



Document de travail

ÉCONOMÉTRIE DE LA PRÉVISION

N° 2005-11
Juillet 2005

Guillaume CHEVILLON

Département analyse et prévision, OFCE
Economics Department, University of Oxford

E-mail : guillaume.chevillon@sciences-po.fr ; téléphone : + 33 1 44 18 54 36. Je remercie Xavier Timbeau et David Hendry pour leurs remarques et suggestions.

Observatoire Français des Conjonctures Économiques

69, Quai d'Orsay 75340 Paris Cedex 07

Tel : 01 44 18 54 00 Fax : 01 45 56 06 15

E-mail: ofce@ofce.sciences-po.fr Web: <http://www.ofce.sciences-po.fr>

Résumé

Cet article présente les résultats récents de l'approche économétrique de la prévision économique. Il s'agit, ici, de déterminer ce qu'on appelle une "bonne" prévision. Nous dérivons une taxinomie des erreurs de prévision afin de comprendre comment générer des prédicteurs robustes vis-à-vis des sources principales d'erreur. Ceci nous permet de dissocier les concepts d'exactitude, de précision et de certitude dans le cadre de modèles dédiés à la prévision. Nous montrons que le critère d'évaluation de l'exactitude est un élément essentiel qu'on ne peut séparer de la construction du modèle. Une application à la prévision des importations françaises de biens et services illustre notre propos.

Mots-clef : Prévision, Critères d'évaluation, Ruptures déterministes.

Codes **JEL** : C32, C5, C53.

Prévoir, c'est établir une déclaration sur les événements à venir en utilisant comme outils le passé et le présent. Pour réussir une telle entreprise, quatre conditions sont à remplir : (1) une certaine régularité dans le fonctionnement du monde est tout d'abord nécessaire, (2) il faut par ailleurs que cette régularité fournisse une information sur l'avenir, (3) que la méthode choisie pour établir la prévision capture une part de cette régularité et (4) enfin qu'elle exclue toutes ou une partie des irrégularités. Les deux premières conditions sont constitutives du système économique, les deux dernières sont fonctions de la méthode de prévision choisie. Il en résulte que les prévisions sont toujours entachées d'erreur et qu'il est possible d'en établir plusieurs pour un même événement à venir (qui constitue l'*objet* de la prévision). Des outils d'analyse doivent donc être développés afin de comparer et de hiérarchiser les prévisions pour discerner ce qui fait qu'on puisse, ou non, en qualifier certaines de "bonnes".

La *qualité* d'une prévision est définie par le critère d'évaluation considéré. Ainsi, si on s'intéresse à l'inflation en France en 2006, il est probable que le prévisionniste soit indifférent entre une erreur de +0,1 % et de -0,1%, i.e. symétrique autour de l'hypothèse centrale. En revanche le sismologue qui s'attache à déterminer la date de la prochaine éruption de l'Etna accordera une valeur nettement (infiniment ?) plus grande à une prévision en-deçà de la date d'éruption ! Il apparaît clair que, dans ces deux contextes, les méthodes choisies se ressentiront des critères retenus pour leur évaluation.

De manière générale, trois qualités sont recherchées : l'*exactitude*, la *précision* et la *certitude*, et on privilégiera l'une ou l'autre selon les cas. Pour bien comprendre leur différence, si nous supposons que l'inflation se révélera en France de 1,95% en 2006 et qu'on dispose en 2005 de deux prévisions de celle-ci (le trait est un peu forcé pour la clarté de l'exposition) : une à $1,935 \pm 6,020$ % et l'autre à $2,0 \pm 0,2$ %. Laquelle préférer ? La première est extrêmement *précise* dans son énoncé, au millième de "pourcent" près ! La seconde l'est nettement moins, à un dixième de pourcent. En revanche, cette dernière est nettement plus *certaine*, car l'intervalle de confiance lui correspondant est assez étroit (en tout 0,4%) contre une incertitude de 12,04% dans le premier cas. Enfin, la première prévision est finalement plus proche – dans son scénario central – de la réalisation. Mais laquelle peut être dite la plus exacte ? Ceci dépend du critère considéré – et donc de l'utilisation faite de cette prévision. Et encore ! Que dire des deux déclarations "l'inflation se portera à 12,314 % en 2006" et "l'inflation sera d'environ 2%". La première est très précise et certaine (du moins dans son assertion, ce qui la rend presque péremptoire), mais très éloignée de la vérité, l'autre est moins précise et certaine, mais plus proche de la réalisation. Toutes ces déclarations peuvent pourtant être qualifiées d'inexactes ! Pour mesurer et hiérarchiser l'exactitude, les économistes utilisent souvent le critère de moyenne quadratique d'erreur de prévision (MQEP, voir définition plus bas) qui combine l'écart à la réalisation (la moyenne d'erreur de prévision) et l'incertitude entourant le scénario central (la variance de l'erreur).

Ces quelques réflexions nous amènent à nous demander comment quantifier a priori l'exactitude d'une prévision. Pour ce faire, des informations sont nécessaires : comment est-elle obtenue ? sous quelle hypothèses ? quelle incertitude envisage-t-elle ? quel est son but, son utilité ? Avec tous ces éléments, il est possible d'analyser une prévision, et c'est

le propos de cet article de montrer comment les outils de l'économétrie permettent d'en établir et d'en contrôler les caractéristiques.

Cet article s'organise comme suit : la première section présente un bref historique des méthodes utilisées en prévision économique afin de dégager ensuite, dans le cadre d'un modèle simple, les principaux apports de l'analyse économétrique, pour grande part tirés des travaux de M. P. Clements et D. F. Hendry. La section 3 en montre toutefois les limites, dont on apporte les solutions dans la partie suivante. Après s'être intéressé aux méthodes d'évaluation des prévisions dans la section 5, nous étudions l'application de ces méthodes à la prévision des importations françaises.

1 La prévision en économie

Pour comprendre les mérites propres des méthodes de prévision, il est essentiel d'en connaître les deux grandes classes, d'ailleurs non mutuellement exclusives : il existe d'une part des modèles purement *statistiques* sans représentation économique et d'autre part des modèles dits *structurels* dont l'ambition est de fournir une image du fonctionnement effectif de l'économie.

Mais afin de bien présenter l'état actuel de la recherche, il nous semble essentiel de rappeler brièvement les développements de la prévision économique au XXème siècle, car chaque grande vague d'avancées s'est brisée contre un écueil qui a permis d'améliorer la compréhension des limitations des diverses méthodes.¹

1.1 Modèles structurels keynésiens

La complexité et l'inter-dépendance des systèmes économiques a longtemps été un frein puissant à toute velléité prévisionnelle, ce qui fit que les premières tentatives ne virent le jour qu'après l'émergence de modèles macroéconomiques simples. Après la publication de la Théorie Générale de J. M. Keynes en 1936 et grâce à de notables avancées techniques de la Statistique, s'ensuivit un vif développement de l'activité économique de construction, d'estimation et d'analyse de modèles visant à représenter le comportement des agents macroéconomiques. Ce bouillonnement intellectuel amena à la création de l'Econometric Society et de sa publication prestigieuse, la revue *Econometrica*,² qui devint le médium naturel de diffusion des idées et avancées de la Commission Cowles à l'université de Chicago au cours des années 1940 et 1950. La facilité d'usage des systèmes d'équations à base de modèles IS-LM et comportant un agent unique pourvu de rationalité dans ses anticipations permit leur généralisation dans les cercles universitaires et gouvernementaux s'attachant à l'analyse et à la prévision des politiques économiques. Les méthodes d'estimation économétrique étant encore peu développées, les principaux débats portaient

¹L'exposition historique doit en grande part à l'article très clair de F. X. Diebold (1998) auquel on réfère vivement le lecteur.

²A l'origine, les deux branches de l'économie mathématique et de l'économétrie ne s'étaient pas encore divisées.

sur des calibrages des coefficients de réaction. Les besoins naissants de planification entraînent la généralisation de modèles input-output qui visaient à prévoir les tensions sur les capacités productives. L'augmentation de la quantité et de la qualité des données disponibles mena à une généralisation de modèles très détaillés de l'économie, comportant des centaines d'équations. Cette abondance d'information ne put empêcher un enlèvement des macro-modèles (comme le fameux modèle Wharton conçu autour de Lawrence Klein) car ils subirent un cuisant échec dans les années 1970 alors que la stagflation apparaissait et que le chômage prenait son envol.

L'apparition de doutes quant au manque de fondements micro-économiques – i.e. représentant les préférences, comportements et anticipations des agents – aboutirent à la remise en cause théorique des “systèmes d'équations”. La critique de la raison pratique vint en l'occurrence de Lucas (1976) qui remit en cause la stabilité des paramètres des modèles non “structurels”. Pour lui, il fallait rechercher les paramètres fondamentaux du comportement économique, ceux qui se révèlent stables. Cette approche tend à privilégier l'élaboration de modèles de faible taille, plus robustes aux chocs et changements de politique économique.

Par ailleurs, les capacités prévisionnelles des macro-modèles se virent bien souvent dépassées par de simples extrapolations statistiques. La macroéconomie purement keynésienne précéda de peu la prévision économétrique structurelle keynésienne dans sa perte de faveur. Une dissociation s'opéra entre les approches de modélisation (pour étudier la réaction de l'économie à une politique) et de prévision (où il s'agit d'obtenir la valeur future, assortie d'un intervalle d'incertitude, d'une ou plusieurs variables).

1.2 Modèles statistiques

A la suite de la baisse d'intérêt pour l'utilisation de modèles théoriques en prévision, l'une des réponses fut une réorientation vers des méthodes alternatives plus empiriques. Un modèle non-structurel élabore des prévisions dites inconditionnelles, i.e. qui s'attachent à déterminer l'évolution de l'économie en supposant que les politiques ne subissent aucune altération – contournant ainsi la critique de Lucas qui niait le pouvoir prévisionnel des modèles économiques en fondant son analyse sur leur manque de *robustesse* (de stabilité) vis-à-vis d'une intervention de politique économique. – alors que les modèles structurels établissent des prévisions conditionnées aux décisions de politique économique. Cette propriété même, conjuguée à la faiblesse des modèles structurels et aux avancées importantes des techniques statistiques a entraîné un fort intérêt pour ces nouvelles méthodes de prévision.

En réalité, les développements techniques étaient pour bonne part déjà connus et utilisés par les économètres car ils dataient d'avant l'émergence de la théorie keynésienne ; la borne initiale fut plantée dans les années 1920 par Slutsky et Yule, lorsque ceux-ci constatèrent que des équations différentielles linéaires, gouvernées par des chocs aléatoires constituaient un cadre simple et puissant pour la représentation et la prévision des variables économiques et financières. Ces équations différentielles (discrètes) stochastiques, appelées autorégressives, permettent d'exprimer une variable à l'instant présent comme la somme pondérée de ses valeurs passées et d'un choc purement aléatoire. L'exemple le plus

simple en est le modèle autorégressif d'ordre 1, noté AR(1), où la variable y à l'instant t , y_t , est donnée par :

$$y_t = \alpha y_{t-1} + \varepsilon_t,$$

où ε_t suit une loi aléatoire (en général Gaussienne) de moyenne nulle et où les propriétés de y_t dépendent de la valeur α . Les processus autorégressifs sont liés de manière proche à ceux dits de moyenne mobile, étudiés aussi par Slutsky et Yule, où, à présent, y_t est exprimé comme la somme pondérée de chocs présent et passés. Le cas non trivial le plus simple en est la moyenne mobile d'ordre 1 (MA(1) pour "moving average") :

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}.$$

Les représentations AR et MA sont, pour ce que est des variables stationnaires³, les deux facettes d'une même dynamique et il est possible de passer de l'une à l'autre, mais au prix d'une perte d'efficacité dans l'estimation. Le très fort développement des modèles du type Slutsky-Yule doit son origine aux travaux, dans les années 1930, de H. Wold qui montra alors que – sous des hypothèses de stationnarité – toutes les séries temporelles peuvent être représentés comme modèles autorégressifs ou de moyenne mobile. Les algorithmes de prévision, d'estimation et de représentation de ces modèles furent développés dans les décennies 1940 et 50 par Wiener, Kolmogorov et Kalman⁴, époque de forte avancée de l'économétrie.

L'heure de la désillusion vis-à-vis des modèles keynésiens coïncida avec la publication par Box et Jenkins du livre référence en matière d'analyse et de prévision des séries temporelles. Une des idées phares de ces auteurs fut l'apport d'une nouvelle méthode pour représenter les tendances. A la place d'une fonction linéaire du temps, ils suggérèrent le concept de tendance stochastique qui est générée par la somme des chocs passés, conférant, de ce fait, à chacun d'entre eux un impact permanent. En termes de prévision, ceci implique une absence de retour à long terme vers une tendance précise, mais, au contraire, la présence à chaque instant d'une tendance locale fluctuante. Les variables présentant un comportement de tendance stochastique sont non-stationnaires et dites intégrées (on parle aussi de présence d'une racine unitaire). La contribution principale de Box et de Jenkins est, toutefois, l'élaboration d'un cadre opérationnel pour la prévision grâce à la clarification d'une méthode de formulation, d'estimation, d'évaluation d'un modèle et de prévision des séries qui reprenne les travaux de Slutsky-Yule et définisse les modèles ARMA, présentant conjointement des propriétés autoregressives et de moyenne mobile.

Les travaux de Box–Jenkins, s'ils ont focalisé essentiellement sur des modèles univariés, ont en revanche préparé un rapprochement de l'économétrie et de l'analyse statistique à la suite de Sims qui proposa au début des années 1980 une nouvelle méthode de

³La stationnarité au sens faible de y_t implique que l'espérance mathématique et la variance de y_t soient indépendantes de la date t .

⁴Pour une présentation des modèles espaces-états et du filtre de Kalman, voir Lemoine & Pelgrin (2003).

représentation des interactions entre variables prenant la forme d'une représentation autorégressive d'un vecteur de séries temporelles inter-dépendantes (i.e. sans différenciation d'endogènes et d'exogènes) : les VAR (pour Vector Auto-Regression) qui permettent, entre autres, d'observer la propagation de chocs dans l'économie (méthode dite de réponse à l'impulsion). Prenons par exemple le cas de deux variables x_t et y_t représentées par un VAR d'ordre 2 (i.e. qui implique des retards des variables jusqu'à $t - 2$, ce qu'on note VAR(2)) et donné par :

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 1/4 & 1/3 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1/5 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{bmatrix},$$

soit

$$\begin{aligned} x_t &= \frac{1}{2}x_{t-1} + \frac{1}{5}x_{t-2} + \varepsilon_{x,t}, \\ y_t &= \frac{1}{4}x_{t-1} + \frac{1}{3}y_{t-1} + \varepsilon_{y,t}. \end{aligned}$$

Ainsi un choc unitaire sur $\varepsilon_{x,t}$ se transmet-il dans l'économie suivant le schéma (on suppose que toutes les autres variables sont nulles pour dissocier l'impact spécifique de ce choc) :

$$\begin{aligned} \text{à } t & : \varepsilon_{x,t} = 1 \rightarrow x_t = 1 \\ \text{à } t + 1 & : x_{t+1} = \frac{1}{2} \text{ et } y_{t+1} = \frac{1}{4} \\ \text{à } t + 2 & : x_{t+2} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{5} = \frac{9}{20} \text{ et } y_{t+2} = \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{4} = \frac{5}{24} \\ \text{pour } k \rightarrow \infty & : x_{t+k} \rightarrow 0 \text{ et } y_{t+k} \rightarrow 0 \end{aligned}$$

L'effet du choc sur $\varepsilon_{x,t}$ équivaut à une variation identique de x_t et l'impulsion se résorbe et les variables retournent vers leur état initial, ainsi que le montre la figure 1.

Depuis les années 1980, de nombreuses méthodes ont été développées sur les fondements des processus VAR. Celles-ci comprennent entre autres le concept de cointégration selon lequel il existe des combinaisons linéaires de variables non-stationnaires qui forment de nouvelles variables stationnaires. Le but des sections suivantes est de montrer de manière simple comment ces méthodes économétriques permettent d'appréhender les problèmes qui se présentent au prévisionniste.

2 Les modèles économétriques de prévision

2.1 Un monde parfait

Notre but ici est d'expliquer les échecs récurrents des prévisions et, de ce fait de tâcher de formuler des modèles *robustes*, i.e. insensibles à certains défauts. Les premiers modèles de prévisions reposaient sur deux hypothèses fondamentales :

1. Le modèle utilisé constitue une représentation exacte de l'économie ;
2. La structure de l'économie a été stable et le restera,

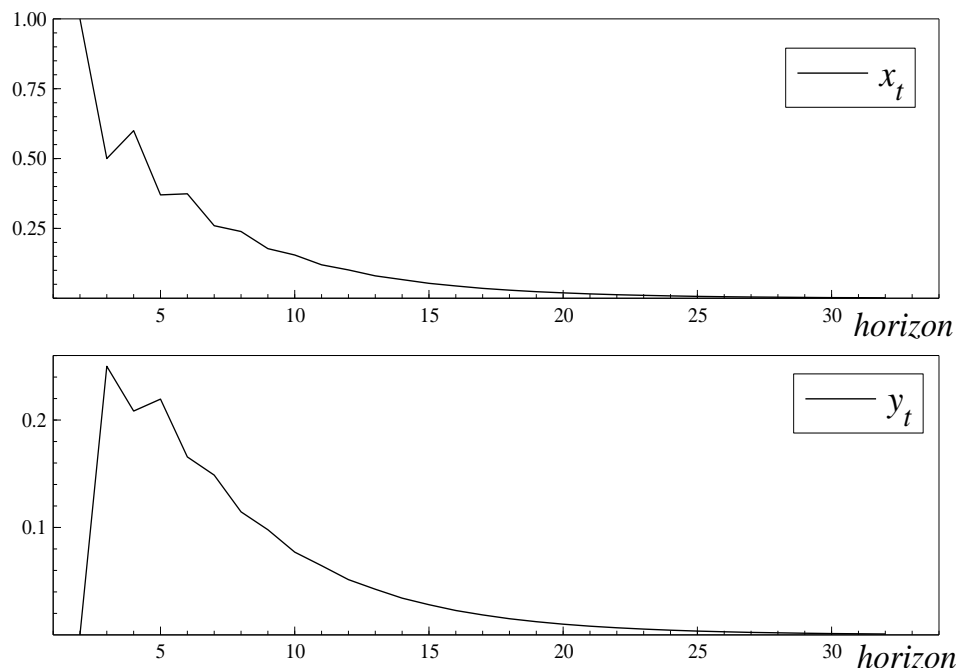


Figure 1 – Réponse d’un système de deux équations à une impulsion unitaire sur x_t et sa propagation dans le système à mesure que l’horizon progresse.

qui fondent ce qu’on appelle à présent la *théorie de l’optimalité*⁵. Prenons l’exemple simple d’un modèle autoregressif d’ordre 1, dit AR(1), où la valeur d’une variable y à l’instant t , y_t , est fonction de sa valeur à la période précédente, y_{t-1} , à laquelle s’ajoute une erreur imprévisible (qui suit une loi Normale d’espérance nulle et de variance σ_ε^2) :

$$y_t = \alpha y_{t-1} + \varepsilon_t, \quad \text{où } \varepsilon_t \sim \mathbf{N}(0, \sigma_\varepsilon^2), \quad (1)$$

avec $|\alpha| < 1$ pour que y_t soit stationnaire (i.e. de moyenne et variance constante). Ceci fournit, pour $h \geq 1$:

$$y_{t+h} = \alpha^h y_t + \sum_{i=0}^{h-1} \alpha^i \varepsilon_{t+h-i},$$

et, notant $\mathbf{E}[\cdot|y_t]$ l’espérance mathématique, sachant y_t , il peut sembler naturel d’utiliser comme prévision de y_{t+h} générée à l’instant t , la valeur $\hat{y}_{t+h|t}$ donnée par :

$$\hat{y}_{t+h|t} = \mathbf{E}[y_{t+h}|y_t] = \alpha^h y_t.$$

La propriété essentielle de l’espérance (conditionnelle) est qu’elle est – selon les hypothèses 1 et 2 ci-dessus – parmi les prévisions $\hat{y}_{t+h|t}$ de y_{t+h} à l’instant t , celle qui minimise le

⁵voir Clements & Hendry (1998) et Makridakis & Hibon (2000).

critère de moyenne quadratique d'erreur de prédiction, défini par :

$$\text{MQEP}_h(\hat{y}_{t+h|t}) = \mathbb{E} \left[(y_{t+h} - \hat{y}_{t+h|t})^2 | y_t \right] = \left(\mathbb{E} [y_{t+h} - \hat{y}_{t+h|t} | y_t] \right)^2 + \text{Var} [y_{t+h} - \hat{y}_{t+h|t} | y_t].$$

En effet pour toute autre prédiction $\tilde{y}_{t+h|t}$:

$$\text{MQEP}_h(\tilde{y}_{t+h|t}) = \text{MQEP}_h(\hat{y}_{t+h|t}) + \mathbb{E} \left[(\hat{y}_{t+h|t} - \tilde{y}_{t+h|t})^2 | y_t \right]$$

et le deuxième terme de cette équation est positif (et nul si et seulement si $\hat{y}_{t+h|t} = \tilde{y}_{t+h|t}$). Dans un tel contexte, le modèle le plus "exact" (ici, dont les paramètres estimés sont les plus précis) fournira les meilleures prévisions. Par extension, selon la théorie de l'optimalité, le modèle qui reflète le mieux les propriétés des variables dans l'échantillon de données dont on dispose fournira les prévisions les plus justes. Par ailleurs, quand on connaît la manière dont sont générées les données, les seules informations qui puissent permettre d'améliorer la précision sont celles concernant des faits à venir (tels que des changements de politique économique, de taux d'intérêts...)

La justesse des prévisions est aussi, dans ce cas, fonction décroissante de l'horizon h car la somme des chocs futurs (les ε_{t+h}) augmente la variance de l'erreur de prédiction (qui converge vers une valeur limite si y_t est stationnaire). Enfin, il est possible de quantifier l'erreur de prédiction car $\text{MQEP}_{h=1} = \text{Var} [\varepsilon_t] = \sigma_\varepsilon^2$. Le prévisionniste serait donc à même de fournir un intervalle de confiance sûr autour de son hypothèse centrale : dans le cadre d'innovations ε_t suivant une loi Normale, un intervalle de confiance à 95% autour de la prédiction s'écrit

$$y_{t+h} \underset{95\%}{\in} \left[\hat{y}_{t+h|t} - 1.96 \sqrt{\text{MQEP}_h(\hat{y}_{t+h|t})}, \hat{y}_{t+h|t} + 1.96 \sqrt{\text{MQEP}_h(\hat{y}_{t+h|t})} \right],$$

où dans le cadre du modèle AR(1), $y_{t+h} - \hat{y}_{t+h|t} = \sum_{i=0}^{h-1} \alpha^i \varepsilon_{t+h-i}$ et donc

$$\text{MQEP}_h(\hat{y}_{t+h|t}) = \text{Var} \left(\sum_{i=0}^{h-1} \alpha^i \varepsilon_{t+h-i} \right) = \frac{1 - \alpha^h}{1 - \alpha} \sigma_\varepsilon^2$$

est bornée (et ceci est toujours le cas lors de la prédiction de variables stationnaires) puisque :

$$\text{MQEP}_h(\hat{y}_{t+h|t}) \underset{h \rightarrow \infty}{\rightarrow} \frac{\sigma_\varepsilon^2}{1 - \alpha}.$$

Le graphique 2 présente une simulation de prédiction sur la base du modèle AR(1) ci-dessus estimé pour 80 observations : la MQEP_h permet d'établir l'incertitude autour de la prédiction et on constate que sur cet exemple, y_{90} sort de l'intervalle de confiance à 95%.

Malheureusement, ces hypothèses ne sont pas vérifiées généralement. Il est possible de fournir de multiples contre-exemples pour chacune des propriétés présentées ci-dessus. C'est pour cette raison que de nouvelles théories se sont révélées nécessaires et ont vu le jour au cours des années 1990.

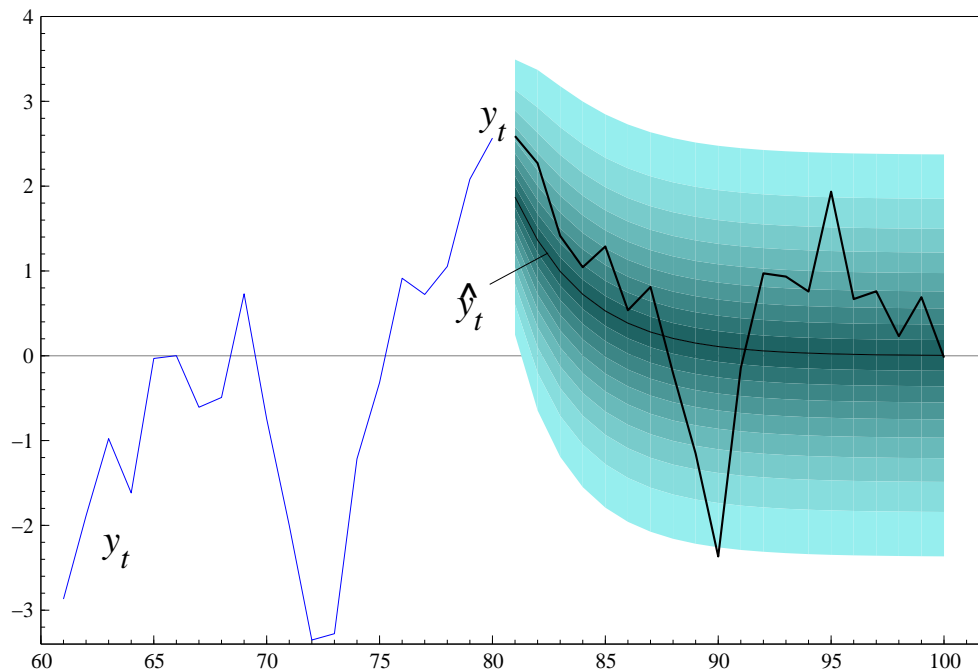


Figure 2 – Donnée simulée y_t suivant un AR(1) et sa prévision \hat{y}_t à partir de la 80ème observation. Les bandes autour de la prévision indique des incertitudes croissantes de 5% en 5%.

2.2 Un cadre pour la prévision

Six facettes sont à prendre en compte dans l'élaboration d'un modèle :

1. la nature du processus probabilistique qui génère les données (noté PGD),
2. le niveau de connaissance à son sujet,
3. la dimension du système,
4. la forme d'analyse,
5. l'horizon de prévision et
6. la complexité mathématique du système.

Ainsi, le PGD peut il être stationnaire, ou alors peut-être ramené à un processus stationnaire (par exemple en utilisant des relations de co-intégration), ou enfin non-stationnaire et sujet à évolutions (comme lorsque les paramètres changent, ou s'il passe par divers régimes...). On peut par ailleurs le connaître de manière sûre (et se retrouver dans le cadre présenté plus haut de la théorie de l'optimalité), ou n'avoir qu'à estimer ses paramètres car on connaît sa forme générale, ou enfin ignorer à la fois la forme et les paramètres du PGD. Le système peut être univarié (scalaire), ou peut constituer un processus vectoriel estimable ou non (auquel cas des hypothèses d'identification sont nécessaires a priori). Il peut s'agir d'une analyse sur des propriétés asymptotiques (à partir de très

grands échantillons) ou, au contraire, se fondant sur une étude d'échantillons de taille finie et en faisant, le cas échéant, appel à des simulations. Il est possible qu'on ne s'intéresse qu'à la prochaine valeur ou à un horizon de prévision s'étendant sur plusieurs périodes. Et enfin, le système étudié peut être linéaire ou non.

Toutes ces hypothèses ont leur importance car les sources d'erreurs résultant du modèle en dépendent ; par ailleurs, la méthode de prévision doit être définie pour pouvoir traiter ces problèmes. Ainsi, au vu des avancées récentes en économie et économétrie, on doit naturellement se placer dans le cadre d'un PGD vectoriel non-stationnaire inconnu mais estimable, soumis à des modifications non-anticipables. On suppose cependant qu'il est possible de se ramener à un modèle linéaire, même si on ne dispose que d'un échantillon de taille réduite et qu'on s'intéresse à des prévisions à divers horizons. Ce type de PGD est un des plus généraux pour lesquels il est possible d'établir une analyse en matière de prévision. Dans la section 6, on s'intéressera à un modélisation des importations françaises de biens et services : il s'agit d'une représentation de leur taux de croissance issue d'un modèle plus complexe prenant en compte l'ensemble des interactions ; ici la croissance des importations est fonction des variations récentes de variables exogènes et d'un déséquilibre à la période précédente entre la demande intérieure, les goulets d'étranglements et une tendance déterministe représentant l'ouverture croissante de l'économie française.

3 Limitations des modèles

Pour comprendre les difficultés rencontrées par les modèles économétriques lorsqu'ils sont appliqués à la prévision, il est utile de décomposer les sources principales d'erreurs selon leur origine afin d'en établir une taxinomie et, ensuite, de les hiérarchiser. Il nous faut dans ce but spécifier un processus probabilistique ayant généré les données, et un modèle de celui-ci – a priori distinct – utilisé en prévision. En macroéconomie, le consensus des vingt dernières années s'est établi autour de données intégrées générées par un système co-intégré et dont les coefficients sont soumis à des variations irrégulières. Le modèle de prévision est souvent formulé en fonction de notions (*d'a priori*) théoriques et il est très certainement empreint d'erreurs de spécification qui peuvent aussi entraîner une mauvaise estimation des paramètres à partir de données potentiellement erronées. Afin d'établir un cadre d'analyse clair, on suppose que le modèle comprend trois composantes : des termes déterministes (des constantes et des tendances linéaires du type $\alpha + \beta t$) dont les valeurs futures sont connues ; des variables stochastiques observées (celles que le modèle tente de caractériser) dont les valeurs futures sont inconnues, et des variables non observées et dont les valeurs passées, présentes et à venir sont inconnues (elles représentent les erreurs du modèle et sont mesurées comme résidus de l'estimation). Chacun de ces éléments est en principe susceptible d'être mal mesuré, mal spécifié, incorrectement estimé, ou de changer de manière non anticipée. On fait ici l'hypothèse qu'on a reformulé le modèle de façon à lui donner une représentation stationnaire (par exemple via un modèle vectoriel de correction d'équilibre). Pour simplifier la présentation, seul un processus de premier ordre est considéré. On suppose par conséquent que le vecteur (de dimension n) de données \mathbf{y}_t

est g n r  par :

$$\mathbf{y}_t = \boldsymbol{\phi} + \mathbf{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \quad \text{o  } \boldsymbol{\varepsilon}_t \sim \text{IN}[\mathbf{0}, \boldsymbol{\Omega}_\varepsilon], \quad (2)$$

o  $\text{IN}[\mathbf{0}, \boldsymbol{\Omega}_\varepsilon]$ indique une distribution Gaussienne multivari e d'esp rance nulle ($\text{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$) et de matrice de variance-covariance ($\text{Var}[\boldsymbol{\varepsilon}_t]$)  gale   $\boldsymbol{\Omega}_\varepsilon$; les  l ments $\boldsymbol{\varepsilon}_t$ et $\boldsymbol{\varepsilon}_{t'}$   des dates diff rentes  tant ind pendants. On suppose, dans ce cadre, que la distribution probabilistique des variables \mathbf{y}_t est stable dans le temps, et que, par cons quent, les param tres $\boldsymbol{\phi}$ et $\mathbf{\Pi}$ sont constants; par ailleurs, afin d' tablir la stationarit  de \mathbf{y}_t (ou en d'autres termes son caract re non int gr , not  $l(0)$), les valeurs propres de $\mathbf{\Pi}$ sont toutes de module inf rieur   1. L'esp rance de \mathbf{y}_t est dans ce cas donn e par $\boldsymbol{\varphi} = \text{E}[\mathbf{y}_t] = \text{E}[\mathbf{y}_{t-1}]$, et ainsi

$$\boldsymbol{\varphi} = \boldsymbol{\phi} + \mathbf{\Pi}\boldsymbol{\varphi},$$

soit $\boldsymbol{\varphi} = (\mathbf{I}_n - \mathbf{\Pi})^{-1} \boldsymbol{\phi}$. Ceci nous permet de reformuler (2) :

$$\mathbf{y}_t - \boldsymbol{\varphi} = \mathbf{\Pi}(\mathbf{y}_{t-1} - \boldsymbol{\varphi}) + \boldsymbol{\varepsilon}_t.$$

A partir d'un  chantillon de donn es de taille T , $\{\mathbf{y}_t, \text{ o  } t = 1, \dots, T\}$, et avec la notation $\hat{\cdot}$ pour les estimateurs et pr visions, il est possible, connaissant la forme du PGD donn e par (2), d'obtenir une pr vision $\hat{\mathbf{y}}_{T+h|T}$ de \mathbf{y}_{T+h} , o  $h = 1, \dots, H$. Celle-ci est obtenue en consid rant que les valeurs futures de $\boldsymbol{\varepsilon}_t$ (i.e. les $\boldsymbol{\varepsilon}_{T+s}$ pour $s \geq 1$) sont en moyenne nulles. En posant $\hat{\boldsymbol{\varphi}} = (\mathbf{I}_n - \hat{\mathbf{\Pi}})^{-1} \hat{\boldsymbol{\phi}}$, $\hat{\mathbf{y}}_{T+h|T}$ est fourni par un mod le autoregressif vectoriel (VAR(1) o  \mathbf{y}_t suit un mod le similaire   celui de l' quation (1)) qui relie les variables repr sent es par le vecteur \mathbf{y}_t   leur pass  :

$$\hat{\mathbf{y}}_{T+h|T} - \hat{\boldsymbol{\varphi}} = \hat{\mathbf{\Pi}}(\mathbf{y}_{T+h-1|T} - \hat{\boldsymbol{\varphi}}),$$

ce qui m ne  

$$\hat{\mathbf{y}}_{T+h|T} = \hat{\boldsymbol{\varphi}} + \hat{\mathbf{\Pi}}^h(\hat{\mathbf{y}}_T - \hat{\boldsymbol{\varphi}}), \quad (3)$$

o  les donn es connues et utilis es lors de la pr vision, $\hat{\mathbf{y}}_T$, sont susceptibles de r visions (il existe potentiellement des erreurs de mesure). Ainsi, $\hat{\mathbf{y}}_T \neq \mathbf{y}_T$ a priori mais, si on suppose que les r visions sont nulles en moyenne, $\text{E}[\hat{\mathbf{y}}_t] = \mathbf{y}_t$. La pr vision correspond donc   la somme de la moyenne $\hat{\boldsymbol{\varphi}}$ et de l' cart vis- -vis de celle-ci de la valeur estim e de \mathbf{y}_T   l'origine de pr vision,   un facteur $\hat{\mathbf{\Pi}}^h$ pr s.

En utilisant (3), on ne pr suppose pas que le mod lisateur connaisse la forme du PGD : il se peut que $\hat{\mathbf{\Pi}}$ soit mal sp cifi  en imposant a priori que certains coefficients soient nuls (ce qui revient   omettre des variables), ou en ajoutant des variables inad quates dans certaines  quations. Comme le syst me est dynamique, le comportement de chaque horizon de pr vision diff re et ce, surtout, lors de chocs sur le syst me. Si une rupture structurelle survient   l'instant T , de sorte que les param tres $(\boldsymbol{\phi}, \mathbf{\Pi})$ deviennent $(\boldsymbol{\phi}^*, \mathbf{\Pi}^*)$ et qu'  partir de $T + 1$, les donn es sont g n r es par :

$$\mathbf{y}_{T+h} = \boldsymbol{\phi}^* + \mathbf{\Pi}^*\mathbf{y}_{T+h-1} + \boldsymbol{\varepsilon}_{T+h}, \quad \text{o  } h \geq 1,$$

alors, en posant $\boldsymbol{\varphi}^* = (\mathbf{I}_n - \boldsymbol{\Pi}^*)^{-1} \boldsymbol{\phi}^*$, le système devient :

$$\begin{aligned} \mathbf{y}_{T+h} - \boldsymbol{\varphi}^* &= \boldsymbol{\Pi}^* (\mathbf{y}_{T+h-1} - \boldsymbol{\varphi}^*) + \boldsymbol{\varepsilon}_{T+h} \\ &= (\boldsymbol{\Pi}^*)^h (\mathbf{y}_T - \boldsymbol{\varphi}^*) + \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\varepsilon}_{T+h-i}. \end{aligned} \quad (4)$$

En utilisant les formules précédentes, et en posant $\widehat{\boldsymbol{\varepsilon}}_{T+h|T} = \mathbf{y}_{T+h} - \widehat{\mathbf{y}}_{T+h|T}$, on obtient l'erreur de prédiction :

$$\widehat{\boldsymbol{\varepsilon}}_{T+h|T} = \boldsymbol{\varphi}^* - \widehat{\boldsymbol{\varphi}} + (\boldsymbol{\Pi}^*)^h (\mathbf{y}_T - \boldsymbol{\varphi}^*) - \widehat{\boldsymbol{\Pi}}^h (\widehat{\mathbf{y}}_T - \widehat{\boldsymbol{\varphi}}) + \sum_{i=0}^{h-1} (\boldsymbol{\Pi}^*)^i \boldsymbol{\varepsilon}_{T+h-i}. \quad (5)$$

Il est possible de réarranger (5) de manière à faire ressortir les différentes influences en négligeant les termes croisés. On note à l'aide d'un indice p les limites, pour un échantillon de taille infinie, des estimateurs : $\boldsymbol{\Pi}_p = \lim_{T \rightarrow \infty} \widehat{\boldsymbol{\Pi}}$ (avec les mêmes mauvaises spécifications). Par ailleurs, on pose deux matrices⁶ \mathbf{C}_h et \mathbf{F}_h et l'opérateur v , qui transforme une matrice de taille $p \times q$ en un vecteur de dimension $pq \times 1$ où les colonnes de la matrice sont simplement placées les unes au-dessus des autres dans le vecteur transformé. Il nous est ainsi possible d'exprimer $\widehat{\boldsymbol{\varepsilon}}_{T+h|T}$ par les décompositions présentées dans le tableau 1. Celui-ci combine des effets dont l'origine peut être différente (une modification des propriétés de $\boldsymbol{\varepsilon}_t$ par exemple). On constate que l'espérance conditionnelle de toutes les composantes est nulle, sauf celle des lignes (ia) et (iia) . Les autres aspects n'induisent donc pas de biais systématique mais altèrent la précision de la prédiction via la variance de son erreur.

Ainsi des erreurs systématiques de prédiction interviennent-elles lorsque la moyenne d'équilibre est non constante ($\boldsymbol{\varphi} - \boldsymbol{\varphi}^* \neq \mathbf{0}$) ou mal spécifiée ($\boldsymbol{\varphi} - \boldsymbol{\varphi}_p \neq \mathbf{0}$). Prenons le cas de cette dernière hypothèse : en principe les résidus d'estimation sont construits de sorte que leur moyenne soit nulle, et donc que $\boldsymbol{\varphi}$ et $\boldsymbol{\varphi}_p$ coïncident, sauf si la moyenne s'est déjà modifiée par le passé de manière inaperçue. Un modèle bien formulé devrait donc éviter cette faille. En revanche, un choc déterministe prenant la forme d'un changement de moyenne induit, en (ia) , une erreur qui s'amplifie avec l'horizon car elle est fonction de $(\mathbf{I} - (\boldsymbol{\Pi}^*)^h)$ et $(\boldsymbol{\Pi}^*)^h \rightarrow \mathbf{0}$ d'après les hypothèses de stationarité (qui nécessitent que $\boldsymbol{\Pi}$, $\boldsymbol{\Pi}^*$, $\boldsymbol{\Pi}_p$ n'aient pas de valeur propre unitaire).

En revanche les composantes dynamiques (ib) et (iib) sont toutes deux multipliées par un terme d'espérance nulle ($\mathbf{y}_T - \boldsymbol{\varphi}$) et n'ont donc qu'un faible impact (qui serait d'ailleurs nul à tous horizons si l'origine de prédiction était égale à sa moyenne). Enfin les effets de variance résultent principalement des problèmes d'estimation (iva) et (ivb) ainsi que de l'accumulation de chocs (v) .

Le but de l'étude du système (2) était de montrer comment analyser la capacité prévisionnelle d'un modèle économétrique et d'apprendre à en tirer les conclusions adéquates. Dans la prochaine section, on s'attachera à montrer comment il est possible de choisir des modèles robustes à certaines erreurs de prédiction.

⁶Pour les définitions, voir Clements & Hendry (2000).

Tableau 1 – Taxinomie des composantes de l’erreur de prévision.

$\widehat{\mathbf{e}}_{T+h T} \simeq \left(\mathbf{I}_n - (\mathbf{\Pi}^*)^h \right) (\boldsymbol{\varphi} - \boldsymbol{\varphi}^*)$	(ia)	changement de moyenne d’équilibre
$+ \left((\mathbf{\Pi}^*)^h - \mathbf{\Pi}^h \right) (\mathbf{y}_T - \boldsymbol{\varphi})$	(ib)	changement de pente
$+ \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) (\boldsymbol{\varphi} - \boldsymbol{\varphi}_p)$	(iia)	équilibre mal spécifié
$+ \left(\mathbf{\Pi}^h - \mathbf{\Pi}_p^h \right) (\mathbf{y}_T - \boldsymbol{\varphi})$	(iib)	mauvaise spécification de la pente
$- \left(\mathbf{\Pi}_p^h - \mathbf{C}_h \right) (\mathbf{y}_T - \widehat{\mathbf{y}}_T)$	(iii)	incertitudes sur l’origine
$- \left(\mathbf{I}_n - \mathbf{\Pi}_p^h \right) \left(\widehat{\boldsymbol{\varphi}}_v - \boldsymbol{\varphi}_p \right)$	(iva)	estimation de la moyenne d’équilibre
$- \mathbf{F}_h \left(\widehat{\mathbf{\Pi}} - \mathbf{\Pi} \right)$	(ivb)	estimation de la pente
$+ \sum_{i=0}^{h-1} (\mathbf{\Pi}^*)^i \boldsymbol{\varepsilon}_{T+h-i}$	(v)	accumulation de chocs stochastiques

4 Obtenir des prévisions robustes

Cette section illustre comment utiliser la taxinomie présentée précédemment dans le cadre de la prévision ; par exemple comment pallier les chocs déterministes, ceux qui entraînent des erreurs d’espérance non nulle. Plusieurs voies s’ouvrent à nous : utiliser des méthodes purement statistiques ou améliorer un modèle existant. La réponse à apporter dépend de l’utilisation qu’on souhaite faire de la prévision. Mais dans les cas exposés ci-dessous, il est important de déterminer la forme du choc que les variables subissent. De manière simple, on s’intéressera donc en premier, dans le cadre du modèle présenté dans la section 3, à un déplacement de la moyenne d’équilibre de $\boldsymbol{\varphi}$ à $\boldsymbol{\varphi}^*$.

4.1 Transformation des variables

La méthode la plus naturelle vise à trouver une transformation des variables qui ne subisse pas les effets du choc. Ici, dans le cas d’un simple passage de $\boldsymbol{\varphi}$ à $\boldsymbol{\varphi}^*$, il est important d’en visualiser les conséquences pour le processus. Pour ce faire, on présente sur la figure 3 des données en présence ou en l’absence d’un tel choc. Il est aisé de constater qu’il s’agit effectivement d’un simple *déplacement* des variables. De ce fait, si on s’intéresse à la variation des données d’une période à la suivante, i.e. à la *différence* de y_t , on constate sur le graphique du bas de la figure que celle-ci n’est quasiment pas perturbée par le choc. Ainsi un modèle de prévision pourra-t-il prendre cet aspect en compte et par exemple s’efforcer de prévoir, non pas le niveau de l’indice des prix, mais son taux de croissance : l’inflation. Mathématiquement, en supposant que le choc intervienne à T :

$$\begin{aligned} \mathbf{y}_T &= \boldsymbol{\phi} + \mathbf{\Pi} \mathbf{y}_{T-1} + \boldsymbol{\varepsilon}_T, \\ \mathbf{y}_{T+1} &= \boldsymbol{\phi}^* + \mathbf{\Pi} \mathbf{y}_T + \boldsymbol{\varepsilon}_{T+1}, \end{aligned} \tag{6}$$

alors

$$\begin{aligned}\Delta \mathbf{y}_T &= \mathbf{\Pi} \Delta \mathbf{y}_{T-1} + \Delta \boldsymbol{\varepsilon}_T, \\ \Delta \mathbf{y}_{T+1} &= (\boldsymbol{\phi}^* - \boldsymbol{\phi}) + \mathbf{\Pi} \Delta \mathbf{y}_T + \Delta \boldsymbol{\varepsilon}_{T+1}, \\ \Delta \mathbf{y}_{T+2} &= \mathbf{\Pi} \Delta \mathbf{y}_{T+1} + \Delta \boldsymbol{\varepsilon}_{T+2}.\end{aligned}\tag{7}$$

Ceci implique que le processus générateur de $\Delta \mathbf{y}_t$ est quasiment constant sur l'échantillon et pratiquement non affecté par le choc (mis à part en $T + 1$).⁷ Il conviendrait donc, dans un tel cas, de modéliser, plutôt que la prévision de \mathbf{y}_{T+1} (par exemple le logarithme de l'indice des prix à la consommation) celle de $\Delta \mathbf{y}_T$ (l'inflation) et ainsi le modèle utilisé s'avèrerait valable pour quasiment toutes les valeurs futures. Cette transformation des variables génère une prévision via (7) identique à celle obtenu par (6) en l'absence de choc, mais permet de s'assurer que l'effet du choc n'est plus permanent mais transitoire (ici seule l'observation à $T + 1$ est affectée). Malheureusement, ce gain de *robustesse* s'acquiert au détriment d'une perte de précision : la variance des résidus de (7) est double de celle de (6). Enfin l'utilisation du taux de croissance au lieu du niveau de la variable se révèle intéressant dans le cadre du choc ci-dessus, lors d'un simple déplacement des variables, mais la transformation d'un choc permanent en transitoire peut se révéler plus difficile comme lorsque la dynamique même de la variable se trouve affectée : si $\mathbf{\Pi}$ devient $\mathbf{\Pi}^*$ car dans ce cas

$$\begin{aligned}\Delta \mathbf{y}_{T+1} &= (\mathbf{\Pi}^* - \mathbf{\Pi}) \Delta \mathbf{y}_T + \Delta \boldsymbol{\varepsilon}_{T+1}, \\ \Delta \mathbf{y}_{T+2} &= \mathbf{\Pi}^* \Delta \mathbf{y}_{T+1} + \Delta \boldsymbol{\varepsilon}_{T+2},\end{aligned}$$

et le passage au taux de croissance ne suffit plus.

4.2 Correction du modèle

L'autre méthode principale pour un choc de $\boldsymbol{\phi}^*$ à $\boldsymbol{\phi}$ consiste à corriger un modèle existant. Ainsi peut-il se révéler judicieux de prendre en compte l'erreur de prévision dans l'échantillon à disposition. En utilisant :

$$\hat{\mathbf{y}}_{t+1|t} = \hat{\boldsymbol{\phi}} + \hat{\mathbf{\Pi}} \mathbf{y}_t,$$

alors sur l'échantillon $t = 1, \dots, T$, i.e. pré-choc, les estimations de $\hat{\boldsymbol{\phi}}$ et de $\hat{\mathbf{\Pi}}$ ne devraient pas être trop éloignées des vraies valeurs des paramètres si le modèle est bien spécifié. Par conséquent, il est probable que les prévisions intra-échantillon (i.e. les valeurs estimées) soient de bonne qualité. En revanche, à partir de $T + 1$, les erreurs vont souffrir du choc. Il est possible de rendre ce modèle plus robuste en ajoutant, par exemple, à une prévision à t , l'erreur faite à la période précédente :

$$\hat{\mathbf{e}}_{t|t-1} = (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}),$$

⁷En pratique, et ainsi que constaté sur la figure 3, en présence et en absence d'un tel choc, les $\Delta \mathbf{y}_t$ sont légèrement modifiés autour de T car $\Delta \mathbf{y}_{T+1}$ intervient dans l'expression régissant $\Delta \mathbf{y}_{T+2}$.

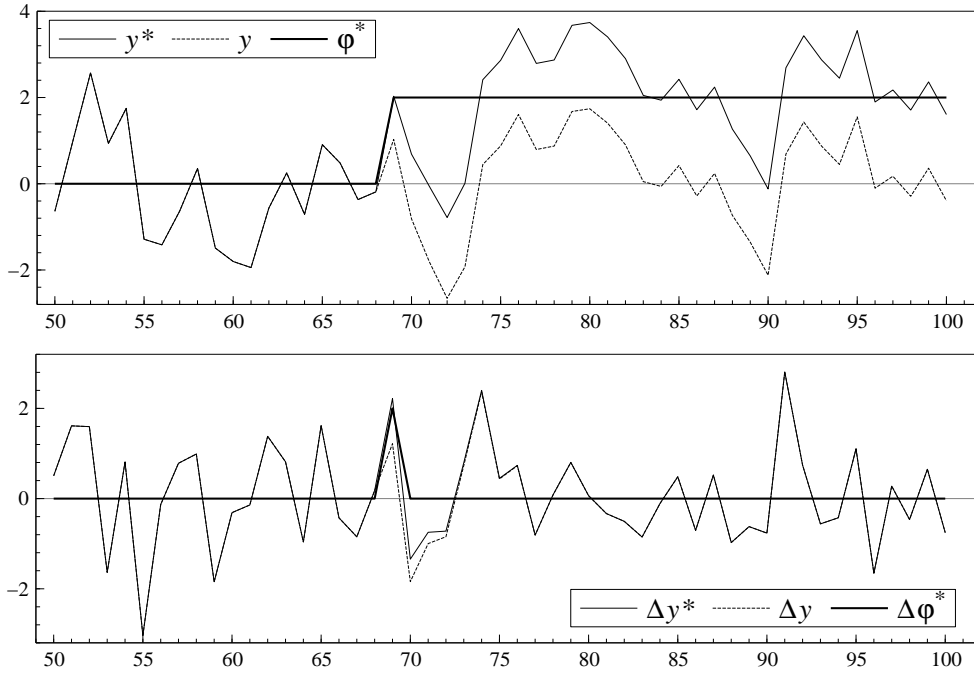


Figure 3 – Impact d’un déplacement de moyenne d’équilibre ($\varphi = 0, \varphi^* = 2$) pour un processus stationnaire et sa première différence. On présente les séries en présence (y_t^*) ou en l’absence (y_t) de choc.

ce qui donne :

$$\widehat{\mathbf{y}}_{t+1|t} = \widehat{\mathbf{y}}_{t+1|t} + (\mathbf{y}_t - \widehat{\mathbf{y}}_{t|t-1}).$$

Cette méthode implique une erreur de prévision :

$$\widehat{\mathbf{e}}_{t+1|t} = \mathbf{y}_{t+1} - \widehat{\mathbf{y}}_{t+1|t} = \widehat{\mathbf{e}}_{t+1|t} - \widehat{\mathbf{e}}_{t|t-1} = \Delta \widehat{\mathbf{e}}_{t+1|t},$$

et donc la correction revient à différencier l’erreur de prévision. Ce faisant, puisque l’erreur d’anticipation est faible avant le choc (et proche de $\boldsymbol{\varepsilon}_T \simeq \mathbf{y}_T - \widehat{\mathbf{y}}_{T|T-1}$ de variance $\boldsymbol{\Omega}_\varepsilon = \boldsymbol{\Sigma}\boldsymbol{\Sigma}'$), on ne contamine pas trop la première prévision :

$$\widehat{\mathbf{y}}_{T+1|T} = \widehat{\boldsymbol{\phi}} + \widehat{\boldsymbol{\Pi}}\mathbf{y}_T + (\mathbf{y}_T - \widehat{\mathbf{y}}_{T|T-1}),$$

mais puisque à $T + 1$ le choc a déjà eu lieu, l’erreur est de toute manière forte si $\|\boldsymbol{\phi}^* - \widehat{\boldsymbol{\phi}}\|^2 > tr(\boldsymbol{\Omega}_\varepsilon)$ ce qui implique que l’amplitude du choc soit supérieure à la variabilité (écart-type) des résidus. L’intérêt de cette méthode devient clair à la période suivante, car si

$$\mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T} \simeq \mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T} \simeq \boldsymbol{\phi}^* - \widehat{\boldsymbol{\phi}},$$

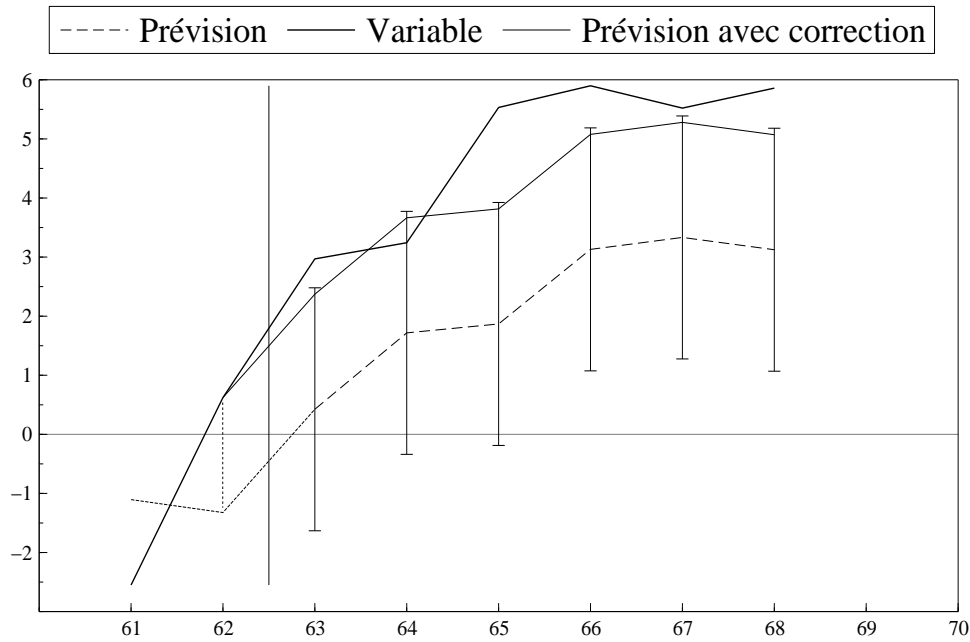


Figure 4 – Prédiction à partir d’un modèle autorégressif avec ou sans correction consistant à faire coïncider la prédiction avec l’origine de prédiction à $T = 62$. Les barres indiquent l’intervalle de confiance autour de la prédiction non corrigée. On constate l’inexactitude de cette dernière, et la pertinence de la correction.

en revanche, l’erreur redevient très faible à $T + 2$ car

$$\begin{aligned} \widehat{\mathbf{y}}_{T+2|T+1} &= \widehat{\boldsymbol{\phi}} + \widehat{\boldsymbol{\Pi}}\mathbf{y}_{T+1} + (\mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T}) \\ &\simeq \boldsymbol{\phi}^* + \widehat{\boldsymbol{\Pi}}\mathbf{y}_{T+1}, \end{aligned}$$

ce qui fait que la correction a “gommé” le déplacement $\boldsymbol{\phi} \rightarrow \boldsymbol{\phi}^*$ à peu de frais et ce, sans connaissance de l’existence ou de l’amplitude du choc : *le modèle s’ajuste de lui même* (ce qui ne signifie pas que le choc n’ait pas eu lieu). L’intérêt de cette méthode, comparée à la transformation des données, est qu’elle permet facilement de conserver les bonnes dynamiques du modèle, et donc, quand on souhaite davantage pouvoir anticiper les mouvements de l’économie qu’obtenir un simple chiffre, elle peut se révéler très fructueuse car elle cible directement la variable utilisée dans le modèle sans transformation. La figure 4 présente les prévisions obtenues par un modèle et sa correction : on constate aisément le gain qu’apporte la modification lorsque le choc intervient à la fin de l’échantillon, i.e. la date origine de la prédiction.

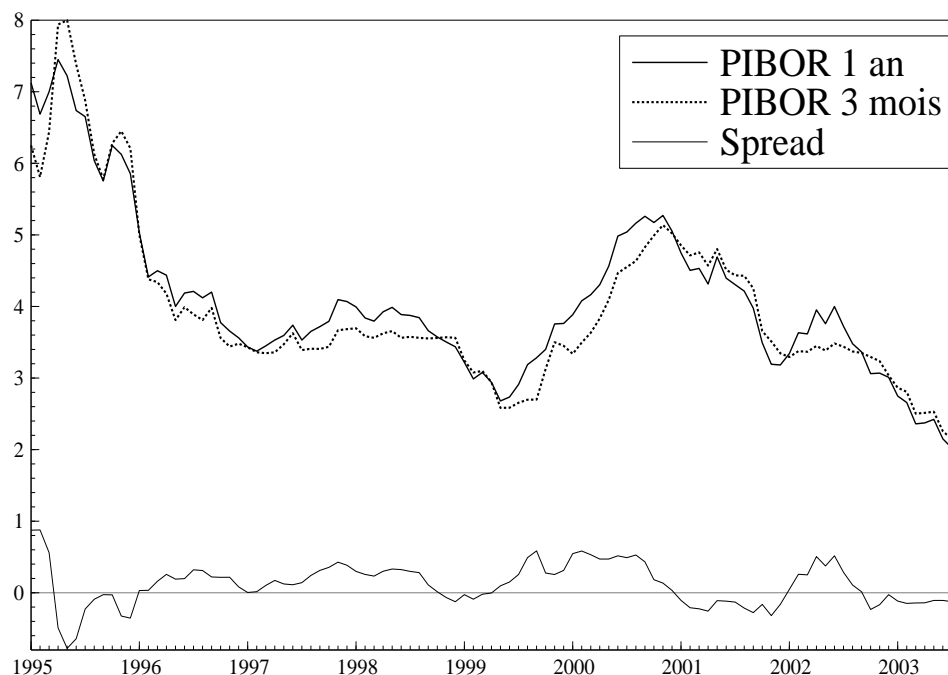


Figure 5 – Taux et spread de taux

4.3 Les relations de causalité

Un résultat *a priori* contre-intuitif résultant de l'analyse des causes d'erreur en matière de prévision concerne l'usage de relations de causalité. Il a été en effet montré théoriquement et validé empiriquement que l'usage de variables intervenant réellement dans le processus générant y_t , la variable à prévoir, n'est pas gage d'amélioration de la précision. Il en va par exemple de l'impact des relations de cointégration, et ce fut l'origine de grandes difficultés : l'introduction de la rémunération des comptes bancaires courants au Royaume-Uni (Banking Act de 1984) a entraîné une modification du coût d'opportunité de détention de monnaie liquide (agrégat M1), ce dont les agents ne se sont que progressivement rendu compte. La banque d'Angleterre n'a pas intégré ce changement dans la conduite de sa politique monétaire (le passage du taux d'intérêt réel à 3 mois à l'écart entre ce taux réel et le taux des dépôts à vue) et a systématiquement sous-estimé le taux de croissance réel de la monnaie car elle a mal pris en compte la modification résultante d'une relation de cointégration entrant dans son modèle (voir Hendry, 2004). La cointégration est le phénomène selon lequel des variables dont le comportement est potentiellement très instable peuvent présenter des relations qui, quant à elles, sont nettement plus stables. Un exemple peut en être le spread de taux entre le PIBOR à un an et à trois mois qui sont représentés sur la figure 5. Le comportement des taux y apparaît erratique (bien qu'il soit possible d'en expliquer les mouvements) tandis que le spread oscille pour sa part autour de zéro. Ainsi les taux peuvent être vus comme intégrés (et imprévisibles par leur passé seul) et leur différence (une combinaison linéaire des deux) comme ayant une moyenne et

une variance constantes (le spread ne sort jamais d'une bande de $\pm 1\%$).

Ainsi, de manière générale, si \mathbf{x}_t est un vecteur de variables intégrées d'ordre 1 (i.e. variables non-stationnaires, dont la différence $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ est stationnaire), il peut exister une combinaison linéaire des variables de \mathbf{x}_t , donnée par le vecteur $\boldsymbol{\beta}$, tel que $\boldsymbol{\beta}'\mathbf{x}_t$ soit stationnaire, i.e. oscillant autour d'un équilibre $E[\boldsymbol{\beta}'\mathbf{x}_t] = \boldsymbol{\mu}$. Ceci nous permet de reformuler le processus sous la forme d'un *mécanisme de correction d'équilibre* où l'écart par rapport au long-terme intervient dans les fluctuations de court terme, ce qui s'écrit dans le cas le plus simple :

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \boldsymbol{\alpha} (\boldsymbol{\beta}'\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t. \quad (8)$$

Dans cet exemple, il est dit que \mathbf{x}_{t-1} cause $\Delta \mathbf{x}_t$ (au sens de Granger). Cependant, comme nous l'avons vu précédemment, si, à la suite d'interventions de politique économique, une modification de $\boldsymbol{\mu}$ intervient (il peut en exister en pratique maintes raisons) qui implique un nouvel équilibre à $\boldsymbol{\mu}^*$ (par exemple en T), l'utilisation de \mathbf{x}_{t-1} dans la prévision entraîne le modèle vers un équilibre erroné. Il est sans doute préférable dans un tel cas de ne pas recourir à la relation de causalité entre $\boldsymbol{\beta}'\mathbf{x}_{t-1}$ et $\Delta \mathbf{x}$, mais d'utiliser un modèle "non-causal" :

$$\Delta \mathbf{x}_t = \boldsymbol{\gamma} + \mathbf{u}_t, \quad (9)$$

où $E[\mathbf{u}_t] = \mathbf{0}$, de sorte que le modèle (8) génère une erreur de $E[\widehat{\boldsymbol{\epsilon}}_{T+1|T}] = -\boldsymbol{\alpha}(\boldsymbol{\mu}^* - \boldsymbol{\mu})$ pour (9) donnant $E[\widetilde{\boldsymbol{\epsilon}}_{T+1|T}] = \mathbf{0}$ car $E[\boldsymbol{\beta}'\mathbf{x}_T - \boldsymbol{\mu}^*] = \mathbf{0}$. Cette technique (9), notée DV, diffère des précédentes, car il ne s'agit pas ici de modéliser le taux de croissance mais en réalité d'imposer une tendance déterministe à partir de l'origine de la prévision :

$$\begin{aligned} \mathbf{x}_{T+1} &= \boldsymbol{\gamma} + \mathbf{x}_T + \mathbf{u}_{T+1} \\ \mathbf{x}_{T+h} &= h\boldsymbol{\gamma} + \mathbf{x}_T + \sum_{i=1}^h \mathbf{u}_{T+i} \end{aligned}$$

ou une simple marche aléatoire dans le cas $\boldsymbol{\gamma} = \mathbf{0}$. Enfin dans certains, il est aussi possible d'utiliser par exemple le modèle DDV

$$\Delta^2 \mathbf{x}_t = \boldsymbol{\tau} + \mathbf{v}_t,$$

qui impose une tendance quadratique.

4.4 Prévision à horizon variable

Si on souhaite obtenir des prévisions à plusieurs horizons, en supposant toujours que le modèle est imparfait, et en reprenant la formulation en VAR (en ignorant la constante par souci de clarté) :

$$\mathbf{y}_t = \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t,$$

de sorte que :

$$\mathbf{y}_{t+h} = \mathbf{\Pi}^h \mathbf{y}_t + \sum_{i=0}^{h-1} \mathbf{\Pi}^i \boldsymbol{\varepsilon}_{t+h-i},$$

alors, la prévision correspondant à h périodes est donnée par

$$\widehat{\mathbf{y}}_{T+h|T} = \widehat{\mathbf{\Pi}}^h \mathbf{y}_T.$$

On constate que si $\widehat{\mathbf{\Pi}} = \mathbf{\Pi} + \boldsymbol{\delta}_{\mathbf{\Pi}}$ est un bon estimateur de $\mathbf{\Pi}$, il n'en va pas nécessairement de même de $\widehat{\mathbf{\Pi}}^h$ pour $\mathbf{\Pi}^h$, car la différence entre les deux est fonction croissante de l'horizon h : $\widehat{\mathbf{\Pi}}^h \simeq \mathbf{\Pi}^h + h\boldsymbol{\delta}_{\mathbf{\Pi}}$. Il peut donc être plus judicieux de s'intéresser directement à la relation entre \mathbf{y}_{t+h} et \mathbf{y}_t en estimant sur l'échantillon :

$$\mathbf{y}_t = \boldsymbol{\Psi}_h \mathbf{y}_{t-h} + \mathbf{u}_{h,t}.$$

$\widetilde{\boldsymbol{\Psi}}_h$ ainsi obtenu permet d'établir une prévision :

$$\widetilde{\mathbf{y}}_{T+h|T} = \widetilde{\boldsymbol{\Psi}}_h \mathbf{y}_T.$$

Ici, le modèle de prévision est formulé de manière à minimiser l'erreur intra-échantillon correspondante. En théorie cette dernière méthode est asymptotiquement optimale (i.e. pour de grands échantillons). En pratique, elle permet souvent de capturer des effets de *mémoire longue* absents des modèles à court horizon : utilisée via un glissement (annuel par exemple sur données trimestrielles), $\Delta_4 y_t = y_t - y_{t-4}$, elle permet de s'extraire de la difficulté des périodicités infra-annuelles de type saisonnier.

4.5 Données inexactes

Lorsque les données sont inexactes, ce qui est toujours le cas pour les variables macroéconomiques vers la partie la plus récente de l'échantillon, l'erreur initiale est transmise à toutes les prévisions alors même que le modèle peut-être adéquat. Il convient donc d'élaborer des méthodes qui permettent de discerner si une anomalie correspond à un choc entraînant une mauvaise spécification du modèle ou si elle résulte simplement d'un problème de mesure des données. Dans le cadre des méthodes précédentes, l'observation d'un fort résidu dans la modélisation de $\widehat{\mathbf{y}}_T$ est-elle due à une donnée inexacte ou cache-t-elle un choc $\boldsymbol{\phi} \rightarrow \boldsymbol{\phi}^*$?

Les révisions de données étant souvent très importantes, l'usage d'enquêtes et de sondages s'avère souvent profitable à l'analyse des données les plus récentes. De nombreuses méthodes ad hoc existent pour prendre en compte cet aspect. Parmi celles-ci, le prévisionniste choisit parfois de procéder à une correction du modèle semblable à celle présentée précédemment, mais qui prend la forme d'un lissage des erreurs récentes de prévision, ou une moyenne de ces dernières. Les méthodes utilisées pour un choc peut être aussi utilisées dans ce cadre. Mais le jugement du prévisionniste est ici essentiel.

4.6 Combiner les prévisions

Nombreux sont les organismes qui établissent des prévisions, et celles-ci sont souvent divergentes. L'expérience a montré qu'en général, la moyenne⁸ des prévisions s'avère rétrospectivement plus précise que la plupart de celles-ci, prises individuellement. Il existe d'ailleurs une publication dédiée à cette tâche : Consensus Forecast. La cause en est intuitivement simple : les prévisions sont très probablement proches de la réalité, mais quand certaines pèchent par optimisme d'autres sont coupables de pessimisme. Ainsi les particularités s'effaceraient-elles quand on les additionne. Il a été plus difficile d'expliquer ceci de manière économétrique. En effet, il apparaît difficile de prouver que la moyenne puisse être plus précise que le meilleur des éléments, et ce d'autant plus que, lors d'un choc, la plus forte probabilité est que toutes les prévisions soient biaisées dans la même direction. Pourtant, lorsque chaque modèle n'utilise qu'une fraction de l'information utile, un même choc peut avoir des conséquences opposées sur les prévisions, justifiant de ce fait l'utilisation de leur moyenne.

5 Evaluer les prévisions

Comment, entre plusieurs prévisions, choisir la bonne ? Cela dépend du critère retenu. On peut supposer par exemple que l'erreur de prévision, $\hat{e}_{T+1|T}$, soit nulle en moyenne ; ceci implique en pratique qu'au fil du temps les anticipations seront également sur- et sous-estimées. Ceci s'écrit, en utilisant l'espérance mathématique :⁹

$$\mathbf{E} [\hat{e}_{T+1|T}] = 0.$$

Un critère d'évaluation qui privilégierait la minimisation de $\mathbf{E} [\hat{e}_{T+1|T}]$ mettrait ainsi l'accent sur la *précision de la prédiction*. Si on estime, en revanche, qu'il n'est pas important d'obtenir un chiffre parfait, mais qu'il est essentiel de ne jamais *trop* se tromper et qu'on souhaite par conséquent pouvoir déterminer un faible intervalle de confiance autour de la prédiction, on peut dans ce cas s'astreindre à minimiser la variance :

$$\mathbf{V} [\hat{e}_{T+1|T}] = \mathbf{E} \left[(\hat{e}_{T+1|T} - \mathbf{E} [\hat{e}_{T+1|T}])^2 \right].$$

En général, en faisant l'hypothèse que l'erreur $\hat{e}_{T+1|T}$ suit une distribution Normale, \mathbf{y}_{T+1} est compris dans 95% des cas dans l'intervalle :

$$\left[\hat{\mathbf{y}}_{T+1|T} - 2\sqrt{\mathbf{V} [\hat{e}_{T+1|T}]}, \hat{\mathbf{y}}_{T+1} + 2\sqrt{\mathbf{V} [\hat{e}_{T+1|T}]} \right].$$

Minimiser $\mathbf{V} [\hat{e}_{T+1|T}]$ revient donc à privilégier la *certitude* (théorique) de $\hat{\mathbf{y}}_{T+1|T}$. En pratique, cependant, $\mathbf{E} [\hat{e}_{T+1|T}] \neq 0$, et donc le prévisionniste doit atteindre un équilibre

⁸moyenne simple, pondérée, voire analyse en composante principale (voir Stock & Watson, 1999)

⁹On suppose pour faciliter la présentation que le modèle est univarié.

entre précision et certitude. Les travaux théoriques ont tendance à se référer à la moyenne quadratique d'erreur de prévision :

$$\text{MQEP} (\widehat{e}_{T+1|T}) = \mathbf{E} [\widehat{e}_{T+1|T}^2] = \mathbf{E} [\widehat{e}_{T+1|T}]^2 + \mathbf{V} [\widehat{e}_{T+1|T}],$$

qui attribue de l'importance aux deux critères. Pour clarifier la différence entre les critères, reprenons l'exemple d'un choc $\phi \rightarrow \phi^*$ à l'instant T dans l'équation (2). Avec les mêmes notations que précédemment (on néglige les erreurs dues à l'estimation des paramètres dans la formule (5)),

$$\begin{aligned} y_t &= \phi + \pi y_{t-1} + \varepsilon_t, \quad t \leq T \\ y_t &= \phi^* + \pi y_{t-1} + \varepsilon_t, \quad t > T, \end{aligned}$$

les prévisions $\widehat{y}_{T+1|T}$ et $\widehat{y}_{T+2|T+1}$ (si on ignore qu'un choc est survenu) génèrent des erreurs $\mathbf{E} [\widehat{e}_{T+1|T}] = \phi^* - \phi = \mathbf{E} [\widehat{e}_{T+2|T+1}]$ et $\mathbf{V} [\widehat{e}_{T+1|T}] = \sigma_\varepsilon^2 = \mathbf{V} [\widehat{e}_{T+2|T+1}]$ de sorte que

$$\text{MQEP} (\widehat{e}_{T+1|T}) = \text{MQEP} (\widehat{e}_{T+2|T+1}) = (\phi^* - \phi)^2 + \sigma_\varepsilon^2$$

En revanche, s'intéresser à la différence $\Delta y_t = \pi \Delta y_{t-1} + \Delta \varepsilon_t$ fournirait une erreur de prévision $\mathbf{E} [\widetilde{e}_{T+1|T}] = \phi^* - \phi$, mais $\mathbf{E} [\widetilde{e}_{T+2|T+1}] = 0$ et $\mathbf{V} [\widetilde{e}_{T+1|T}] = 2\sigma_\varepsilon^2 = \mathbf{V} [\widetilde{e}_{T+2|T+1}]$ ce qui donne une MQPE supérieure à celle du modèle en niveau pour $T + 1$

$$\text{MQEP} (\widetilde{e}_{T+1|T}) = (\phi^* - \phi)^2 + 2\sigma_\varepsilon^2,$$

mais au prix d'une réduction pour les horizons suivants :

$$\text{MQEP} (\widetilde{e}_{T+2|T+1}) = 2\sigma_\varepsilon^2.$$

La prévision de y_{T+1} est ici moins précise, mais plus certaine que celle de Δy_{T+1} (espérance plus élevée, moindre variance). Selon les valeurs des paramètres et les priorités du prévisionniste l'une ou l'autre méthode serait à privilégier : la prévision en différence se révèle préférable selon ce critère et dans ce cadre pourvu que $|\phi^* - \phi| > \sigma_\varepsilon$. De manière générale, lorsqu'on modifie un modèle pour affiner la précision de la prévision, ceci se fait au détriment de la certitude l'entourant. Il est, par conséquent, essentiel de toujours préciser dans quel but, pour quelle utilité, une prévision est réalisée car selon le type de jugement qui lui est appliqué – il en va des critères comme des Pyrénées – elle peut très facilement se trouver vérité en-deça, erreur au-delà.

6 Prévoir les importations françaises

Afin d'observer comment s'appliquent les méthodes présentées, nous nous intéressons ici à un modèle d'importations françaises de biens et services, tel qu'utilisé dans le modèle de l'économie française de l'OFCE (*emod.fr*, voir Chauvin et al, 2002). Il s'agit d'un mécanisme à correction d'équilibre selon lequel s'établit un équilibre de long terme entre les importations de biens et services en volume (M_t), la demande intérieure (D_t), le prix

relatif des importations (hors pétrole) vis-à-vis du prix de la valeur ajoutée française (P_t^M/P_t^{VA}) et la durée d'utilisation des équipements dans l'industrie (DUE_t , représentant les tensions sur la capacité productive de l'économie)¹⁰. En notant en lettres minuscules les logarithmes des variables, cet équilibre, (estimé de 1983(2) à 1999(4), selon une méthode à une étape, voir Banerjee et al, 1995) est représenté par la relation de cointégration :

$$c_t = m_t - d_t + 0.087 (p_t^M - p_t^{VA}) - 0.84due_t + 3.56 - 3.7 \times 10^{-3}t.$$

L'élasticité des importations à leur prix relatif est nettement plus faible que celle retenue dans le modèle emod (0.6), mais nous la conservons cependant. L'équation d'importations régit leur taux de croissance ($\Delta m_t = m_t - m_{t-1}$) en fonction de ceux des variables mentionnées précédemment et du déséquilibre c_{t-1} selon : (erreurs standards entre parenthèses, valeurs-p entre crochets)

$$\Delta m_t = \underbrace{-6.44 \times 10^{-3}}_{(2.23 \times 10^{-3})} - \underbrace{0.51}_{(0.096)} c_{t-1} + \underbrace{1.20}_{(0.20)} \Delta d_t + \underbrace{0.85}_{(0.20)} \Delta d_{t-1} + \underbrace{0.94}_{(0.33)} \Delta due_t + \varepsilon_t \quad (10)$$

$$R^2 = 0.66 \quad \hat{\sigma}_\varepsilon = 1.03\% \quad SC = -8.78 \quad \chi_{nd}^2 = 0.61 [0.74] \quad F_{ar} = 0.57 [0.69].$$

Le modèle (10) n'est pas rejeté par les tests de spécification, est stable sur la période d'estimation et représente bien les variations des importations dans l'échantillon (voir figure 6). En revanche, si les fluctuations du taux de croissance sont effectivement prévues à partir de 2000 ($\Delta \hat{m}_t$), leur niveau et amplitude présentent des erreurs. Afin d'observer comment améliorer les prévisions en pratique, nous allons tour à tour utiliser des ajustements différents de (10), notée MCE. La méthode la plus simple vise à s'abstraire des problèmes de dynamique et à poser $\Delta \hat{m}_t^{DV} = \Delta m_{t-1}$, ainsi le taux de croissance d'hier perdure aujourd'hui (méthode DV). Sinon, la méthode IC (*intercept correction*, ou correction du biais), corrige la prévision $\Delta \hat{m}_t$ obtenue par (10) de l'erreur de prévision observée à la période précédente :

$$\Delta \hat{m}_t^{IC} = \Delta \hat{m}_t + (\Delta m_t - \Delta \hat{m}_{t-1}).$$

Sinon, comme nous avons vu que les modèles à correction d'erreur peuvent présenter des faiblesses si le niveau de l'équilibre de long terme se modifie, nous recourrons à un modèle n'utilisant que la dynamique de court terme (modèle CT^e) :

$$\Delta \hat{m}_t^{CT^e} = \underbrace{-1.00 \times 10^{-3}}_{(2 \times 10^{-3})} + \underbrace{1.06}_{(0.20)} \Delta d_t + \underbrace{0.62}_{(0.19)} \Delta d_{t-1} + \underbrace{0.46}_{(0.26)} \Delta due_t \quad (11)$$

$$R^2 = 0.47 \quad \hat{\sigma}_\nu = 1.2\% \quad SC = -8.67 \quad \chi_{nd}^2 = 0.50 [0.78] \quad F_{ar} = 1.43 [0.23]$$

Enfin deux modèles purement statistiques peuvent se révéler utiles, soit un processus autorégressif d'ordre 1, AR(1), estimé avant 2000 et qui ne représente pas très bien les variations :

$$\begin{aligned} \Delta m_t &= \tau + \rho \Delta m_{t-1} + \nu_t, \\ &= \underbrace{8.92 \times 10^{-3}}_{(2.18 \times 10^{-3})} + \underbrace{0.26}_{(0.11)} \Delta m_{t-1} + \nu_t \end{aligned}$$

$$R^2 = 0.07 \quad \hat{\sigma}_\nu = 15.6\% \quad SC = -8.23 \quad \chi_{nd}^2 = 0.44 [0.80] \quad F_{ar} = 2.70 [0.03]$$

¹⁰Données INSEE sauf Banque de France pour la durée d'utilisation des équipements.

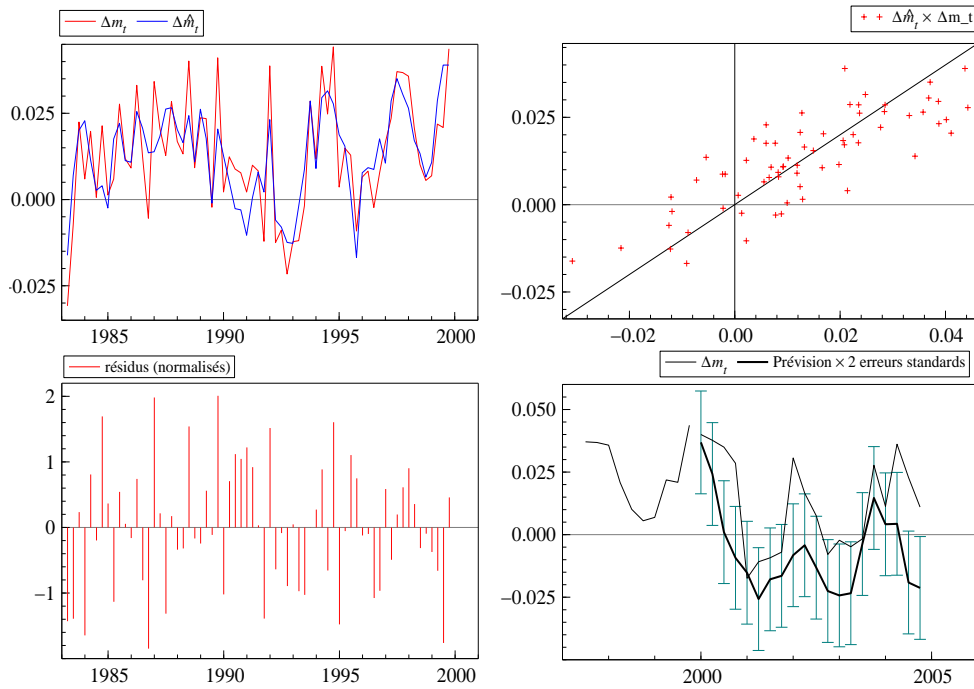


Figure 6 – Estimation et prévision (2000-2004) du taux de croissance des importations françaises.

ou un modèle ARMA plus général : par estimation on aboutit au modèle ARMA(2, 3)

$$\Delta m_t = \underset{(3.86 \times 10^{-3})}{12.5 \times 10^{-3}} - \underset{(.12)}{0.23} \Delta m_{t-1} + \underset{(0.11)}{0.41} \Delta m_{t-2} + v_t + \underset{(0.10)}{0.45} v_{t-1} + \underset{0.09}{0.49} v_{t-3}$$

$$R^2 = 0.26 \quad \hat{\sigma}_v = 1.47\% \quad SC = -5.44 \quad \chi_{nd}^2 = 0.20 [0.90] \quad F_{ar} = 1.18 [0.33]$$

Ces deux modèles fournissent respectivement $\Delta \hat{m}_t^{AR}$ et $\Delta \hat{m}_t^{ARMA}$. Les prévisions obtenues par ces différents modèles sont présentées figure 7 où à partir de 2000 sont reportés le taux de croissance effective, la prévision ainsi que les bornes d'un intervalle de confiance à 95% (en pointillés). Les modèles DV et AR présentant des erreurs standards très élevées, nous avons également placé deux graphiques sans intervalles de confiance. Le modèle CT existe sous deux formes : l'une où les coefficients sont repris de l'équation (10), notée CT, et l'autre en utilisant (11), notée CT^e. Deux éléments semblent importants : (1) les modèles "économétriques" perçoivent mieux les retournements et pics de la série et les corrections IC et CT, CT^e ont l'avantage sur MCE de ne pas s'éloigner durablement des valeurs à prévoir – ils sont d'ailleurs plus souvent de même signe que Δm_t ; (2) les méthodes "statistiques", AR, ARMA et DV, s'auto-correctent rapidement mais peuvent générer de fortes erreurs (cf 2001(1) et 2002(1) pour DV) et peuvent présenter une très forte incertitude.

Il est difficile de hiérarchiser les techniques de prévision à la seule vue de la figure 7. Nous nous intéressons donc à des critères d'évaluation et reportons ci-dessous les MQEP

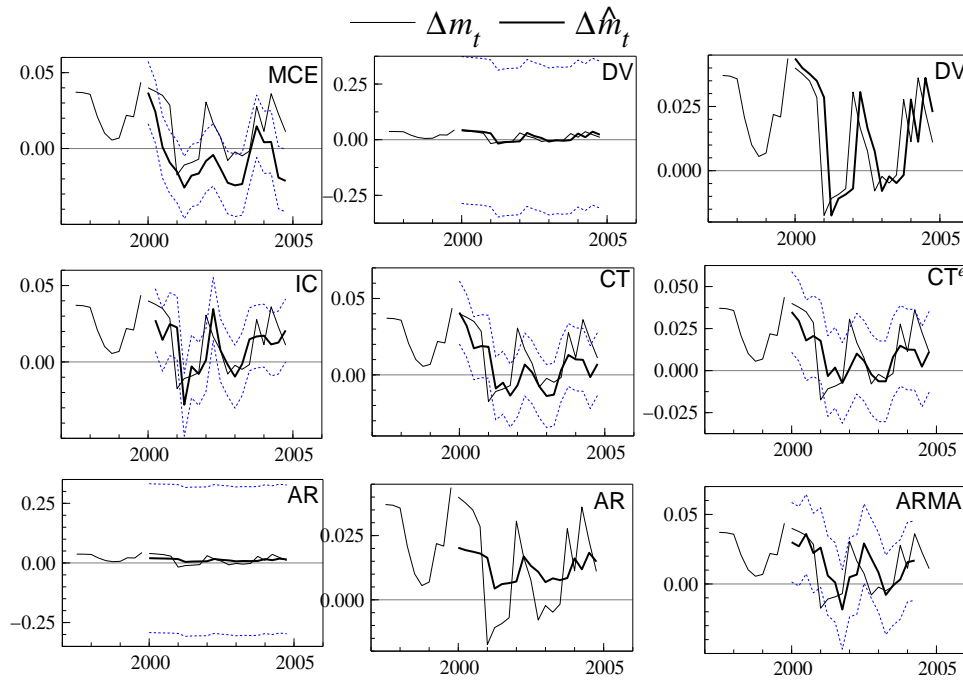


Figure 7 – Prédiction (2000-2004) du taux de croissance des importations françaises.

et MAEP (Moyenne absolue d'erreur de prédiction, $E[|\hat{e}_{T+1|T}|]$). Comme la fonction de distribution des erreurs n'est pas connue (car elle impliquerait de connaître le véritable mécanisme ayant généré Δm_t), nous utilisons comme approximation de l'espérance la moyenne des erreurs entre 2000(1) et 2004(4), soit 20 observations :

en %	CT ^e	IC	CT	AR	ARMA	DV	MCE
$\sqrt{\text{MQEP}}$	1.43	1.58	1.59	1.61	1.69	1.77	2.31
MAEP	1.05	1.21	1.15	1.40	1.32	1.28	1.94

La méthode la plus précise, au sens de la MQEP, est ici de loin CT^e, la technique utilisant les variations de court terme, suivie par IC qui utilise le modèle complet mais corrigé de son erreur récente. Les modèles autorégressifs se situent assez bien eux aussi pour des méthodes n'utilisent que la passé récent de la série sans aucun apport de théorie économique. La méthode DV est meilleure en prenant en compte la moyenne absolue car il lui arrive de générer de très fortes erreurs, ce que la moyenne quadratique pénalise fortement. Enfin, le mécanisme à correction d'erreur s'avère le moins précis dans cet exercice et les techniques alternatives présentées améliorent effectivement la qualité de la prédiction, jusqu'à diviser l'erreur par deux en moyenne : environ 1 point de pourcentage d'amélioration pour CT^e !

7 Conclusion

Il est suprenant de remarquer le retard que l'analyse de la prévision a sur les autres domaines économiques et économétriques. Ceci tient évidemment à son objet même : l'avenir. Il est pourtant réconfortant de constater comment les développements se nourrissent les uns les autres. En effet, s'il n'est pas pensable que les échecs récurrents en matière de prévision disparaissent complètement, il est clair – on l'espère du moins après lecture de cet article – que la compréhension de ceux-ci progresse et qu'il est possible, de ce fait, d'envisager de les réduire à l'avenir. Ainsi, si une bonne modélisation des comportements économiques est essentielle, ce n'est pourtant pas un gage de bonne prévision car l'échec provient de la modification de ceux-ci. La théorie de l'optimalité s'écroule quand le monde change. En revanche, une bonne compréhension des phénomènes entraînant des altérations dans un modèle permet d'essayer de les pallier, même sans avoir connaissance de leur apparition. On aboutit ainsi à des prévisions *robustes*. Dans les exemples présentés ici, il est apparu que les chocs sont difficiles à prendre en compte *ex ante*, mais qu'il est du moins possible de développer des méthodes (modélisation du taux de croissance, correction...) qui permettent de limiter l'incidence du choc sur les prévisions ultérieures ou qui fournissent des prévisions ni très précises ni très erronées (les méthodes DV et DDV), apportant de ce fait une réponse différente au problème des chocs.

La conclusion de notre analyse ne peut se formuler qu'en une incitation à formaliser le processus de prévision afin de pouvoir anticiper son échec, quantifier son inexactitude, faciliter sa reproduction, et progresser dans sa compréhension.

Références

- Banerjee, A., J. Dolado, J. Galbraith, and D. Hendry (1993). *Co-Integration, Error-Correction, and the Econometric Analysis of Non-stationary Data*. Oxford and New York : Oxford University Press.
- Banerjee, A., D. F. Hendry, and G. E. Mizon (1996). The econometric analysis of economic policy. *Oxford Bulletin of Economics and Statistics* 58, 573–600.
- Bhansali, R. J. (2002). Multi-step forecasting. See Clements and Hendry (2002), pp. 206–221.
- Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis, Forecasting and Control* (2nd ed.). San Francisco, CA : Holden-Day. First published, 1970.
- Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business and Economic Statistics* 11(2), 121–135.
- Chauvin, V., G. Dupont, E. Heyer, and X. Timbeau (2002). Le modèle France de l'OFCE. *Revue de l'OFCE* 81, 245–300.
- Chevillon, G. and D. F. Hendry (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting* 21, 201–18.
- Clements, M. P. and D. F. Hendry (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting* 12, 617–637.

- Clements, M. P. and D. F. Hendry (1995a). Forecasting in cointegrated systems. *Journal of Applied Econometrics* 10, 127–146.
- Clements, M. P. and D. F. Hendry (1995b). Forecasting in macro-economics. In *Time Series Models in Econometrics, Finance and Other Fields*, pp. 101–141. London : Chapman and Hall.
- Clements, M. P. and D. F. Hendry (1995c). Macro-economic forecasting and modelling. *Economic Journal* 105, 1001–1013.
- Clements, M. P. and D. F. Hendry (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics* 58, 657–683.
- Clements, M. P. and D. F. Hendry (1998). *Forecasting Economic Time Series*. Cambridge : Cambridge University Press.
- Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-Stationary Economic Time Series*. Cambridge, MA : The MIT Press.
- Clements, M. P. and D. F. Hendry (2000). An historical perspective on forecast errors. *National Institute Economic Review* 177, 70–82.
- Clements, M. P. and D. F. Hendry (2001). Forecasting with difference-stationary and trend-stationary models. *Econometrics Journal* 4, S1–S19.
- Clements, M. P. and D. F. Hendry (Eds.) (2002). *A Companion to Economic Forecasting*. Oxford : Blackwell Publishers.
- Clements, M. P. and D. F. Hendry (2004). Pooling of forecasts. *Econometrics Journal* 7, 1–31.
- Diebold, F. X. (1998). The past, present and future of macroeconomic forecasting. *Journal of Economic Perspectives* 12, 172–195.
- Ericsson, N. R. and J. Marquez (1998). A framework for economic forecasting. *Econometrics Journal* 1, C228–C226.
- Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operations Research Quarterly* 20, 199–207.
- Harvey, A. C. (1993). *Time Series Models* (2nd ed.). Hemel Hempstead : Harvester Wheatsheaf. First edition : 1981.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford : Oxford University Press.
- Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics* 11, 45–65.
- Hendry, D. F. (2004). Robustifying forecasts from equilibrium-correction models. Economic working paper 2004-w14, Nuffield College.
- Hendry, D. F. and J. A. Doornik (1997). The implications for econometric modelling of forecast failure. *Scottish Journal of Political Economy* 44, 437–461. Special Issue.
- Johnston, H. N., L. Klein, and K. Shinjo (1974). Estimation and prediction in dynamic econometric models. In W. Sellekaerts (Ed.), *Essays in honor of Jan Tinbergen*. London : Macmillan.
- Lemoine, M. and F. Pelgrin (2003). Introduction aux modèles espace-état et au filtre de Kalman. *Revue de l'OFCE* 86, 223–229.

- Lucas, R. E. (1976). Econometric policy evaluation : A critic. In K. Brunner and A. Meltzer (Eds.), *The Phillips Curve and Labor Market*, pp. 19–46. Amsterdam : North-Holland.
- Makridakis, S. and M. Hibon (2000). The M3 competition : Results, conclusions and implications. working paper, INSEAD.
- Sampson, M. (1991). The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. *Journal of Applied Econometrics* 6, 67–76.
- Schumpeter, J. (1933). The common sense of econometrics. *Econometrica* 1(1), 5–12.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* 48(1), 1–48.
- Stock, J. H. and M. W. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time-series. In R. F. Engle and H. White (Eds.), *Cointegration, Causality and Forecasting : A Festschrift in honour of Clive W. J. Granger*, pp. 1–44. Oxford : Oxford University Press.
- Wallis, K. F. (1993). Comparing macroeconomic models : A review article. *Economica* 60, 225–237.
- Wold, H. O. A. (1938). *A study in the analysis of stationary time series*. Stockholm : Almqvist and Wicksell.