activité passée	I(0)	I(1)	I(0)	-
Enquête	Bâtiment	Activité prévue	I(0)	I(1)	I(1)	-
Enquête	Bâtiment	Effectifs passés	I(1)	I(1)	I(0)	-
Enquête	Bâtiment	Carnets de commandes	I(1)	I(1)	I(1)	1-L
Enquête	Bâtiment	TUC	I(1)	I(1)	I(1)	1-L
Enquête	Commerce de détail	Perspectives générales d'activité	I(1)	I(1)	I(0)	-
Enquête	Commerce de détail	Ventes passées	I(0)	I(0)	I(0)	-
Enquête	Commerce de détail	Commandes	I(0)	I(0)	I(1)	-
Enquête	Commerce de détail	Emploi prévu	I(0)	I(0)	I(0)	-
Enquête	Ménages	Situation financière perso – évolution passée	I(0)	I(0)	I(0)	-
Enquête	Ménages	Situation financière perso – évolution prévue	I(0)	I(0)	I(0)	-
Enquête	Ménages	Niveau de vie en France – évolution passée	I(1)	I(1)	I(0)	-
Enquête	Ménages	Niveau de vie en France - évolution attendue	I(0)	I(0)	I(0)	-
Enquête	Ménages	Opportunité de faire des achats importants	I(1)	I(1)	I(1)	1-L
Enquête	Industrie	Demande - évolution passée	I(0)	I(0)	I(0)	-
Enquête	Industrie	Demande - évolution prévue	I(0)	I(0)	I(0)	-
Enquête	Services	Résultats d'exploitation passés	I(0)	I(0)	I(0)	-
Enquête	Services	Résultats d'exploitation prévus	I(0)	I(0)	I(0)	-
Réel	Ménages	Immatriculations de véhicules particuliers neufs	I(0)	I(0)	I(0)	(1-L)log
Réel	Ménages	Consommation de produits manufacturés	I(1)	I(1)	I(1)	(1-L)log
Réel	Ménages	Conso de biens durables	I(1)	I(1)	I(1)	(1-L)log
Réel	Ménages	Conso d'automobile	I(0)	I(0)	I(1)	(1-L)log
Réel	Ménages	Conso de biens d'équipement logement	I(1)	I(1)	I(1)	(1-L)log
Réel	Ménages	Conso de textile	I(1)	I(0)	I(1)	(1-L)log
Réel	Ménages	Conso d'autres produits manufacturés	I(1)	I(1)	I(1)	(1-L)log
Réel	Ménages	Vente de détail hors automobile	I(1)	I(1)	I(1)	(1-L)log
Réel	Ménages	Taux de chômage	I(1)	I(1)	I(0)	1-L
Réel	Ménages	Taux de chômage des moins de 25 ans	I(1)	I(1)	I(0)	1-L
Réel	Ménages	Emplois vacants	I(1)	I(1)	I(1)	1-L
Réel	Bâtiment	Mises en chantier total	I(0)	I(0)	I(1)	(1-L)log
Réel	Bâtiment	Mises en chantier collectif	I(1)	I(0)	I(1)	(1-L)log
Réel	Bâtiment	Mises en chantier individuel	I(1)	I(0)	I(1)	(1-L)log
Réel	Bâtiment	Mises en chantier résidentiel	I(0)	I(0)	I(1)	(1-L)log
Réel	Bâtiment	Permis construire total	I(1)	I(0)	I(1)	(1-L)log
Réel	Bâtiment	Permis construire collectif	I(1)	I(0)	I(1)	(1-L)log
Réel	Bâtiment	Permis construire individuel	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	BE IPI	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	CZ IPI	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	C1 IPI	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	C2 IPI	I(0)	I(0)	I(1)	(1-L)log
Réel	Industrie	C3 IPI	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	CL1 IPI	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	CL2 IPI	I(1)	I(0)	I(1)	(1-L)log
Réel	Industrie	C5 IPI	I(1)	I(1)	I(1)	(1-L)log
Réel	Industrie	DE IPI	I(0)	I(0)	I(1)	(1-L)log
Réel	Industrie	F IPI	I(1)	I(0)	I(1)	(1-L)log
Nominal	Bourse	CAC 40	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	SP500	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	FTSE	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	DAX	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	Eurostoxx50	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	Nikkei	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	Price Earning Ratio (US)	I(1)	I(1)	I(1)	(1-L)log
Nominal	Bourse	Indice de volatilité Vix	I(0)	I(0)	I(0)	log
Nominal	Monnaie	M1	I(1)	I(1)	I(1)	(1-L)log
Nominal	Monnaie	M2	I(1)	I(1)	I(1)	(1-L)log
Nominal	Monnaie	M3	I(1)	I(1)	I(1)	(1-L)log
Nominal	Monnaie	Prêts (en valeur)	I(0)	I(0)	I(2)	(1-L)log
Nominal	Taux d'intérêt	Taux d'intérêt immobilier	I(1)	I(1)	I(1)	(1-L)
Nominal	Taux d'intérêt	Taux d'intérêt à 3 mois	I(1)	I(1)	I(1)	(1-L)
Nominal	Taux d'intérêt	Taux d'intérêt à 1 an	I(1)	I(1)	I(1)	(1-L)
Nominal	Taux d'intérêt	Taux d'intérêt à 10 ans	I(1)	I(1)	I(1)	(1-L)
Nominal	Taux d'intérêt	Pente des taux France	I(0)	I(0)	I(0)	-
Nominal	Taux d'intérêt	Pente des taux US	I(1)	I(1)	I(0)	-
Nominal	Prix	Or	I(1)	I(1)	I(1)	(1-L)log
Nominal	Prix	Pétrole	I(1)	I(1)	I(1)	(1-L)log
Nominal	Prix	Matières premières	I(1)	I(1)	I(1)	(1-L)log
Nominal	Prix	IPC	I(1)	I(1)	I(1)	(1-L)log
International	Taux de change	Euro/dollar	I(1)	I(1)	I(0)	(1-L)log
International	Taux de change	Euro/uk	I(1)	I(1)	I(0)	(1-L)log
International	Taux de change	Euro/yen	I(0)	I(0)	I(1)	(1-L)log
International	Taux de change	Euro/yuan	I(1)	I(1)	I(1)	(1-L)log
International	Taux de change	Taux effectif de l'euro	I(1)	I(1)	I(1)	(1-L)log
International	Taux de change	Taux effectif réel de l'euro	I(1)	I(1)	I(0)	(1-L)log
International	Allemagne	Situation actuelle (IFO)	I(0)	I(0)	I(0)	-
International	Allemagne	Perspectives d'activité à 6 mois (IFO)	I(0)	I(0)	I(0)	-
International	Allemagne	Situation actuelle (ZEW)	I(0)	I(0)	I(0)	-
International	Allemagne	Perspectives économiques (ZEW)	I(0)	I(0)	I(0)	-
International	Allemagne	IPI manufacturé	I(1)	I(1)	I(1)	(1-L)log
International	US	Ventes de détail (en valeur)	I(1)	I(1)	I(1)	(1-L)log
International	US	IPI manufacturé	I(1)	I(1)	I(1)	(1-L)log
International	US	Emploi	I(1)	I(1)	I(1)	(1-L)
International	US	Taux de chômage	I(1)	I(1)	I(1)	(1-L)
International	US	PMI manufacturé	I(0)	I(0)	I(0)	-

Annexe : Modèle espace-état et filtre de Kalman

Cette annexe décrit la représentation espace état, puis le filtre de Kalman qui permet d'estimer les variables cachées (les facteurs dans notre application) à tout instant t conditionnellement aux variables observées jusqu'à t . L'algorithme de lissage permet ensuite d'affiner l'estimation en prenant en compte à tout instant l'ensemble des observations de Y^{28} .

A.1 Le modèle espace - état

Le modèle espace-état du processus multivarié Y_t est représenté par le système d'équations (1) et (2) suivant :

$$\begin{cases} Z_{t+1} = AZ_t + \varepsilon_t & (1) \\ Y_t = CZ_t + \eta_t & (2) \end{cases} \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \approx N\left(0, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}\right)$$

avec $Z_0 \sim N(\mu, P)$. L'équation (1) est l'équation d'état et l'équation (2) est l'équation de mesure.

On appelle par ailleurs :

Y_t	la variable de mesure à la date t
Z_t	la variable d'état à la date t
ε_t	le vecteur d'innovations à la date t
η_t	le vecteur des erreurs de mesure à la date t
A	la matrice de transition
C	la matrice de mesure
CZ_t	le signal à la date t

On considère ici le système sous sa *forme canonique* car on suppose que :

$$E(\varepsilon_t \eta_s) = E(\varepsilon_t Z_0) = E(\eta_t Z_0) = 0 \quad \forall t, s = 1, \dots, T$$

On suppose ici par ailleurs que les matrices A et C ne dépendent pas du temps (le système est dit *invariant par rapport au temps*).

A.2 Le filtre de Kalman

Le filtre de Kalman permet d'estimer à chaque instant t le vecteur d'état conditionnellement à la variable Y observée jusqu'à la date t , soit $Z_{t|t}^* = E(Z_t | Y_0, \dots, Y_t)$, $t=1, \dots, T$. Cet algorithme consiste à itérer les cinq étapes suivantes pour $t=1, \dots, T$:

$$\begin{cases} (1) & Z_{t|t}^* = Z_{t|t-1}^* + K_t (Y_t - CZ_{t|t-1}^*) \\ (2) & \Sigma_{t|t} = (I - K_t C) \Sigma_{t|t-1} \\ (3) & Z_{t+1|t}^* = AZ_{t|t}^* \\ (4) & \Sigma_{t+1|t} = A \Sigma_{t|t} A' + Q \\ (5) & K_t = \Sigma_{t|t-1} C' (C \Sigma_{t|t-1} C' + R)^{-1} \end{cases}$$

²⁸ Vous trouverez plus de détails et des éléments de démonstration par exemple dans le chapitre 15 de l'ouvrage de C. Gourieroux et A. Monfort, *Séries temporelles et modèles dynamiques*.

avec $Z_{t|t}^* = E(Z_t | Y_0, \dots, Y_t)$ l'estimation courante du vecteur d'état, $Z_{t+1|t}^* = E(Z_{t+1} | Y_0, \dots, Y_t)$ la prévision de Z_t pour l'instant $t+1$ faite en t et $\Sigma_{t+1|t} = V(Z_t - Z_{t+1|t}^*)$ la variance de l'erreur de prévision. La matrice K_t est appelée *matrice de gain*. Les deux premières équations actualisent l'estimation de Z_t (équation 1) et la précision correspondante (équation 2). Dans l'équation (1), l'estimation courante du vecteur d'état est obtenue à partir de l'estimation faite à la date $t-1$, à laquelle on ajoute un terme prenant en compte l'information nouvelle contenue dans l'observation de Y_t ; ce terme, qui se déduit d'un programme de minimisation de la variance de l'erreur courante, s'exprime simplement en fonction de l'estimation à la date $t-1$ de Z_t , de la variable de mesure Y_t , et de la matrice de gain K_t , définie par l'équation (5). Les équations (3) et (4) mettent à jour l'estimation de la variable d'état en $t+1$ à partir de l'équation d'état.

L'initialisation de l'algorithme nécessite la connaissance de μ et de P . En effet, $Z_{1|0}^* = \mu$ et $\Sigma_{1|0} = P$. Il faut donc avoir un *a priori* sur Z_0 pour que le processus converge.

A.3 L'algorithme de lissage

L'algorithme de lissage permet ensuite d'affiner l'estimation des états en conditionnant l'estimation non pas à la seule information présente et passée mais à l'ensemble de l'information. On itère cette fois-ci en arrière les calculs suivants pour $t=T-1$ à 1.

$$\begin{cases} Z_{t|T}^* = Z_{t|t}^* + F_t (Z_{t+1|T}^* - Z_{t+1|t}^*) & \text{avec } F_t = \Sigma_{t|t} A' \Sigma_{t+1|t}^{-1} \\ \Sigma_{t|T} = \Sigma_{t|t} + F_t (\Sigma_{t+1|T} - \Sigma_{t+1|t}) F_t' \end{cases}$$

Les calculs sont initialisés en T par les produits du filtre.

L'ensemble des calculs précédents (filtre et lissage) suppose que les matrices A , C , Q et R sont connues. Si tel n'est pas le cas, la log-vraisemblance du modèle peut être construite au cours des itérations du filtre et l'on pourra estimer les paramètres inconnus par maximum de vraisemblance. Dans notre application, on utilise comme matrices A et C les estimateurs de la première étape et comme matrices Q et R les matrices de variance covariance des résidus associés.