

OPTIM : un outil de prévision trimestrielle du PIB de la France

**Karim BARHOUMI, Véronique BRUNHES-LESAGE,
Laurent FERRARA, Bertrand PLUYAUD
et Béatrice ROUVREAU**

*Direction des Analyses macroéconomiques et de la Prévision
Service du Diagnostic conjoncturel*

Olivier DARNÉ

*Direction des Analyses macroéconomiques
et de la Prévision*

*Service du Diagnostic conjoncturel
Université Paris X–Nanterre (EconomiX)*

Cet article présente une version révisée du modèle OPTIM proposé par Irac et Sédillot (2002) et utilisé à la Banque de France afin de prévoir les taux de croissance du PIB français et de ses différentes composantes sur deux trimestres (en cours et suivant). Le modèle OPTIM est destiné à être utilisé tous les mois en intégrant de l'information économique mensuelle à l'aide d'équations d'étalonnage des variables des comptes trimestriels. Pour chaque composante du PIB (côtés offre et demande), des équations d'étalonnage sont spécifiées en utilisant une approche "general-to-specific" mise en œuvre de manière automatique par Hoover et Perez (1999) et améliorée par Krolzig et Hendry (2001). Cette approche permet de sélectionner automatiquement de manière optimale des variables explicatives adéquates parmi un large ensemble d'informations économiques. Une analyse en prévision dynamique est menée afin d'estimer la capacité prévisionnelle du nouveau modèle OPTIM. Les résultats obtenus montrent que le modèle fournit de meilleures prévisions que des modèles statistiques simples.

Mots-clés : Prévision, taux de croissance du PIB,
modèle d'étalonnage, approche "general-to-specific"

Codes JEL : C52, C53, E20

La connaissance dans les délais les plus brefs de la situation économique en cours d'un pays, ainsi que de ses perspectives de croissance, est un enjeu central pour les décideurs économiques et politiques. La mesure la plus large de l'activité économique d'un pays reste le PIB trimestriel dont les chiffres officiels sont publiés plusieurs semaines après la fin du trimestre par les comptes nationaux. Par exemple, en France, le délai de publication par l'INSEE est d'environ 45 jours. Il est alors utile pour une institution comme la Banque de France de pouvoir établir un diagnostic conjoncturel de l'économie française avant la publication des résultats des comptes trimestriels.

Dans cette optique, ce papier présente une nouvelle version du modèle OPTIM¹, développé par Irac et Sédillot (2002), utilisé par la Banque de France afin de prévoir le taux de croissance trimestriel du PIB de la France et de ses différentes composantes. Son objectif principal est de produire tous les mois une évaluation, aussi précise que possible, du taux de croissance du PIB, pour les trimestres coïncident et futur. À partir d'un grand nombre de séries de fréquence mensuelle, l'approche consiste dans une première étape à prévoir les principales composantes des comptes nationaux (à la fois du côté offre et demande), puis dans une deuxième étape à les agréger de manière à prévoir le PIB dans son ensemble. Une telle décomposition a l'avantage de fournir des informations quantitatives permettant une analyse plus détaillée de la situation économique.

De nombreuses méthodes quantitatives ont été proposées dans la littérature afin de prévoir l'activité réelle de court terme d'une économie. Actuellement, deux grands types d'approche émergent. Une première approche consiste à effectuer des prévisions avec des modèles à facteurs dynamiques comme cela a été développé, par exemple, dans les travaux de Stock et Watson (2002, 2006), Forni *et alii* (2003), Schumacher (2007). Cette approche utilise un nombre élevé de variables qui sont décomposées en un nombre réduit de facteurs communs et en composantes idiosyncratiques. La méthodologie des modèles à facteurs dynamiques est de plus en plus populaire dans la littérature économétrique, en particulier dans une optique de prévision macroéconomique. Toutefois, cette approche est souvent perçue par les conjoncturistes comme une « boîte noire » dans le sens

où les résultats obtenus sont difficiles à interpréter d'un point de vue économique (voir sur ce point Banbura et Rünstler, 2007). La seconde approche est peut-être moins innovante sur le plan économétrique mais reste plus informative pour interpréter les fluctuations économiques. Il s'agit de construire des équations linéaires d'étalonnage des variables des comptes trimestriels à partir d'informations mensuelles. Ce type de modèles, connu sous le nom de "bridge models", a été utilisé par plusieurs auteurs pour prévoir le taux de croissance du PIB (cf. Sédillot et Pain, 2003 ; Rünstler et Sédillot, 2003 ; Baffigi *et alii*, 2004 ou Diron, 2006). C'est cette dernière approche que nous retenons.

|| Objectifs

Le premier objectif visé est de construire un modèle d'étalonnage qui combine différentes équations pour les principales composantes du PIB de la France. Le PIB est désagrégé selon les composantes de l'offre et celles de la demande. Pour chaque composante, un modèle d'étalonnage est estimé à partir de variables sélectionnées par une procédure automatique, dite « du général au spécifique » : en fonction des résultats de tests statistiques, un modèle ne retenant que les variables les plus pertinentes est extrait d'un modèle de départ le plus large possible. Cette procédure de sélection permet au conjoncturiste d'exploiter un grand nombre de séries macroéconomiques et elle apporte un cadre économétrique robuste, transparent et systématique à la sélection de variables explicatives dans les équations de prévision. Les composantes sont ensuite agrégées afin de retrouver le chiffre du PIB.

Pour réaliser cet exercice, nous utilisons deux types de variables mensuelles. Le premier type de variables, dites "hard", sont des indicateurs quantitatifs de l'activité réelle, souvent d'origine comptable (indice de production industrielle, consommation des ménages en produit manufacturés, etc.), alors que le second type de variables, dites "soft", correspond aux données d'enquêtes réalisées auprès des ménages et des chefs d'entreprise. Les variables du type "soft" présentent plusieurs avantages par rapport aux séries dites "hard". En particulier, elles sont rapidement disponibles, c'est-à-dire bien plus rapidement que les données relatives aux principaux agrégats macroéconomiques,

¹ On renvoie à l'article de Barhoumi *et alii* (2007) pour les résultats détaillés du modèle.

et (très) peu révisées. De plus, les enquêtes de conjoncture apportent des indications précoces sur le passé récent et les perspectives d'évolution à court terme du comportement des acteurs économiques.

À partir de ce modèle d'étalonnage estimé, l'objectif est de fournir tous les mois une prévision du taux de croissance trimestriel du PIB pour le trimestre en cours (ou « trimestre coïncident ») et le suivant (« trimestre futur »)². Au fur et à mesure que l'ensemble des informations augmente au sein du trimestre, le modèle doit être capable de fournir une prévision dont l'incertitude tend à se réduire. Nous construisons un modèle pour prévoir le PIB du trimestre en cours et un autre modèle pour prévoir celui du trimestre suivant, en essayant de sélectionner dans chaque cas les indicateurs les plus pertinents.

2| Stratégie de modélisation et sélection des données

2| I Une prévision détaillée

Les variables intégrées dans le modèle OPTIM sont issues des comptes trimestriels publiés par l'INSEE. La décomposition du PIB est effectuée aussi bien du côté de l'offre que du côté de la demande. La prévision de PIB proprement dite est établie à partir des prévisions des différentes composantes de la production et s'appuie donc uniquement sur l'approche par l'offre. Toutefois, la modélisation des composantes de la demande apporte un éclairage complémentaire et reste essentielle pour l'analyse conjoncturelle.

La décomposition retenue ici n'est pas exhaustive : certaines composantes, difficiles à prévoir par des équations fondées sur des raisonnements économiques, n'ont pas été modélisées. C'est en particulier le cas de la contribution à la croissance du PIB de la variation des stocks, qui est directement calculée comme la différence entre la croissance du PIB et la somme des contributions des autres composantes de la demande. De plus, certaines sous-composantes, telles que la production de services non marchands ou l'investissement immatériel, n'ont également pas été

modélisées. Enfin, les consommations intermédiaires sont ignorées : la prévision du PIB est établie à partir de la production prévue et non de la valeur ajoutée. En effet, des données mensuelles essentielles, telles que l'indice de la production industrielle, se rapportent explicitement à la production, tandis que l'information disponible sur les consommations intermédiaires et la valeur ajoutée est limitée. Au final, les composantes et sous-composantes modélisées sont les suivantes.

Du côté de la demande

- Consommation des ménages, calculée par agrégation des prévisions de :
 - consommation des ménages en produits agricoles et agroalimentaires ;
 - consommation des ménages en produits manufacturés ;
 - consommation des ménages en énergie ;
 - consommation des ménages en services.
- Consommation des administrations publiques.
- Investissement, calculé par agrégation des prévisions de :
 - investissement des entreprises en matériel et équipement ;
 - investissement des entreprises en construction ;
 - investissement des ménages ;
 - investissement des administrations publiques.
- Exportations.

Du côté de l'offre

- Importations.
- Production, calculée par agrégation des prévisions de :
 - production de produits agricoles et agroalimentaires ;

2 Par trimestre en cours ou coïncident, on entend le premier trimestre pour lequel le résultat du PIB n'a pas encore été publié.

- production de produits manufacturés ;
- production d'énergie ;
- production en construction ;
- production de services marchands.

Le PIB est prévu par une équation d'étalonnage sur la production totale.

Ainsi, la production, la consommation et l'investissement sont agrégés au moyen d'équations estimées sur leurs sous-composantes : les poids des sous-composantes sont estimés par les coefficients de l'équation d'agrégation, qui reflètent une moyenne des poids sur la période d'estimation. De manière à limiter l'écart entre ces poids estimés et les poids actuels, des périodes d'estimations relativement courtes ont été choisies pour les équations d'agrégation (six ans).

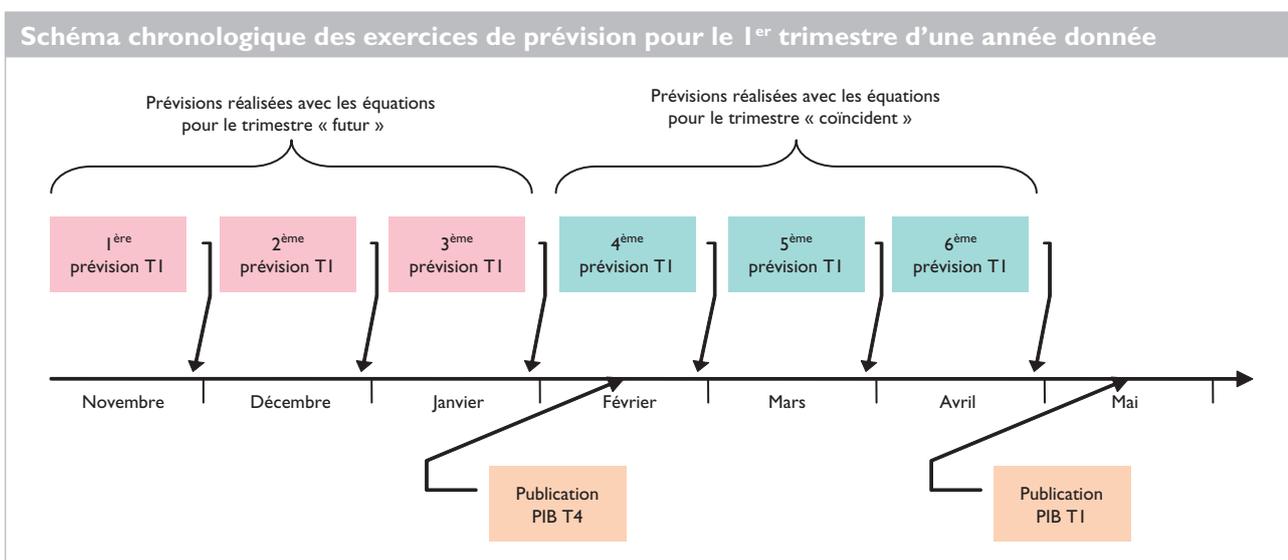
2|2 Des exercices mensuels

Ce modèle a été conçu pour être utilisé chaque mois et est fondé sur des données explicatives disponibles en fréquence mensuelle (la version précédente du modèle visait plutôt une fréquence trimestrielle et utilisait donc en grande partie des informations trimestrielles). Les prévisions étant réalisées chaque mois pour le trimestre coïncident et le trimestre

suivant, six exercices de prévision permettent au total de prévoir la croissance du PIB pour un trimestre donné. De manière conventionnelle, les prévisions sont réalisées en fin de mois (juste après la publication des enquêtes de l'INSEE dans l'industrie et les services et de celles de la Commission européenne et juste avant la première réunion du mois du Conseil des gouverneurs de la Banque centrale européenne), la dernière des six prévisions étant effectuée 15 jours avant la publication des données du PIB, soit 30 jours après la fin du trimestre à prévoir. Le schéma ci-dessous présente le déroulement chronologique des exercices de prévision pour le PIB du premier trimestre d'une année donnée.

D'une manière générale, pour chaque composante du PIB, une seule et même équation est utilisée pour les trois prévisions du trimestre coïncident et une autre pour les trois prévisions du trimestre suivant. La seule exception concerne les équations pour la production du trimestre coïncident : l'indice de la production industrielle (IPI) apporte une information pertinente pour prévoir la production, mais est publié avec un délai plus important que les données d'enquête. Par conséquent, des équations sans IPI et avec IPI sont estimées, les unes ou les autres étant utilisées selon l'information disponible à la date de réalisation de l'exercice de prévision.

Les séries mensuelles sont converties en fréquence trimestrielle prenant comme valeur pour chaque trimestre la moyenne des valeurs sur les trois mois correspondants. Sur le dernier trimestre, lorsque les



valeurs d'une série ne sont connues que pour un ou deux mois, on affecte comme valeur au trimestre la moyenne des trois derniers mois connus.

2|3 Une base de données large

Les équations sont estimées en fréquence trimestrielle mais reposent sur de l'information mensuelle. Ainsi, les séries disponibles uniquement en fréquence trimestrielle ont été exclues. Lorsque des données ne sont pas disponibles sur certains mois du trimestre en cours, la valeur manquante d'une série pour ce trimestre est fixée comme égale à la moyenne mobile des trois derniers mois connus.

Les séries utilisées correspondent aussi bien à des données macroéconomiques mensuelles (telles que la consommation des ménages en produits manufacturés) qu'à des données d'enquête. En ce qui concerne les équations d'importation et d'exportation, un traitement spécifique a été appliqué aux séries d'enquête de la Commission européenne : pour les différentes questions de l'enquête, les séries correspondantes pour chaque pays ne sont pas prises en compte séparément, mais de manière agrégée, en utilisant comme pondérations les poids des pays dans les importations de la France (les poids pour les exportations sont proches). Seules les données relatives aux principaux partenaires commerciaux européens sont prises en compte (soit 59,3 % du total des importations de la France pour 2004, qui est l'année prise comme référence).

Le tableau de l'annexe 2 présente les sources des séries utilisées pour chacune des équations du modèle.

2|4 Des équations interprétables économiquement

Bien qu'OPTIM ne soit pas un modèle structurel, la signification économique des relations économétriques mises en évidence est prise en compte, afin que celles-ci puissent apporter une contribution au diagnostic économique. De fait, l'objectif des exercices de prévision n'est pas seulement de fournir une prévision précise du PIB mais également d'élaborer un

scénario économique cohérent. Ainsi, les coefficients des équations sélectionnées doivent présenter des signes reflétant une logique économique : par exemple, une relation négative entre le chômage et la consommation des ménages. Une sélection des équations fondée sur des raisonnements économiques permet également de limiter le risque d'obtenir des régressions fallacieuses ou peu robustes.

2|5 La méthode de sélection des données

La méthode de sélection des données se veut aussi robuste que possible et facilement reproductible. En effet, les relations statistiques peuvent s'éroder avec le temps et il est utile de pouvoir les réestimer facilement. La sélection des variables se fait de la manière suivante :

- un ensemble de séries, correspondant à une première source d'information, est présélectionné. La source en question est choisie pour le lien qu'elle présente *a priori* avec la variable modélisée : par exemple, l'enquête mensuelle de conjoncture dans les services de la Banque de France, pour la consommation de services. Les séries présélectionnées doivent présenter une forte corrélation avec la variable à expliquer et, au contraire, une corrélation la moins forte possible entre elles (de manière à apporter des informations complémentaires) ;
- parmi cet ensemble de séries, les variables les plus pertinentes, retardées ou non, sont sélectionnées *via* une procédure automatique dénommée Gets ("*general-to-specific*"), élaborée par Hoover et Perez (1999) et mise en application par Krolzig et Hendry (2001). Cette procédure, mise en œuvre à l'aide du module Grocer (cf. Dubois, 2003) intégré au logiciel Scilab, effectue de manière systématique un certain nombre de tests statistiques permettant de sélectionner une équation aussi fiable que possible. La forme générale des équations est la suivante :

$$y_t = \alpha + \sum_{i=1}^m \beta_i Y_{t-i} + \sum_{j=1}^q \sum_{i=0}^k \delta_{ji} X_{j,t-i} + \varepsilon_t$$

où Y est la variable expliquée et X_j la $j^{\text{ème}}$ variable explicative ;

- un nouvel ensemble de séries est ensuite présélectionné à partir d'une autre source et fusionné avec les séries isolées par la procédure Gets. Cette même procédure est appliquée au nouvel ensemble et permet d'aboutir à une nouvelle équation. Cette opération est renouvelée de manière à obtenir des équations *a priori* satisfaisantes statistiquement pour différentes combinaisons de sources d'information qui paraissent intéressantes (par exemple, pour la consommation en services : enquête mensuelle de conjoncture dans les services de la Banque de France et enquête auprès des ménages de l'INSEE). Cette procédure d'augmentation progressive de l'ensemble d'information est liée au fait qu'en pratique, la sélection par la méthode Gets ne peut s'opérer que parmi un nombre limité de séries (une vingtaine au maximum) ;

- le choix de l'équation finale se fait en comparant les qualités statistiques et l'interprétabilité économique des différentes équations, mais surtout en comparant les performances en prévision, en simulant les équations sur le passé, par une procédure de prévisions récursives ("*rolling forecast*"). Ainsi, l'équation est estimée jusqu'à une date donnée et une prévision est réalisée sur la période suivante, puis l'équation est estimée de nouveau jusqu'à la date suivante et simulée en prévision sur une période supplémentaire et ainsi de suite. La mesure de la qualité de l'équation en prévision est obtenue par le RMSE³. Ces simulations sont réalisées en tenant compte de la disponibilité des variables explicatives : chaque estimation et chaque prévision sont effectuées en excluant pour les variables explicatives les données qui n'étaient pas

connues à la date correspondant à la simulation (par exemple, lors du premier exercice de prévision sur le premier trimestre de 2000, considéré comme un trimestre coïncident, l'IPI n'est connu que jusqu'en décembre 1999). Cette méthode permet de comparer les performances d'équations fondées sur des sources différentes, ayant des délais de publication différents : une équation fondée sur des données d'enquête peut ainsi s'avérer moins performante qu'une équation fondée sur l'IPI pour retracer en historique les évolutions de la variable à expliquer, mais plus performante en prévision, du fait du délai de publication plus court des enquêtes, qui permet d'avoir une information plus complète au moment de la prévision.

3 | Résultats

Au total, plus d'une trentaine d'équations ont été spécifiées et estimées sur le PIB et ses composantes. L'ensemble des équations est présenté dans Barhoumi *et alii* (2007). Seules quatre d'entre elles sont reportées ci-dessous à titre d'exemple : deux du côté de l'offre et deux du côté de la demande. Ces équations présentent des caractéristiques recherchées lors de la conception du modèle : elles mobilisent une large palette d'indicateurs et sont facilement interprétables d'un point de vue économique.

- L'équation de production de biens manufacturés pour le trimestre coïncident incluant l'IPI manufacturier, estimée sur la période T1 1991 – T4 2006, est la suivante :

$$PMANU_GT = 0,082 + 0,371PMANU_GT_{t-4} + 0,910IPI_MANUF_GT_t + 0,038\Delta (PRODPREV_MANUF)_t$$

t - stat (0,989) (7,099) (13,980) (3,334)

$R^2 = 0,83$; $SE = 0,55$; $DW = 2,27$

3 Racine de la moyenne arithmétique des carrés des erreurs, ou "Root mean squared error", définie par $RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (PIB_{prévu,t} - PIB_{réalisé,t})^2}$.

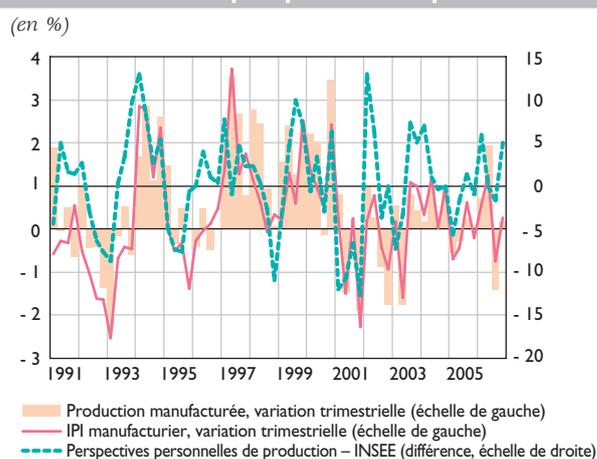
Les RMSE sont mesurés à partir des derniers chiffres du PIB connus à la date de l'écriture du document et non pas à partir des chiffres du PIB fournis en temps réel par les premiers résultats de l'INSEE. L'analyse des performances du modèle OPTIM sur la série des premières publications sera entreprise ultérieurement.

où $PMANU_GT$ est le taux de croissance trimestriel de la production manufacturière, IPI_MANUF_GT est le taux de croissance trimestriel de l'IPI manufacturier et $PRODPREV_MANUF$ est la réponse à l'enquête mensuelle de conjoncture dans l'industrie menée par l'INSEE relative aux perspectives de production personnelles. La présence de la production manufacturière retardée de quatre trimestres permet de prendre en compte un résidu de saisonnalité annuelle restant dans la série désaisonnalisée. Le critère du R^2 confirme que cette équation est de bonne qualité statistique ⁴.

Cette équation présente l'avantage de reposer sur deux types de variable : une série macroéconomique mensuelle, l'IPI, qui apporte une information fiable sur l'évolution de la variable à prévoir, mais n'est disponible que quarante jours après la fin du mois considéré ⁵, et une série issue d'une enquête, qui est moins corrélée avec la variable à prévoir, mais est disponible en fin de mois. Si l'une des deux sources utilisées donne un faux signal sur l'évolution de la conjoncture, celui-ci peut être tempéré par l'évolution de l'autre indicateur. Par ailleurs, les deux séries ont trait à l'évolution de la production, mais sont complémentaires dans la mesure où l'IPI porte sur l'évolution de la production constatée tandis que la série d'enquête utilisée est le solde d'opinion relatif aux perspectives de production.

Le graphique 1 met en évidence la très forte corrélation entre la production manufacturée mesurée dans les comptes trimestriels et l'IPI manufacturier. Toutefois, en pratique, lors de la réalisation de la première prévision sur le trimestre coïncident, l'IPI n'est connu que sur le dernier mois

Graphique 1 Production manufacturée, IPI manufacturier et perspectives de production



Source : INSEE

du trimestre précédent et le recours à une équation alternative, fondée uniquement sur des données d'enquête est utile. Les simulations effectuées ont ainsi montré qu'une équation où l'IPI est remplacé par le solde d'opinion relatif à la production passée issu de l'enquête INSEE dans l'industrie (disponible dès la fin du mois) apporte une meilleure précision lors des deux premières prévisions du trimestre (RMSE de 0,82, contre 1,14 pour le premier exercice et de 0,79, contre 1,07 pour le deuxième). Pour le troisième exercice, en revanche, le résultat est inversé (RMSE de 0,79, contre 0,71 pour l'équation avec IPI).

• L'équation de production de services marchands pour le trimestre coïncident excluant l'IPI, estimée sur la période T1 1991 – T4 2006, est la suivante :

$$\begin{aligned}
 PSERM_GT = & 1,53 + 0,24PSERM_GT_{t-1} - 0,20PSERM_GT_{t-4} + 0,04\Delta(EVACT_SV)_t + 0,03\Delta(NIVTRES_SV)_{t-1} - 0,02OPPEPAR_t - 0,012CHOMPREV_t \\
 t - stat & (5,69) \quad (2,11) \quad (-2,10) \quad (5,21) \quad (3,22) \quad (-3,47) \quad (-4,55) \\
 R^2 = & 0,71 ; SE = 0,31 ; DW = 1,96
 \end{aligned}$$

⁴ La qualité statistique des modèles a été vérifiée avec un certain nombre de diagnostics statistiques (cf. Barhoumi et alii).

⁵ L'IPI pour l'ensemble du trimestre t n'est connu que quelques jours avant la première publication du PIB du trimestre t par l'INSEE.

où $PSERM_GT$ est le taux de croissance trimestriel de la production de services marchands, $NIVTRES_SV$ et $EVACT_SV$ sont les réponses relatives, respectivement, au niveau de trésorerie et à l'activité passée dans les services marchands dans l'enquête de la Banque de France, et $OPPEPAR$ et $CHOMPREV$ sont les réponses relatives, respectivement, à l'opportunité d'épargner et au chômage prévu dans l'enquête auprès des ménages menée par l'INSEE. On introduit les premiers et seconds retards de la série du taux de croissance de la production de services afin de blanchir la série des résidus. La qualité de l'équation mesurée par le critère du R^2 est élevée. La présence dans l'équation de la variable endogène retardée de quatre trimestres est due à la trace d'une saisonnalité encore présente dans les données corrigées des variations saisonnières. Les paramètres estimés des séries d'enquêtes possèdent bien le signe attendu, que ce soit pour les deux séries d'enquêtes auprès des ménages ou pour les deux séries d'enquêtes dans les services. En effet, il semble naturel qu'une hausse de l'opportunité d'épargner ou du chômage anticipé agisse négativement sur la production de services.

L'équation repose sur deux sources complémentaires : l'enquête dans les services porte sur la situation des entreprises et de leur activité, tandis que l'enquête auprès des consommateurs apporte un éclairage sur la demande adressée aux entreprises. La variété des séries utilisées permet de ne pas se focaliser sur une seule information qui pourrait s'avérer erronée et facilite l'analyse économique des évolutions de la production : si le solde d'opinion relatif à l'activité passée est directement lié à l'évolution de la production, la série de niveau de trésorerie est liée à la situation financière des entreprises, l'opportunité d'épargner donne une indication sur la demande potentielle des ménages (en fonction des revenus, des évolutions financières, des prix...), tandis que le chômage prévu permet d'appréhender les effets de la situation du marché du travail sur la demande.

• L'équation relative aux exportations pour le trimestre coïncident, estimée sur la période T1 1991 – T4 2006, est la suivante :

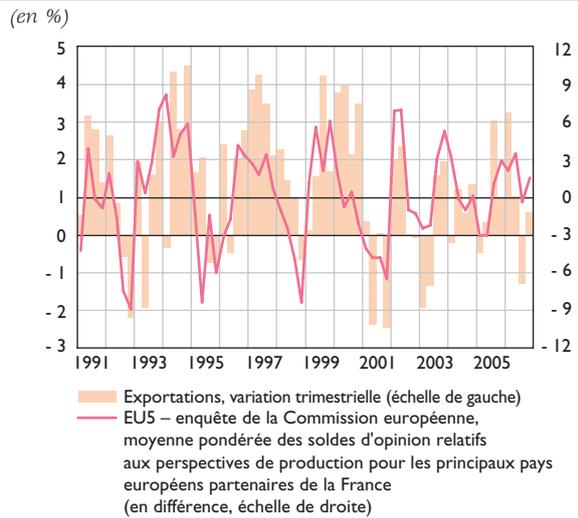
$$EXPORT_GT = 0,171 + 0,136\Delta(EU5)_t + 0,132EVCOM_I_{t-1}$$

<i>t - stat</i>	(0,74)	(3,37)	(6,12)
-----------------	--------	--------	--------

$R^2 = 0,53 ; SE = 1,26 ; DW = 2,16$

où $EXPORT_GT$ est le taux de croissance trimestriel des exportations et $EVCOM_I$ est la réponse relative

Graphique 2 Exportation et perspectives de production dans les principaux pays européens partenaires de la France



Sources : INSEE, Commission européenne ; calculs : Banque de France

à l'évolution des commandes étrangères issue de l'enquête mensuelle dans l'industrie de la Banque de France (en différences). La variable EU5 a été créée à partir des réponses relatives aux perspectives de production pour les mois à venir dans l'enquête de la Commission européenne pour différents pays européens, qui comptent parmi les principaux partenaires commerciaux de la France. Il s'agit d'une moyenne pondérée par l'importance relative de chacun de ces pays dans le commerce extérieur de la France. Cette variable approche donc d'une certaine manière la demande adressée par ces pays à la France.

L'équation relative aux importations pour le trimestre coïncident, estimée sur la période T1 1991 – T4 2006, est la suivante :

$$IMPORT_GT = 0,234 + 0,193\Delta(EU5)_t + 0,125EVCOM_I_{t-1}$$

<i>t - stat</i>	(1,47)	(5,71)	(6,64)
-----------------	--------	--------	--------

$R^2 = 0,56 ; SE = 1,09 ; DW = 1,89$

où $IMPORT_GT$ est le taux de croissance trimestriel des importations et $EVCOM_I$ est la réponse relative à l'évolution des commandes globales issue de l'enquête mensuelle dans l'industrie de la Banque de France (en différences).

Cette équation présente la même structure que l'équation d'exportations ; seule la variable relative

à l'évolution des commandes étrangères a été remplacée par l'évolution des commandes globales issue de l'enquête mensuelle dans l'industrie de la Banque de France. Dans le cas de l'équation d'importations, la présence de la variable EU5 peut s'expliquer de deux manières : d'une part, une hausse des importations de la France a un impact direct sur l'activité économique de ses principaux partenaires et d'autre part, une hausse des importations françaises se produit généralement dans une phase d'expansion de l'économie, or les cycles économiques de la France et de ses voisins européens sont très proches.

Performances relatives en prévision

En termes de prévision, les résultats obtenus concernent l'exercice de prévisions récursives sur les trimestres coïncident et futur, du premier trimestre 2000 au quatrième trimestre 2006, soit 28 observations. Afin d'établir des comparaisons avec un modèle de référence, nous effectuons des prévisions sur chaque composante à l'aide d'une approche naïve (la valeur prévue est la valeur précédente) et d'un modèle autorégressif de type AR(p), avec un nombre de retards $p \leq 4$ (cf. Annexe 1).

S'agissant du trimestre coïncident, les RMSE sont plus faibles que ceux des prédicteurs naïfs et autorégressifs

pour toutes les équations d'OPTIM. De plus, on observe que, comme cela était attendu, la précision des prévisions croît au fil du temps (les RMSE diminuent entre la première et la troisième prévision). En ce qui concerne le PIB agrégé, les RMSE pour les première, deuxième et troisième prévisions sont de, respectivement, 0,27 point de pourcentage, 0,25 point de pourcentage et 0,23 point de pourcentage. Des tests statistiques⁶ montrent que les prévisions effectuées sur les composantes améliorent significativement les prévisions des modèles de référence, excepté pour la production d'énergie et de construction (deux premiers exercices).

S'agissant du trimestre futur, on note que les RMSE sont naturellement plus élevés que ceux obtenus pour le trimestre coïncident, traduisant une plus grande incertitude. Toutefois, le modèle OPTIM permet d'améliorer la qualité de prévision par rapport à une prévision naïve et à un modèle autorégressif dans tous les cas, sauf pour les prévisions de consommation des ménages et d'importation (deux premiers exercices). On remarque également que pour la plupart des composantes et surtout pour le PIB, les RMSE des prévisions ne s'améliorent pas sensiblement au fil des prévisions. Ceci semble indiquer qu'au cours d'un trimestre, l'apport d'information sur le trimestre à venir est négligeable.

La version révisée du modèle OPTIM permet de prévoir le PIB français et ses principales composantes à partir d'équations d'étalonnage couvrant une large palette d'indicateurs conjoncturels. Les principaux traits du modèle sont une fréquence mensuelle pour les prévisions et une sélection des variables qui, d'une part, se fonde sur une approche systématique, dite "general-to-specific" et qui, d'autre part, tient compte de la disponibilité des informations, à travers un exercice de simulation dynamique.

En termes de RMSE, le nouveau modèle améliore significativement les prévisions de référence (modèles autorégressif et « naïf ») et appuie ainsi le diagnostic conjoncturel de l'économie française dont dispose la Banque de France pour participer aux décisions de l'Eurosystème.

6 On utilise le test de Diebold-Mariano modifié par Harvey, Leybourne et Newbold (1997).

Bibliographie

Baffigi (A.), Golinelli (R.) et Parigi (G.) (2004)

"Bridge model to forecast the euro area GDP", *International Journal of Forecasting*, n° 20, p. 447-460

Banbura (M.) et Rünstler (G.) (2007)

"A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP", Banque centrale européenne, *Working Paper*, n° 751

Barhoumi (K.), Brunhes-Lesage (V.), Darné (O.), Ferrara (L.), Pluyaud (B.) et Rouvreau (B.) (2007)

"Monthly forecasting of French GDP: a revised version of the OPTIM model", article présenté au colloque "Forecasting short-term economic developments and the role of econometric models", Banque du Canada, 25-26 octobre

Diron (M.) (2006)

"Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data", Banque centrale européenne, *Working Paper*, n° 622

Dubois (E.) (2003)

"GROCER: an econometric toolbox for Scilab", Mimeo, <http://dubois.ensae.net/grocer.html>

Dubois (E.) et Michaux (E.) (2006)

« Étalonnages à l'aide d'enquêtes de conjoncture : de nouveaux résultats », *Économie et Prévision*, n° 172, p. 11-28

Forni (M.), Hallin (M.), Lippi (M.) et Reichlin (L.) (2003)

"Do financial variables help forecasting inflation and real activity in the euro area?", *Journal of Monetary Economics*, n° 50, p. 1243-1255

Harvey (D.), Leybourne (S.) et Newbold (P.) (1997)

"Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, n° 13, p. 281-291

Hoover (K.) et Perez (S.) (1999)

"Data mining reconsidered: encompassing and the general-to-specific approach to specification search", *Econometrics Journal*, n° 2, p. 167-191

Irac (D.) et Sédillot (F.) (2002)

"Short run assessment of French economic activity using OPTIM", Notes d'études et de recherche, n° 88, Banque de France, <http://www.banque-france.fr/gb/publications/telnomot/ner/ner88.pdf>

Krolzig (H.-M.) et Hendry (D. F.) (2001)

"Computer automation of general-to-specific model selection procedures", *Journal of Economic Dynamics and Control*, n° 25, p. 831-866

Rünstler (G.) et Sédillot (F.) (2003)

"Short-term estimates of GDP by means of monthly data", Banque centrale européenne, *Working Paper*, n° 176

Schumacher (C.) (2007)

"Forecasting German GDP using alternative factor models based on large datasets", *Journal of Forecasting*, n° 26, p. 271-302

Sédillot (F.) et Pain (N.) (2003)

"Indicator models of real GDP growth in selected OECD countries", *Economic Department Working Paper*, n° 364, OCDE

Stock (J. H.) et Watson (M. W.) (2002)

"Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, n° 20, p. 147-162

Stock (J. H.) et Watson (M. W.) (2006)

"Forecasting with many predictors", dans "Handbook of Economic Forecasting", Elliot G., Granger C. W. J. et Timmermann A. (eds)

Annexe I

Comparaison des performances d'OPTIM avec un modèle autorégressif et un modèle naïf, à l'aide des RMSE

Trimestre coïncident					
Composante	1 ^{ère} prévision	2 ^{ème} prévision	3 ^{ème} prévision	Modèle AR	Modèle naïf
PIB	0,27	0,25	0,23	0,38	0,51
Production agricole et agroalimentaire	0,54	0,54	0,45	0,57	0,68
Production manufacturée	0,82	0,79	0,71	1,28	1,73
Production d'énergie	1,44	1,34	1,21	1,44	2,52
Production en construction	0,62	0,60	0,55	0,67	0,76
Production de services	0,44	0,39	0,34	0,45	0,59
Consommation des ménages	0,26	0,19	0,19	0,33	0,45
Consommation des administr. publiques	0,23	0,23	0,23	0,23	0,28
Investissement	0,80	0,77	0,71	0,87	1,24
Exportations	1,46	1,32	1,27	1,62	2,07
Importations	1,23	1,13	1,13	1,31	1,54
Trimestre futur					
Composante	1 ^{ère} prévision	2 ^{ème} prévision	3 ^{ème} prévision	Modèle AR	Modèle naïf
PIB	0,30	0,30	0,30	0,37	0,48
Production agricole et agroalimentaire	0,66	0,65	0,67	0,77	1,08
Production manufacturée	1,09	1,06	1,06	1,24	1,57
Production d'énergie	1,51	1,51	1,51	1,51	2,07
Production en construction	0,54	0,61	0,61	0,72	0,77
Production de services	0,42	0,42	0,42	0,45	0,54
Consommation des ménages	0,37	0,37	0,34	0,33	0,48
Consommation des administr. publiques	0,24	0,24	0,24	0,24	0,28
Investissement	0,80	0,81	0,80	0,85	1,24
Exportations	1,58	1,53	1,45	1,75	2,33
Importations	1,54	1,50	1,31	1,41	1,54

Annexe 2

Tableau synthétique des sources utilisées pour prévoir les composantes du PIB

	Données d'enquête									Données macroéconomiques mensuelles				
	Enquête Banque de France - industrie	Enquête Banque de France - services	Enquête Banque de France - commerce de détail	Enquête INSEE - industrie	Enquête INSEE - services	Enquête INSEE - construction	Enquête INSEE - consommateurs	Enquête Commission européenne	Enquête FNTP - travaux publics	Indice de la production industrielle	Consommation des ménages en produits manufacturés	Indice des prix à la consommation	Consommation d'électricité	Mises en chantier de logements neufs
Délai de publication <i>(nombre de jours après la fin du mois concerné)</i>	+ 15 *	+ 15 *	+ 15 *	+ 0 *	+ 0 *	+ 0 *	+ 0 *	+ 0 *	+ 35	+ 40	+ 25	+ 20	+ 1	+ 30
Production agricole et agroalimentaire	C - F									C - F				
Production manufacturée				C - F						C				
Production d'énergie										C - F		C		
Production en construction				C - F						C				
Production de services		C			F		C - F			C**				
Consommation en alimentation	C - F											C - F		
Consommation de produits manufacturés	F		F				F				C			
Consommation d'énergie	Trimestre futur modélisé par un modèle autorégressif											C		
Consommation de services		C - F					C - F							
Consommation des APU	Modélisée par un modèle autorégressif													
Investissement des entreprises en équipement	C - F													
Investissement des entreprises en construction						C - F								
Investissement des ménages						C - F								C
Investissement des APU									C - F					
Exportations	C - F							C						
Importations	C - F		F					C						

C = source utilisée dans l'équation pour le trimestre coïncident

F = source utilisée dans l'équation pour le trimestre futur

* Dans les enquêtes Banque de France, les questions portent explicitement sur l'activité au cours du mois précédent le mois de collecte des données, tandis que dans les enquêtes de l'INSEE et de la Commission européenne, les répondants sont interrogés sur leur activité « au cours des trois derniers mois ».

** L'indice de la production industrielle dans le secteur manufacturier est utilisé pour la prévision de production de services.