

Réévaluation des modèles d'estimation précoce de la croissance

Françoise Charpin

OFCE, Centre de recherche
en économie de Sciences Po,
Université Paris II

La *Revue de l'OFCE* est ouverte aux chercheurs et à tous les spécialistes en économie et en sociologie. La revue s'assure de la rigueur des propos qui sont tenus mais les jugements et opinions exprimés par les auteurs, y compris quand ils appartiennent à l'OFCE, n'engagent qu'eux-mêmes et non les institutions auxquelles ils appartiennent.

Dans le n° 108 de la Revue de l'OFCE, des modèles d'estimation précoce de la croissance française ont été proposés et évalués en pseudo temps réel sur la période 2001-2007. La crise financière a quelque peu dégradé leur performance. Le changement de base des comptes trimestriels en mai 2011 modifie aussi sensiblement les résultats car il affecte la croissance du PIB sur toute la période d'estimation des modèles. Ainsi le moment est venu de faire le point sur les outils de prévision à court terme présentés dans l'article de F. Charpin (2009). C'est aussi l'occasion de les confronter à d'autres modèles et d'évaluer l'ensemble de nos outils en pseudo temps réel sur la période allant du 1^{er} trimestre 2001 au 1^{er} trimestre 2011.

francoise.charpin@ofce.sciences-po.fr

Mots clé : Préviation du PIB. Modèles à facteurs.

Dans l'article de F. Charpin (2009), l'estimation de la croissance est réalisée à l'aide de deux modèles à facteurs, l'un est construit à partir de données d'enquêtes et de séries financières, l'autre inclut en plus des données mensuelles réelles (indice de la production industrielle, consommation manufacturière et exportations). La méthodologie utilisée dans ces deux modèles s'écartent notablement de celle des modèles à facteurs développés actuellement¹, introduits par Doz, Giannone et Reichlin (2006, 2007), comme par exemple ceux présentés pour la France dans Barhoumi, Darné et Ferrara (2010) et dans Bessec et Doz (2011), et pour la zone euro dans Bańbura et Rünstler (2011). Cependant en termes d'erreur quadratique moyenne, nos modèles s'avèrent plus précis que ceux de Barhoumi, Darné et Ferrara (2010) et Bessec et Doz (2011), ce qui ne nous incite pas à changer de méthodologie.

Nos facteurs sont calculés à partir d'un faible nombre de séries sélectionnées par l'algorithme de la régression LARS (Efron, Hastie, Johnstone et Tibshirani, 2004). L'algorithme classe les séries en fonction de la cible à prévoir (ici, la croissance du PIB) selon leur pouvoir prédictif. Les séries sont intégrées dans les facteurs avec leur éventuelle avance sur la croissance du PIB. C'est l'algorithme qui renseigne sur cette éventuelle avance. Par exemple, la variation de l'opinion sur l'activité prévue dans le bâtiment a une avance de trois trimestres sur la croissance du PIB et c'est avec cette avance que la série entrera dans les facteurs. Le passage par les facteurs est un moyen de régler le problème de colinéarité entre les variables, ce qui permet de garder un peu plus de variables que dans une régression ordinaire, afin de ne pas faire dépendre la prévision d'un faible nombre de prédicteurs.

Dans les modèles à facteurs développés actuellement, toutes les séries mensuelles disponibles entrent dans les facteurs, par exemple elles sont au nombre de 93 dans Bessec et Doz (2011). Il n'y a aucune pré-sélection des séries en fonction de la cible à prévoir et les facteurs ainsi calculés peuvent servir à prévoir n'importe quelle cible. Or un certain nombre d'articles ont souligné le fait qu'introduire des séries non directement reliées à la cible peut s'avérer nuisible parce que ces séries introduisent du bruit qui accroît la volatilité de la prévision (Boivin et Ng 2006, Bai et Ng 2008). Concernant la France, les résultats de Barhoumi, Darné et Ferrara (2010) vont dans ce sens : les meilleurs résultats sont obtenus avec des facteurs construits à partir de 20 séries plutôt qu'avec 51 séries ou qu'avec 140 séries². Deuxième différence avec notre méthode, les séries composant les facteurs sont introduites sans retard dans les facteurs. La dynamique des modèles à facteurs passe donc uniquement par les facteurs. Ainsi, on perd le caractère avancé ou coïncident de chaque série avec l'activité. Enfin cette dynamique n'est pas facile à déterminer car elle l'est empiriquement et non par des tests.

1. Comme il est décrit dans Charpin (2009).

2. Les auteurs choisissent ces 3 ensembles d'information auxquels ils appliquent 4 méthodes de détermination des facteurs.

Rappelons qu'un modèle à facteurs dynamiques peut être ramené à un modèle à facteurs statiques et qu'alors l'analyse en composantes principales permet d'obtenir une approximation des facteurs qui converge vers leur vraie valeur (Stock et Watson, 2002). C'est pourquoi Barhoumi, Darné et Ferrara (2010) comparent les modèles à facteurs statiques et dynamiques. Ils concluent que les meilleurs résultats sont obtenus avec des facteurs statiques à la Stock et Watson (2002)³. Notons que les erreurs quadratiques moyennes présentées sont élevées, au mieux 0,23 point, et ceci sur une période (2000 T1–2007 T4) d'avant la crise, et lorsque les données mensuelles du trimestre à prévoir sont connues sur le trimestre.

Signalons enfin que Bessec et Doz (2011) mentionnent qu'un des avantages de leur méthode est que les facteurs peuvent être déterminés jusqu'à la fin du mois T alors que certaines séries mensuelles ne sont disponibles que jusqu'en $(T-1)$ ou $(T-2)$. Il n'est donc pas nécessaire de recourir à des modèles auxiliaires pour prévoir les valeurs manquantes. Cette commodité indéniable a tout de même un prix : l'erreur quadratique moyenne de leur meilleur modèle s'accroît fortement dès qu'il manque 2 mois d'IPI et 1 mois de consommation manufacturière puisqu'elle passe de 0,19 en information complète⁴ à 0,34 avec les valeurs manquantes que l'on vient de citer⁵.

Les deux premières sections sont consacrées à la réévaluation des modèles à facteurs présentés en 2009, que nous désignons respectivement par *FH* (Factor Hard) pour celui qui comporte des variables réelles et par *FS* (Factor Soft) pour celui qui n'en comporte pas. Nous ne réévaluons leur performance que lorsqu'ils sont utilisés en « nowcasting », c'est-à-dire pour estimer le PIB au plus tôt 2 mois avant sa parution pour *FH* et 2 mois et demi avant pour *FS*. On connaît alors sur le trimestre à prévoir 1 mois des variables réelles et 2 mois des variables d'enquête. En section 3, dans le but de comparaison, nous évaluons les performances d'un « bridge model », noté *BH* (Bridge Hard), c'est-à-dire un modèle de régression dont les régresseurs ne sont plus des facteurs mais des séries individuelles en particulier des variables réelles. Enfin, en section 4, nous nous intéressons à un « ECM bridge model », noté *ECM-BH*, c'est-à-dire un modèle à correction d'erreurs basé sur une relation de cointégration entre le PIB et les éléments de la demande considérés dans tous nos modèles, à savoir la consommation manufacturière et les exportations.

3. Cela n'incite donc pas à estimer des facteurs selon la méthode de Doz, Giannone et Reichlin (2006, 2007).

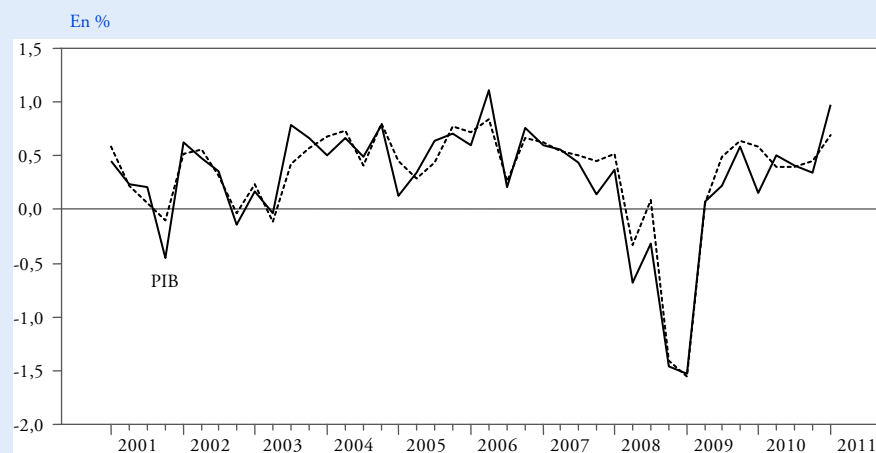
4. Nous entendons par là : lorsque les données mensuelles du trimestre à prévoir sont connues, c'est-à-dire juste avant la publication du PIB ou juste après celle des indices de la production industrielle du dernier mois du trimestre.

5. Dans leur travail, la période de calcul de l'erreur quadratique moyenne est 2000T1-2009T3. Dans notre cas, avec l'ancienne base des comptes et sur la période de calcul 2001T1-2009T3, l'erreur quadratique moyenne passe de 0,175 (en information complète) à 0,235 lorsque les mêmes valeurs manquent.

1. Réévaluation en nouvelle base du modèle à facteurs FH

La sélection des variables entrant dans la construction des facteurs n'a aucune raison d'être figée comme nous l'avons mentionné en 2009. On peut vérifier la pertinence de la sélection de temps à autre. Par exemple, concernant la zone euro et dans le cadre d'un travail pour Eurostat (Charpin et Mazzi, 2010), elle est revue une fois par an. La sélection qui figure dans le tableau 1 (Charpin, 2009) a donné de bons résultats jusqu'au deuxième trimestre 2009. À partir du troisième trimestre 2009, la forte surestimation de la croissance du PIB nous a amenés à revoir la sélection (Annexe, tableau A1). Sept variables sur quatorze ont été remplacées. Il s'agit de variables d'enquêtes car les quatre variables réelles sont toujours les mêmes (indices de la production industrielle et de la production manufacturière, consommation manufacturière et exportations). Depuis, cette nouvelle sélection est conservée. Nous n'avons pas revu cette sélection après l'introduction très récente de la nouvelle base. Mais les simulations en pseudo⁶ temps réel qui sont présentées par la suite sont réalisées avec la nouvelle série de PIB. En information complète, c'est-à-dire juste avant la parution du PIB⁷, la simulation (du premier trimestre 2001 jusqu'au premier trimestre 2011) est reportée sur le graphique 1. L'erreur quadratique moyenne est de 0,185 point. Elle s'est dégradée par rapport à celle présentée dans Charpin (2009). L'erreur moyenne est négative, mais non significativement au seuil 5 %, due à des surestimations en 2007 T4, 2008 T2, 2008 T3, 2009 T3 et 2010 T1. Les deux trimestres de récession 2008 T4 et 2009 T1 sont par contre correctement prévus.

Graphique 1 : Estimation de la croissance trimestrielle juste avant sa parution (modèle FH)



Source : INSEE et calculs OFCE.

6. Dans une analyse en « pseudo » temps réel, les séries sont tronquées pour ne tenir compte que des valeurs disponibles à la date où la prévision est calculée, mais les séries sont celles disponibles à la mi-mai 2011. Ainsi, les révisions des séries sont ignorées.

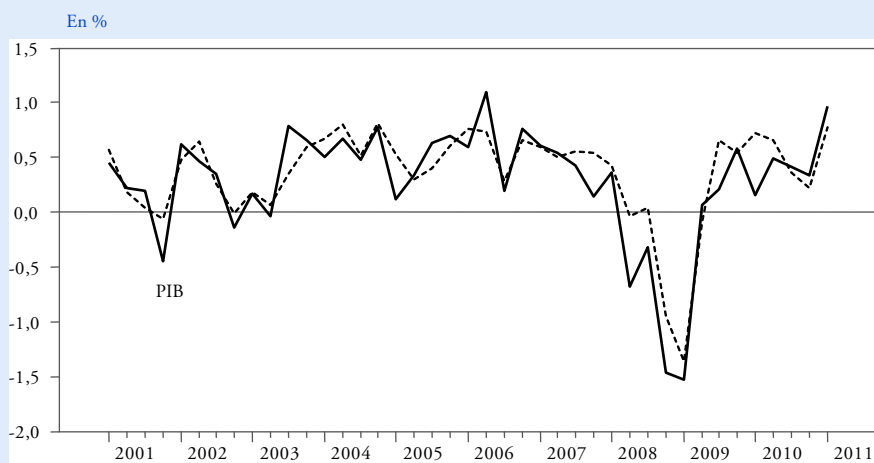
7. Après la sortie des IPI.

Nous allons maintenant examiner les erreurs quadratiques moyennes en information incomplète lorsqu'on dispose d'au moins 1 mois de données réelles du trimestre à prévoir, ce qui correspond *grosso modo* aux dates suivantes :

- (1) 15 jours avant la parution du PIB où il manque 1 mois des indices de production industrielle et d'exportations ;
- (2) 1 mois avant la parution du PIB où il manque 1 mois des indices de production industrielle, des exportations et de la consommation manufacturière ;
- (3) 1 mois et demi avant la parution du PIB où il manque 2 mois des indices de production industrielle, des exportations et 1 mois de la consommation manufacturière ;
- (4) 2 mois avant la parution du PIB où il manque 2 mois des indices de production industrielle, des exportations, de la consommation manufacturière et 1 mois d'enquête.

Pour prévoir les valeurs manquantes des variables mensuelles réelles, on utilise des régressions auxiliaires décrites dans l'article de Charpin (2009). Ces régressions comportent toujours 1 ou 2 termes autorégressifs et des soldes d'opinion. Elles ont un pouvoir prédictif très médiocre car les taux de croissance mensuels de ces grandeurs ont une forte volatilité et on n'en explique qu'environ 30 à 40 %. Concernant les données d'enquête, on a au plus 1 mois à prévoir dans le cas (4). Nous prenons alors comme valeur trimestrielle, la moyenne des deux premiers mois du trimestre. Les erreurs quadratiques moyennes sont reportées dans le tableau ci dessous qui donne les résultats obtenus avec tous nos modèles. Pour le modèle *FH*, elles passent de 0,195 à 0,25. Sur le graphique 2 sont reportées les prévisions faites 2 mois avant la parution du PIB (cas 4). Relativement au graphique 1, on note une dégradation des estimations en 2008 T2, 2008 T4, 2009 T3 et 2010 T1.

Graphique 2 : Estimation de la croissance trimestrielle 2 mois avant sa parution (modèle *FH*)



Source : INSEE et calculs OFCE.

Tableau : Erreurs quadratiques moyennes des estimations en pseudo temps réel selon le modèle, sur la période 2001 T1 – 2011 T1

En point				
Dates d'estimation	<i>FH</i>	<i>FS</i>	<i>BH</i>	<i>ECM-BH</i>
Juste avant la parution du PIB	0,185	0,24	0,20	0,18
(1) 15 jours avant la parution du PIB	0,195	0,24	0,22	0,20
(2) 1 mois avant la parution du PIB	0,195	0,24	0,22	0,205
(3) 1 mois et demi avant la parution du PIB	0,23	0,24	0,26	0,245
(4) 2 mois avant la parution du PIB	0,25	0,25	0,27	0,25

Source : Calculs OFCE.

2. Réévaluation en nouvelle base du modèle à facteurs *FS*

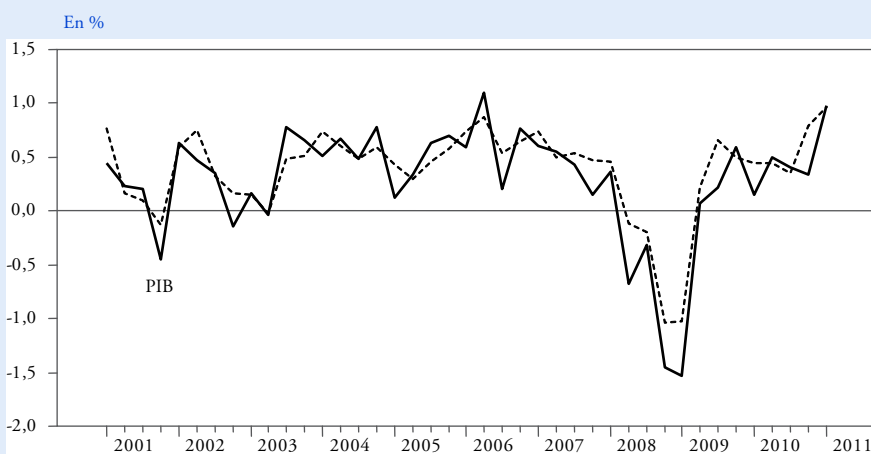
Les modèles ne contenant aucune variable réelle sont incapables de retracer la chute de la croissance de 2008 T4 et 2009 T1, comme un certain nombre de conjoncturistes l'ont constaté. Celui présenté dans l'article de Charpin (2009) n'échappe pas à la règle entraînant des surestimations de l'ordre de 0,8 point de la croissance pour ces deux trimestres. Nous avons donc cherché à améliorer ce modèle en examinant (après coup) si l'algorithme LARS permettait de sélectionner des soldes d'opinion plus aptes à décrire la récession. Rappelons que l'algorithme classe un ensemble de soldes d'opinion figurant sous plusieurs formes (niveau et variation) et avec plusieurs retards. À cet ensemble nous avons ajouté une nouvelle forme, non linéaire, du type $\Delta y' / |\Delta y|$, où Δy désigne la variation du solde d'opinion y ⁸. L'utilisation de ce type de variable est suggérée dans l'article de Cornec (2010). En examinant les sélections qu'on aurait pu obtenir depuis le troisième trimestre 2009, les dix variables données dans le tableau A2 de l'annexe sont retenues. Cette sélection contient trois variables non linéaires.

En information complète, c'est-à-dire 1 mois et demi avant la parution du PIB, la simulation (du premier trimestre 2001 jusqu'au premier trimestre 2011) est reportée sur le graphique 3. L'erreur quadratique moyenne est de 0,24 point. La surestimation des deux trimestres de récession a diminué mais reste forte (respectivement de 0,4 point en 2008 T4 et de 0,5 point en 2009 T1). Si, dans la liste du tableau A2, les variables $\Delta y' / |\Delta y|$ sont remplacées par Δy , la sous-estimation des deux trimestres de récession est plus forte (respectivement de 0,5 et 0,7) et l'erreur quadratique moyenne passe à 0,26 point. L'introduction de variables non linéaires a donc conduit à une petite amélioration. Enfin, avec seulement deux mois d'enquête sur le trimestre (soit 2 mois avant la parution du PIB), l'erreur quadratique moyenne est de 0,25 point (tableau 1) presque inchangée, et

8. Utiliser une telle variable dans une régression revient à dire que le coefficient de régression de Δy dépend de l'amplitude de la variation.

graphiquement on obtient des prévisions très voisines de celles reportées sur le graphique 3.

Graphique 3 : Estimation de la croissance trimestrielle 1 mois et demi avant sa parution (modèle FS)



Source : INSEE et calculs OFCE.

3. Un modèle simple *BH* serait-il aussi précis que le modèle *FH* ?

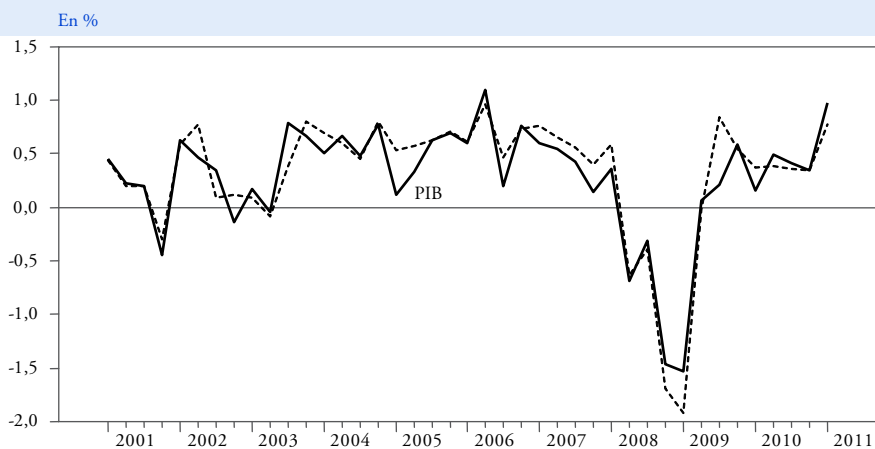
Cette réévaluation de modèles est aussi l'occasion d'en tester d'autres, en particulier de revenir à des modèles de régression plus simples où les régresseurs ne sont plus des facteurs mais des séries (« bridge model »). Pour estimer le modèle que nous aurions utilisé aux dates de prévision, nous examinons les classements des variables donnés par l'algorithme LARS sur diverses périodes d'estimation 1989-2000, 1989-2001, ..., 1989-2009. Puis nous retenons certaines variables figurant dans les 20 premières séries classées. La colinéarité entre les séries ne permet d'en retenir qu'un faible nombre parmi ces 20, mais fait aussi que plusieurs choix peuvent convenir pour une même période d'estimation. Ces modèles alternatifs, qui donnent des résultats équivalents sur la période d'estimation, ne donnent par contre pas nécessairement les mêmes résultats en prévision. C'est le grand défaut de ces modèles simples, la multiplicité des possibles, à période d'estimation donnée.

Les régresseurs retenus vont dépendre de la période d'estimation : un jeu pour la période 2001 T1–2006 T4, un autre, pour la période 2007 T1–2007 T4, puis un autre, pour la période 2008 T1–2009 T2, et enfin, un dernier, pour la période 2009 T3–2010 T4. Toutes ces régressions ont en commun 5 variables, dont 3 réelles (taux de croissance de l'indice de la production manufacturière, de la consommation manufacturière et des exportations en volume), 1 variable financière (le taux de croissance de l'indice boursier français⁹ retardé d'un

trimestre) et 1 solde d'opinion de l'enquête auprès des ménages (l'intention de faire des achats important, en variation et retardée de 2 trimestres). S'ajoute à cela 2 ou 3 soldes d'opinion qui varient selon la période. Dans la première période, il s'agit de l'indicateur synthétique dans les services (en niveau et en variation). Dans la deuxième, l'opinion sur l'évolution de la production prévue dans l'industrie manufacturière, l'opinion des ménages sur leur future capacité à épargner (retardée de 3 trimestres). Dans la troisième, l'opinion sur l'évolution de la production prévue dans l'industrie manufacturière (comme précédemment), l'opinion des ménages sur leur situation financière future. Dans la quatrième, on ajoute aux deux précédentes l'activité prévue dans l'enquête bâtiment (retardée de 3 trimestres). Remarquons que sur l'ensemble des périodes, le taux de croissance du prix réel du pétrole retardé de 4 trimestres est toujours significatif mais nous ne l'avons pas retenu. En effet il engendre une surestimation des prévisions sur la période de crise (durant laquelle le prix du pétrole baisse mais l'activité aussi, contrairement à l'effet qu'il a dans la régression).

En information complète, c'est-à-dire juste avant la parution du PIB, l'analyse en pseudo temps réel donne une erreur quadratique moyenne de 0,20 (tableau 1) donc plus élevée que le modèle à facteurs contenant des variables réelles. On a reporté sur le graphique 4 les estimations obtenues et on peut observer la différence avec le graphique 1 : 2008 T2 et T3 sont mieux estimés, mais 2008 T4, 2009 T1, et 2009 T3, nettement moins bien.

Graphique 4 : Estimation de la croissance trimestrielle juste avant sa parution (modèle BH)

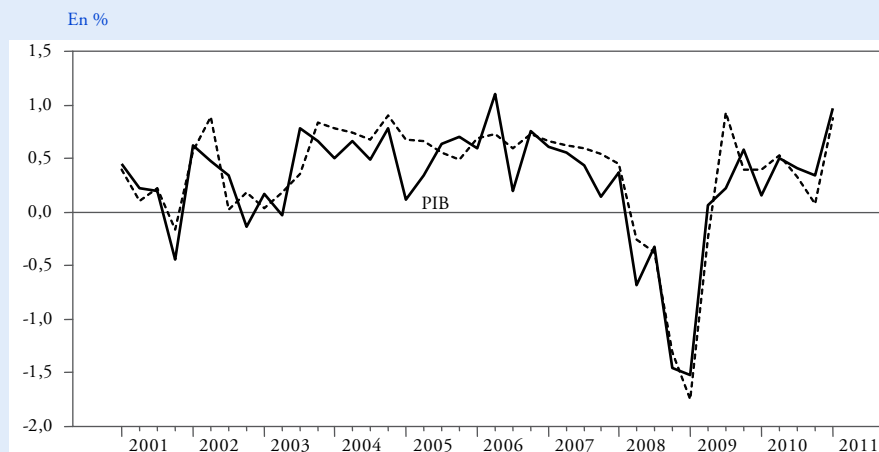


Source : INSEE et calculs OFCE.

9. Représenté par l'indice MSCI.

Les erreurs quadratiques moyennes lorsque l'on ne dispose pas de toute l'information passent de 0,22 à 0,27. Sur le graphique 5 sont reportées les prévisions faites 2 mois avant la parution du PIB.

Graphique 5 : Estimation de la croissance trimestrielle 2 mois avant sa parution (modèle BH)



Source : INSEE et calculs OFCE.

4. Un modèle à correction d'erreurs (ECM-BH)

On constate qu'entre le logarithme du PIB et le logarithme des deux variables réelles du modèle précédent, représentant une partie de la demande (la consommation manufacturière et les exportations), il existe une relation de cointégration. On peut chercher à exploiter cette relation d'équilibre en estimant un modèle à correction d'erreurs. Notons que la consommation manufacturière et les exportations sont connues avant le PIB, ce qui leur donne le statut de variables exogènes.

Nous avons trouvé un modèle à correction d'erreurs qui convient pour toutes les périodes d'estimation, ce qui n'était pas le cas du modèle précédent. La variation logarithmique du PIB est déterminée par (1) les logarithmes retardés d'un trimestre du PIB, de la consommation manufacturière et des exportations, (2) les variations logarithmiques de la consommation manufacturière, des exportations et l'indice de la production manufacturière, (3) la variation logarithmique de l'indice boursier français retardée d'un trimestre, (4) trois soldes d'opinion, à savoir, la variation de l'opinion sur l'évolution de la production prévue dans l'industrie manufacturière, la variation de l'opinion des ménages sur leur capacité à épargner¹⁰ retardée de deux

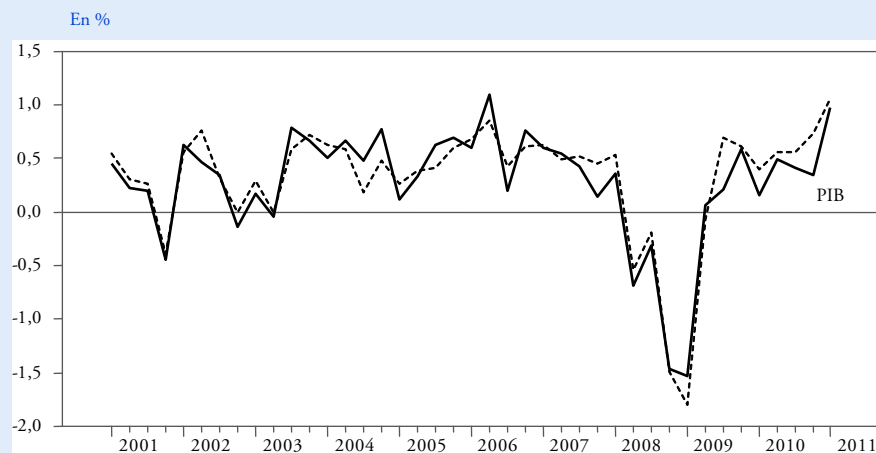
10. Il s'agit de la variable non linéaire associée.

trimestres et la variation de l'indicateur de confiance dans le bâtiment¹¹ retardée de deux trimestres.

On a vérifié qu'il y avait bien cointégration dans ce modèle à correction d'erreurs sur toutes les périodes d'estimation, la première étant 1989 T1–2000 T4, et la dernière 1989 T1–2010 T4. Pour cela, il faut tester la significativité du coefficient du logarithme retardé du PIB avec la table de Ericsson et MacKinnon (2002)¹².

En information complète, c'est-à-dire juste avant la parution du PIB, l'analyse en pseudo temps réel donne une erreur quadratique moyenne de 0,18 (tableau 1) donc comparable à celle du modèle à facteurs contenant des variables réelles. On a reporté sur le graphique 6 les estimations obtenues. À partir de 2009 T3, les prévisions sont toujours au-dessus de la réalité. Mais sur l'ensemble de la période il n'y a pas de biais significatif.

Graphique 6 : Estimation de la croissance trimestrielle juste avant sa parution (modèle *ECM-BH*)



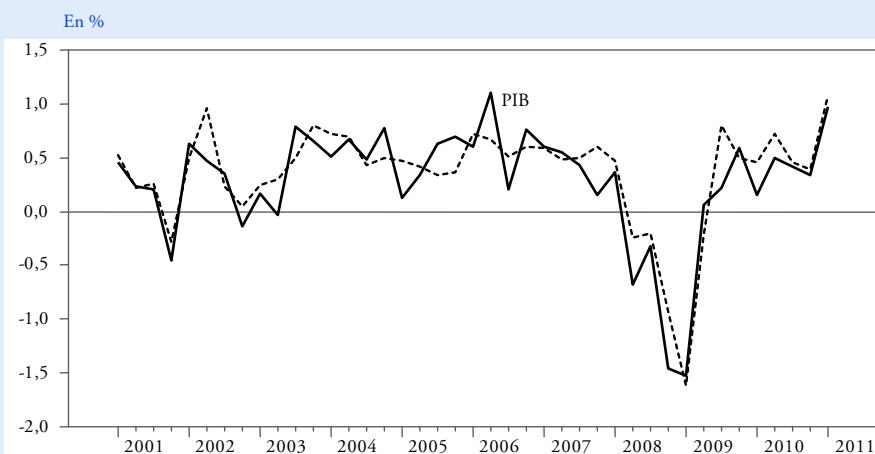
Source : INSEE et calculs OFCE.

Les erreurs quadratiques moyennes lorsque l'on ne dispose pas de toute l'information passent de 0,20 à 0,25. Sur le graphique 7, sont reportées les prévisions faites 2 mois avant la parution du PIB.

11. Enquête construction de la Commission européenne, puisque l'INSEE ne produit pas d'indicateur synthétique du bâtiment.

12. *Econometrics Journal*, vol. 5, pp. 285-318.

Graphique 7 : Estimation de la croissance trimestrielle 2 mois avant sa parution (modèle *ECM-BH*)



Source : INSEE et calculs OFCE.

■ Conclusion

L'estimation à très court terme de la croissance du PIB (« nowcast ») s'avère un exercice difficile. Même à la veille de la parution du PIB, les modèles sont assez peu précis. En effet, le contenu de nos modèles ne couvre pas tous les éléments de l'égalité ressources-emplois, en particulier, les variations de stock ne sont pas présentes dans l'information que nous utilisons. Il y a bien dans les enquêtes une question sur le niveau des stocks (dans l'industrie manufacturière, le commerce de détail et le bâtiment) mais ces soldes d'opinion ne permettent pas d'anticiper les variations de stock des comptes trimestriels. Ces variations ont par ailleurs été fortement revues dans la nouvelle base, non pas tant leurs fluctuations, mais le niveau général. Ainsi des variations précédemment négatives deviennent positives.

Nos deux meilleurs modèles contiennent trois variables réelles. L'un, à facteurs, est basé sur 14 variables, l'autre, à correction d'erreurs, sur 7 variables. Nous nous écartons de la tendance actuelle du « nowcasting » consistant à considérer un grand nombre de variables, mais nous constatons que c'est sans perte de précision.

Par rapport à notre article de 2009, deux éléments nouveaux ont été introduits : l'utilisation de variables non linéaires et l'exploitation d'une relation de cointégration.

Références bibliographiques

- Bai J. et S. Ng, 2008, « Forecasting Economic Time Series Using Targeted Predictors », *Journal of Econometrics*, 146, pp. 304-317.
- Bañbura M. et G. Rünstler, 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., O. Darné et L. Ferrara, 2010, « Are disaggregate data useful for factor analysis in forecasting French GDP ? », *Journal of Forecasting*, 29(1-2), pp. 132-144.
- Bessec M. et C. Doz, 2011, « Prévission de court terme de la croissance du PIB français à l'aide de modèles à facteurs dynamiques »,
- Boivin J. et S. Ng, 2006, « Are more data always better for factor analysis? », *Journal of Econometrics*, vol 132, issue 1, mai, pp. 169-194.
- Charpin F. et G. L. Mazzi, 2010, « Early estimate of the quarterly GDP growth of euro area », 30th CIRET Conference, New York.
- Charpin F., 2009, « Estimation précoce de la croissance: de la regression LARS au modèle à facteurs », *Revue de l'OFCE*, n° 108, pp. 31-48.
- Cornec M., 2010, « Constructing a conditional GDP fan chart with an application to French business survey data », 30th CIRET Conference, New York.
- Doz, C., D. Giannone et L. Reichlin, 2006, « A quasi maximum likelihood approach for large approximate dynamic factor models », *WP ECB*, n° 674, à paraître dans *Review of Economics and Statistics*.
- Doz, C., D. Giannone et L. Reichlin, 2007, « A two-step estimator for large approximate dynamic factor models based on Kalman filtering », CEPR discussion paper, n° 5724, à paraître dans *Journal of Econometrics*.
- Efron B., T. Hastie I. Johnstone et R. Tibshirani, 2004, « Least Angle Regression », *Annals of Statistics* 32:2, pp. 407-499.
- Stock J.H. et M.W. Watson, 2002, « Forecasting Using Principal Components from a Large Number of Predictors », *Journal of the American Statistical Association*, 97, pp. 1167-1179.

ANNEXE

Tableau A1 : Les séries entrant dans les facteurs à partir de 2009 T3

Définition des séries et source	Transformation	Avance (en trimestres)
Indice de la production manufacturière (En volume, Insee)	Taux de croissance	0
Indice de la production industrielle (En volume, Insee)	Taux de croissance	0
Exportations (Service des douanes, calcul OFCE pour le volume)	Taux de croissance	0
Consommation en produits manufacturés (En volume, Insee)	Taux de croissance	0
Opinion sur l'évolution de la production prévue (Enquête industrie manufacturière, Insee)	Niveau	0
Opinion sur l'évolution de la production prévue (Enquête industrie manufacturière, Insee)	Variation	0
Opinion sur le niveau des stocks (Enquête industrie, Insee)	Niveau	0
Indicateur de confiance (Enquête industrie manufacturière, Insee)	Niveau	0
Opinion des ménages sur leur situation financière future (Enquête auprès des ménages, Insee)	Niveau	0
Opinion des ménages sur la possibilité d'épargner (Enquête auprès des ménages, Insee)	Niveau	3
Opinion sur les effectifs prévus dans le bâtiment (Enquête industrie du bâtiment, Insee)	Variation	1
Indicateur de confiance (Enquête construction, CE)	Variation	1
Indicateur de confiance (Enquête construction, CE)	Variation	2
Opinion sur l'activité prévue dans le bâtiment (Enquête industrie du bâtiment, Insee)	Variation	3

Source : Insee et Commission européenne.

Tableau A2 : Les séries entrant dans les facteurs du modèle FS

Définition des séries et source	Transformation	Avance (en trimestres)
Opinion sur l'évolution de la production prévue (Enquête industrie manufacturière, Insee)	Niveau	0
Opinion sur l'évolution de la production prévue (Enquête industrie manufacturière, Insee)	Variation	0
Indicateur de confiance (Enquête industrie manufacturière, Insee)	Niveau	0
Indicateur synthétique (Enquête industrie manufacturière, Insee)	Variation	0
Opinion des ménages sur leur situation financière future (Enquête auprès des ménages, Insee)	Niveau	0
Indicateur synthétique (Enquête dans les services, Insee)	Variation	0
Opinion des ménages sur la possibilité d'épargner (Enquête auprès des ménages, Insee)	Niveau	3
Opinion sur les effectifs prévus dans le bâtiment (Enquête industrie du bâtiment, Insee)	Variation non linéaire $\Delta y' \Delta y $	1
Opinion sur l'activité passée dans le bâtiment (Enquête industrie du bâtiment, Insee)	Variation non linéaire $\Delta y' \Delta y $	2
Opinion sur les carnets de commande dans le bâtiment (Enquête dans l'industrie du bâtiment, Insee)	Variation non linéaire $\Delta y' \Delta y $	2

Source : Insee.