

Prévisions d'inflation par des modèles vectoriels autorégressifs

Thomas J. Jordan, Chef de la Direction de la recherche,
Banque nationale suisse, Zurich

Peter Kugler, Professeur, Université de Bâle

Carlos Lenz, Professeur, Université de Bâle

Marcel R. Savioz, Conseiller scientifique, Direction de la recherche,
Banque nationale suisse, Zurich

Dans la nouvelle stratégie de politique monétaire de la Banque nationale suisse (BNS), la prévision de l'inflation joue un rôle clé. D'une part, elle est à la base des décisions de politique monétaire, d'autre part, elle constitue un important outil de communication avec le public. Afin d'améliorer encore la transparence de sa politique monétaire, la Banque nationale a décidé de publier différentes études expliquant le calcul des prévisions d'inflation et leur influence sur les décisions de politique monétaire. Dans ce contexte, elle a déjà publié les articles de Jordan et Peytrignet (2001), ainsi que de Stalder (2001) dans son deuxième numéro du Bulletin trimestriel de 2001.

Le premier de ces exposés examine en détail la raison pour laquelle la Banque nationale fonde sa politique sur la prévision de l'inflation, en quoi cette prévision influe sur les décisions et comment il convient d'interpréter les prévisions publiées. Il décrit aussi la façon dont la Banque nationale recourt à une prévision qui repose sur un large consensus. Cette prévision se fonde aussi bien sur divers indicateurs que sur des modèles macroéconomiques.

Parmi les modèles que la BNS utilise, il faut mentionner les modèles économétriques traditionnels et les modèles vectoriels autorégressifs (modèles VAR). Les premiers se fondent sur une conception explicite du fonctionnement de l'économie, les relations entre les variables étant modélisées à l'aide d'équations de comportement. Dans le cas des modèles VAR, on part en revanche du principe que la connaissance du fonctionnement de l'économie est limitée. Pour cette raison, aucune structure n'est imposée aux modèles ou cette structure est réduite au minimum. Contrairement aux modèles traditionnels dans lesquels une distinction est faite entre variables exogènes et endogènes, les modèles VAR considèrent toutes les variables comme endogènes.

Stalder (2001) décrit en détail le plus grand modèle structurel traditionnel auquel recourt la BNS. Il s'agit d'un modèle néo-keynésien. Le présent exposé complète la présentation des modèles de prévision de l'inflation utilisés par la BNS par une présentation des modèles VAR développés par la BNS. Ceux-ci comportent deux catégories, à savoir les modèles VAR non structurels et structurels.

Les modèles VAR non structurels se caractérisent par l'absence d'une préconception du fonctionnement de l'économie. En particulier, le mécanisme de transmission des impulsions de politique monétaire n'est pas modélisé explicitement. Les prévisions faites par de tels modèles comportent implicitement les effets d'une politique monétaire qui correspond au comportement historique moyen de la banque centrale. A cette occasion, l'évolution de l'instrument de politique monétaire fait également l'objet d'une prévision. Celle-ci est explicite lorsque l'instrument est une variable du modèle et implicite dans le cas contraire. Comme la politique monétaire prévue correspond simplement au comportement historique moyen de l'institut d'émission et qu'ainsi aucune condition n'est imposée au sujet de son évolution, ces prévisions sont souvent appelées non conditionnelles.

Les modèles VAR non structurels ne permettent pas de faire des prévisions de politique monétaire dépendant d'hypothèses différentes. Les effets d'une politique qui s'écarte du comportement historique moyen de la BNS ne peuvent de même pas être calculés; les modèles VAR non structurels n'identifiant ni le mécanisme de transmission ni la fonction de réaction de la banque centrale. Cette limitation peut être dépassée en transformant des modèles VAR non structurels en modèles structurels (modèles SVAR). En se basant sur un minimum d'a priori quant au fonctionnement de l'économie – c'est-à-dire sur un nombre minimum de restrictions fondées sur des principes économiques généralement acceptés – un modèle SVAR permet de retracer la transmission à l'économie d'une impulsion de politique monétaire. Il permet donc de calculer des prévisions pour diverses hypothèses quant à l'évolution de la politique monétaire. Ces simulations constituent des prévisions conditionnelles. Elles donnent la possibilité de prévoir les variables du modèle sur la base d'une évolution donnée de la politique monétaire, qui peut donc temporairement diverger du comportement historique moyen de l'institut d'émission.

Tant les prévisions conditionnelles que non conditionnelles sont des parties importantes de l'analyse de la politique monétaire. Dans la pratique, les prévisions non conditionnelles forment le point de départ de la discussion. Elles indiquent les perspectives d'inflation lorsque l'institut d'émission se comporte comme par le passé. Afin d'analyser les effets d'hypothèses alternatives de politique monétaire et prendre une décision en la matière, l'institut d'émission a cependant besoin de prévisions conditionnelles.

Le présent exposé se divise en quatre parties. La première est une introduction générale à la méthodologie du VAR. Le lecteur familiarisé avec les modèles VAR peut passer directement à la deuxième partie. Elle montre la façon dont des modèles VAR non structurels permettent d'établir des prévisions non conditionnelles, et leur utilisation à la BNS. La troisième partie présente les modèles VAR structurels, décrit le modèle utilisé par la BNS et montre comment des simulations de politique monétaire et des prévisions conditionnelles peuvent être effectuées avec ce modèle. Les conclusions figurent dans la quatrième partie.

1 Introduction succincte à la méthode VAR

Ce chapitre a pour but de faciliter la compréhension des modèles VAR. La section 1.1 donne un bref aperçu historique, puis passe à la question de l'identification des modèles, question essentielle pour distinguer la méthode VAR des méthodes traditionnelles. La section 1.2 est consacrée à une description générale de ces modèles.

1.1 Contexte historique

Tinbergen a développé dès l'avant-guerre les premiers modèles macroéconomiques de prévision.¹ Ce sont ensuite Koopmans et ses collaborateurs à la *Cowles Foundation* qui ont formulé, dans les années cinquante, la base méthodologique sur laquelle reposent ces modèles.² Ainsi, la méthode économétrique des modèles traditionnels est souvent nommée *Cowles Commission Approach*. Ces modèles sont des systèmes d'équations dynamiques simultanées, qui comportent des équations décrivant le comportement des agents économiques ainsi que des identités comptables.

La méthode de la Commission Cowles se caractérise par une séparation stricte des tâches de la théorie économique et de l'économétrie. D'un côté, la théorie économique définit les équations du modèle, c'est-à-dire fixe les relations entre les variables. Elle détermine notamment quelles variables figurent dans certaines équations et quelles variables en sont exclues. De l'autre côté, il incombe à l'économétrie d'estimer les paramètres (de choisir éventuellement la forme supposée des relations) et de déterminer le processus d'ajustement progressif à l'équilibre (structure des retards échelonnés).

Cette division des tâches a été remise en question au milieu des années septante. A l'époque, les modèles macroéconomiques traditionnels n'ont pas été à même de prévoir la récession mondiale déclenchée par le premier choc pétrolier. Cette situation a démontré que les modèles étaient mal construits, malgré des recherches théoriques menées pendant des années pour les améliorer. La méthode de spécification des modèles macroéconomiques traditionnels a alors été critiquée de façon radicale.³ A partir des années septante, de nouvelles voies se sont ainsi ouvertes à la recherche économétrique.⁴

L'une d'elle consistait à restreindre le rôle de la théorie économique pour spécifier les modèles économétriques. Les modèles basés sur des techniques

1 Dès 1936, Tinbergen a présenté un modèle de 27 équations pour les Pays-Bas et un modèle de 84 équations pour les Etats-Unis.

2 Voir Koopmans et Hood (1953).

3 Voir Liu (1960) qui a fait œuvre de précurseur et Sims (1980).

4 Voir un aperçu chez Kirchgässner et Savioz (1997).

de séries temporelles étaient particulièrement appropriés à cet objectif, puisqu'ils ne nécessitent que peu de théorie économique. La recherche dans le domaine de l'économétrie des séries temporelles en a été stimulée jusqu'à aujourd'hui. Dès 1970, Box et Jenkins ont développé les modèles de séries temporelles ARIMA. En 1975, Granger et Newbold ont démontré que de petits modèles de séries temporelles ARIMA ont souvent fourni de meilleures prévisions que les grands et onéreux modèles économétriques traditionnels d'alors. Dans les années quatre-vingt, on est parvenu à des percées décisives dans la conception de l'économétrie des séries temporelles, notamment par les travaux de Dickey et Fuller (1981) sur les variables intégrées, d'Engle et Granger (1987) sur les variables cointégrées, de Sims (1980) sur les modèles de séries temporelles multivariées (modèles VAR) et de Johansen (1988) sur le recours aux relations de cointégration dans les modèles multivariés de séries temporelles. Ces méthodes ont été développées en particulier pour le traitement des questions liées à la politique monétaire.

En substance, la critique de Sims (1980) à l'encontre des modèles macroéconomiques traditionnels porte sur la manière dont ceux-ci sont identifiés. Indépendamment de l'approche économétrique choisie, un modèle doit être *identifié* de manière à permettre une interprétation économique de ses résultats. Des prévisions conditionnelles ne sont ainsi possibles qu'après identification du modèle. Habituellement, dans les manuels d'économétrie le problème de l'identification est illustré à l'aide du marché d'un produit agricole tel que le café. L'offre de café aux Etats-Unis dépend des conditions météorologiques du Brésil. Toutefois, les conditions météorologiques ne jouent guère de rôle dans la demande de café aux Etats-Unis. Cette situation permet d'exclure un facteur, les conditions météorologiques, de l'équation qui détermine la demande de café. Si les conditions météorologiques changent en Amérique du Sud, l'équilibre du marché se déplace le long de la courbe de la demande de café. Celle-ci peut donc être *identifiée* et ses paramètres peuvent être déterminés empiriquement. Il devient alors possible de faire des prévisions du prix du café conditionnelles à un changement météorologique. Si non seulement l'offre, mais aussi la demande de café dépendaient du temps en Amérique du Sud, les variations de prix et de quantités le long de la courbe de demande ne se distingueraient plus de celles le long de la courbe d'offre. Les variations de prix et de quantités observées ne se laisseraient plus traduire en terme de déplacements

de l'équilibre le long de la courbe de demande (notion d'identification de la courbe de demande). La pente de la courbe de demande serait inconnue et les prévisions conditionnelles aux conditions météorologiques deviendraient difficiles voire impossibles.⁵

Selon l'approche de la Commission Cowles, l'identification s'effectue grâce à des *restrictions d'exclusion*. Par une telle restriction, le coefficient d'une variable explicative potentielle est fixé à zéro dans l'équation qui détermine une variable endogène du modèle. Ainsi, toute valeur explicative est déniée à cette variable dans la détermination de la variable endogène. L'identification a ainsi lieu simultanément à la spécification du modèle, en ayant recours aux connaissances de la théorie économique. Sims (1980) taxe de *peu plausible* l'identification d'après la méthode de la Commission Cowles. Il soutient que l'on ne peut guère faire appel à la théorie économique pour justifier de telles restrictions d'exclusion. Au contraire, la théorie de l'équilibre général décrit l'économie comme un système dans lequel tout est lié. L'identification d'une équation de demande en donne un exemple. Il est difficile d'identifier une équation seule, car toutes les équations de demande devraient théoriquement présenter les prix de tous les biens comme variables explicatives. Si des restrictions d'exclusion apparaissent néanmoins dans un système d'équations de demande, Sims suppose que l'économètre recourt alors davantage à sa propre version d'une théorie psychologique ou sociologique qu'à la théorie économique.⁶

Le rôle des anticipations en économie est une autre raison qui conduit à remettre en question l'identification selon l'approche de la Commission Cowles. Sims recourt précisément à l'exemple du marché du café pour démontrer que, dans le cas d'anticipations rationnelles, les restrictions d'exclusion des modèles économétriques traditionnels ne sont pas adéquates et les modèles ne sont donc pas identifiés. Des conditions météorologiques défavorables au Brésil peuvent en effet entraîner des achats de café par les consommateurs et les courtiers qui anticipent la hausse des prix. Si l'on part d'anticipations rationnelles tant la courbe d'offre que la courbe de demande dépendent donc du temps au Brésil. Il en résulte que les variations de prix et de quantités utilisées par l'économètre pour estimer la valeur des paramètres de la courbe de demande n'ont pas nécessairement lieu le long de celle-ci. L'équation estimée n'est pas véritablement une courbe de demande de café et ne doit pas être considérée comme *structurelle*. En fait, elle résulte de l'action combinée des

5 Voir une discussion plus détaillée chez Sims (1980).

6 Voir Sims (1980), page 3.

courbes d'offre et de demande qui se déplacent simultanément. Elle constitue de ce fait ce que les économètres appellent une *forme réduite* du modèle d'offre et de demande. Par conséquent, les prévisions conditionnelles calculées sur la base d'une équation de demande estimée ainsi sont erronées.

Le développement de l'approche VAR par Sims (1980) est une réponse aux difficultés susmentionnées. Sims a montré qu'une méthode de spécification et d'identification moins problématique était possible dans le cadre des modèles vectoriels autorégressifs.

1.2 Bases théoriques des modèles VAR

Un modèle VAR(p) peut être décrit de la manière suivante:

$$(1) \quad y_t = D_1 y_{t-1} + D_2 y_{t-2} + \dots + D_p y_{t-p} + \varepsilon_t, \\ E(\varepsilon_t \varepsilon_t') = \Omega,$$

où y_t constitue un vecteur de n variables endogènes au moment t . Le vecteur peut, par exemple, inclure le taux d'inflation, la variation des taux d'intérêt, l'expansion de la masse monétaire ou d'autres variables:

$$y_t = \begin{bmatrix} \text{taux d'inflation,} \\ \text{variation des taux d'intérêt,} \\ \text{expansion de la masse monétaire,} \\ \dots \end{bmatrix}.$$

Les matrices D_i , pour $i = 1, \dots, p$, contiennent les coefficients du modèle. L'ordre p du modèle VAR indique combien de vecteurs retardés y_{t-i} figurent dans le modèle. Le vecteur ε_t représente les influences non systématiques, dénommées également impulsions ou innovations. La matrice de variance-covariance Ω indique la structure de corrélation des influences non systématiques ε_t .⁷

Dans le cas du modèle VAR bivarié d'ordre 1 que nous utiliserons plusieurs fois à titre d'illustration, l'équation (1) devient:

$$y_t = D_1 y_{t-1} + \varepsilon_t \\ \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \\ = \begin{bmatrix} d_{11} y_{1t-1} + d_{12} y_{2t-1} + \varepsilon_{1t} \\ d_{21} y_{1t-1} + d_{22} y_{2t-1} + \varepsilon_{2t} \end{bmatrix}.$$

Le vecteur des variables endogènes ($y_t = (y_{1t} \ y_{2t})'$) s'explique par le modèle (les coefficients de la matrice D_1), la position initiale ($y_{t-1} = (y_{1t-1} \ y_{2t-1})'$) de la période précédente et les influences non systématiques ($\varepsilon_t = (\varepsilon_{1t} \ \varepsilon_{2t})'$). Chaque variable dépend de l'évolution passée de l'ensemble des variables. Les relations simultanées entre les variables, c'est-à-dire les effets réciproques qui se manifestent au cours de la période, seront saisies par la matrice de variance-covariance suivante:

$$\Omega = \begin{bmatrix} \text{var}(\varepsilon_{1t}) & \text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) \\ \text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) & \text{var}(\varepsilon_{2t}) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

Supposons que la première variable du vecteur y_t représente le taux d'inflation et la seconde le taux d'intérêt déterminé par la politique monétaire. Com-

7 Voir également Stock et Watson (2001) pour une description simple de la méthode VAR. Afin de simplifier la notation, les variables sont exprimées en déviation par rapport à la tendance. Pour la même raison, la notation ne distingue pas les paramètres véritables des paramètres estimés.

ment faut-il, dans ce cas, interpréter les coefficients de la matrice D_1 et Ω ? Une innovation du taux d'intérêt ne se répercute directement sur l'inflation que si la covariance avec l'innovation de l'inflation est différente de zéro: $\text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) \neq 0$. Si l'inflation a sa propre dynamique, le premier coefficient de la première ligne de la matrice (d_{11}) est différent de zéro. L'inertie du taux d'inflation dépend de la taille de ce coefficient. Si le coefficient d_{11} est proche de un, une inflation forte restera forte. Si la politique monétaire influe sur l'inflation avec retard, le coefficient d_{12} doit être différent de zéro. L'ampleur de l'effet dépend de la taille du coefficient. Ce retard peut provenir du fait qu'un changement du taux d'intérêt ne se répercute sur les prix qu'après un certain temps. Toutefois, il peut aussi être imputable au fait que la politique monétaire fixe le taux d'intérêt au moment t en anticipation de l'inflation au moment $t+1$.

Dans une économie, la relation entre l'inflation et le taux d'intérêt est trop complexe pour être capturée par les quatre coefficients de la matrice D_1 . Néanmoins, les interrelations dynamiques complexes que l'on observe dans une économie peuvent être décrites par des modèles VAR comportant plusieurs variables et d'ordre p supérieur à un.

Revenons à la partie systématique du modèle (1) VAR(p):

$$(2) \quad \hat{y}_t = D_1 y_{t-1} + D_2 y_{t-2} + \dots + D_p y_{t-p}$$

Les coefficients des matrices D_i $i=1,2,\dots,p$ ne doivent pas être interprétés comme des paramètres de comportement structurels, car ils sont des fonctions complexes de coefficients de comportement sous-jacents. La composante systématique \hat{y}_t du modèle VAR(p) représente la prévision de la variable y_t . Cette prévision ne se base que sur des informations disponibles au moment $t-1$, à savoir la position initiale y_{t-i} , $i=1,\dots,p$ et les matrices D_i $i=1,2,\dots,p$. Ainsi, le vecteur des influences non systématiques ε_t représente l'erreur de prévision

$$y_t - \hat{y}_t = \varepsilon_t$$

du modèle. Le vecteur des influences non systématiques ε_t rassemble les informations qui se rapportent aux variables pour la période en cours t , et qui ne sont pas prévisibles par le modèle. C'est pourquoi les composantes de ce vecteur sont nommées innovations. Dans la méthode VAR, ces innovations figurent au centre de l'analyse. Si, comme ci-dessus dans le modèle VAR(1), la seconde variable du vecteur y_t représente le taux d'intérêt, l'innovation de la seconde équation de (1) représente la variation non

prévisible du taux d'intérêt. Un changement de politique monétaire inattendu constitue une telle variation imprévisible: La politique monétaire se fonde sur des informations disponibles *durant* le trimestre en cours t , tandis que le taux d'intérêt attendu \hat{y}_{2t} se rapporte à des informations de la période précédente $t-1$. Une modification imprévue du taux d'intérêt peut également provenir d'un changement inattendu de l'inflation.

L'équation (1) illustre qu'au moment t , les variables endogènes y_t dépendent de leur passé, y_{t-i} pour $i=1,2,\dots,p$, et de l'innovation en cours ε_t . Il en va ainsi de toutes les périodes précédentes. En raison du caractère récursif de l'équation (1), le vecteur des variables endogènes peut être représenté uniquement comme une fonction des innovations passées:⁸

$$(3) \quad y_t = \varepsilon_t + C_1 \varepsilon_{t-1} + C_2 \varepsilon_{t-2} + \dots$$

L'équation (3) forme le noyau de la méthode VAR. Elle représente le vecteur de variables endogènes y_t comme la moyenne pondérée des innovations actuelles et antérieures. Il s'agit de la représentation «moyenne mobile» (*Vector Moving Average* ou VMA) du modèle. L'influence des innovations ε_{t-i} sur le vecteur de variables endogènes y_t , i périodes dans l'avenir, est décrite par la matrice des coefficients C_i . Les matrices C_i indiquent les réactions des variables y_t aux innovations et sont des fonctions complexes des matrices D_i , $i=1,2,\dots,p$.

Toutefois, les matrices C_i sont difficiles à interpréter s'il existe une corrélation entre les innovations: il n'est guère judicieux de considérer séparément la réaction d'une variable à une innovation déterminée s'il faut prévoir que le changement de cette innovation entraînera simultanément un changement de toutes les innovations corrélées avec elle. Pour résoudre ce problème, Sims a proposé simplement et pragmatiquement de représenter de manière récursive la corrélation des innovations, méthode devenue routinière dans les analyses VAR. Dans ce contexte, l'innovation de la première variable du système VAR, ε_{1t} , est interprétée comme un changement exogène. Le deuxième changement exogène est le résidu de la régression de l'innovation de la deuxième variable ε_{2t} sur la première. Si nous appliquons cette méthode récursivement jusqu'à la $n^{\text{ème}}$ innovation, ε_{nt} , nous obtenons la représentation des innovations comme fonction linéaire récursive de n variables non corrélées, désignés par u ci-après:

8 La stabilité du modèle VAR est considérée comme une condition.

$$\begin{aligned}
 \varepsilon_{1t} &= u_{1t}, \\
 \varepsilon_{2t} &= a_{21}u_{1t} + u_{2t}, \\
 (4) \quad &\dots \\
 &\dots \\
 \varepsilon_{nt} &= a_{n1}u_{1t} + \dots + a_{n,n-1}u_{n-1t} + u_{nt}.
 \end{aligned}$$

En remplaçant les innovations de l'équation (3) par les chocs non corrélés de l'équation (4), nous obtenons une représentation du modèle qu'il est possible d'interpréter directement. Celle-ci indique les effets des chocs autonomes u sur les variables du modèle (fonction de réponse aux impulsions). La matrice triangulaire A_0 des coefficients de (4) peut également être calculée par une décomposition de Choleski de la matrice Ω des covariances des innovations. Un modèle VAR dont les innovations sont décomposées selon (4) est appelé modèle VAR *récuratif*.

2 Prévisions de base par modèle VAR

Ce chapitre est consacré aux prévisions qui peuvent être calculées par des modèles VAR non structurels. La section 2.1 contient quelques remarques préalables relatives à l'élaboration de prévisions VAR et au rôle que ces prévisions peuvent jouer dans la politique monétaire. Dans la section 2.2, nous examinons la spécification des modèles VAR non structurels utilisés par la BNS. La section 2.3 explique comment combiner les prévisions de divers modèles VAR pour améliorer les prévisions.

2.1 Quelques remarques préalables

Les prévisions d'inflation calculées grâce à un modèle VAR non structurel s'expliquent le plus simplement au moyen d'un modèle VAR(1).

$$\begin{aligned}
 y_t &= D_1 y_{t-1} + \varepsilon_t, \\
 E[\varepsilon_t \varepsilon_t'] &= \Omega.
 \end{aligned}$$

Si l'on nomme le présent T , l'état de l'économie aux moments $T+1$ et $T+2$ est donné par l'expression

$$y_{T+1} = D_1 y_T + \varepsilon_{T+1}$$

et

$$\begin{aligned}
 y_{T+2} &= D_1 y_{T+1} + \varepsilon_{T+2} \\
 &= D_1 (D_1 y_T + \varepsilon_{T+1}) + \varepsilon_{T+2} \\
 &= D_1^2 y_T + D_1 \varepsilon_{T+1} + \varepsilon_{T+2}
 \end{aligned}$$

Les innovations futures ε_{T+1} et ε_{T+2} sont inconnues au moment T où la prévision est établie. L'espérance mathématique de ces innovations est donc égale à zéro:

$$E_T(\varepsilon_{T+1}) = E_T(\varepsilon_{T+2}) = 0.$$

Ainsi, les prévisions d'inflation des moments $T+1$ et $T+2$ sont:

$$\begin{aligned}
 \hat{y}_{T+1} &= E_T(y_{T+1}) = D_1 y_T, \\
 \hat{y}_{T+2} &= E_T(y_{T+2}) = D_1^2 y_T
 \end{aligned}$$

et la prévision du moment $T+i$ est

$$(5) \quad \hat{y}_{T+i} = D_1^i y_T.$$

Si les prévisions sont destinées à la politique monétaire, l'horizon de prévision doit dépasser le délai moyen à partir duquel les instruments de politique monétaire agissent. La Banque nationale suisse se fonde sur l'idée que la majeure partie de l'effet d'un choc de politique monétaire se manifeste dans

un délai de trois ans. La période de prévision est donc de $i = 12$ trimestres et la prévision la plus longue est, par conséquent,

$$\hat{y}_{T+12} = D_1^{12} y_T.$$

Dans les prévisions calculées par des modèles VAR non structurels, la politique monétaire n'est pas déterminée par un scénario, mais correspond au comportement historique moyen de la banque centrale. Contrairement aux prévisions conditionnelles, basées sur certaines hypothèses relatives à l'évolution future de la politique monétaire, les prévisions de base sont non conditionnelles.⁹ Si la période de prévision i est assez longue, les prévisions de modèles VAR non structurels, comme les prévisions non conditionnelles de tout modèle économétrique défini de façon appropriée, indiquent toujours une convergence avec la moyenne historique des variables lorsque les variables sont stationnaires (voir l'équation (5)).¹⁰ Les prévisions de ces modèles sont néanmoins intéressantes à court et moyen terme. Premièrement, elles forment une première base pour juger les besoins d'action en politique monétaire. Les prévisions \hat{y}_{T+i} indiquent l'évolution attendue en prenant en compte l'état initial de l'économie et le comportement passé moyen de la banque centrale en matière de politique monétaire. Ainsi, les prévisions non conditionnelles révèlent si le comportement habituel en matière de politique monétaire suffit à maintenir à moyen terme l'inflation dans le cadre souhaité.

Deuxièmement, les prévisions de modèles VAR non structurels sont athéoriques. Ainsi, l'évaluation d'un éventuel besoin d'agir ne se base pas sur un modèle théorique spécifique. L'évaluation est ainsi indépendante d'une préconception quant à la structure et à la manière de fonctionner de l'économie. En raison de leur caractère athéorique, ces prévisions servent fréquemment de *référence* dans la comparaison avec d'autres.

9 Il faut être conscient que chaque prévision est conditionnelle à certains égards. Même les prévisions des modèles VAR non structurels dépendent des informations disponibles. On préfère donc employer la notion de prévision de base (baseline forecast) à celle de prévision non conditionnelle. (Voir à ce sujet Canova (1995), page 100).

10 La raison de la convergence est que le modèle est estimé par les données d'un régime de politique monétaire dans lequel l'inflation est l'objectif prioritaire de la politique monétaire. De plus, l'inflation historique moyenne ne devrait pas s'écarter fortement du niveau souhaité. Ces conditions impliquent que la situation soit stable et que $\lim_{j \rightarrow \infty} D_1^j = 0$.

2.2 Spécification des modèles VAR

Pour ses prévisions de base, la Banque nationale ne recourt pas à un seul modèle VAR. Elle se base sur ceux qui ont fourni récemment les meilleures prévisions et calcule une prévision combinée de ces modèles.

La spécification d'un modèle VAR repose sur quatre décisions: (i) sélection des variables, (ii) détermination de l'ordre du modèle VAR, (iii) choix de composantes déterministes éventuelles et (iv) traitement des caractéristiques tendanciennes des variables (degré d'intégration et propriétés de cointégration).

L'ordre des modèles VAR non structurels utilisés par la BNS est de quatre ou cinq. Cet ordre suffit pour saisir la dynamique des variables et, partant, de disposer d'équations débarrassées d'une éventuelle autocorrélation des résidus. Chaque équation VAR comprend une constante et des variables muettes (en cas de données non corrigées des variations saisonnières). Aucune composante déterministe additionnelle n'est insérée dans les modèles VAR. Les modèles dont les variables sont en première différence ne tiennent pas compte des tendances et des relations de cointégration éventuelles entre variables intégrées; tous les modèles sont ainsi estimés aussi bien en niveau qu'en première différence.¹¹

Le choix des variables entrant dans les modèles se fait en plusieurs étapes. Un groupe de variables potentiellement informatives quant aux perspectives futures de l'inflation est tout d'abord déterminé. Selon les connaissances théoriques concernant le processus de transmission, les études empiriques ainsi que les recherches internes à la BNS,¹² 11 variables au moins font partie de ce groupe. Outre de l'inflation, mesurée par l'indice des prix à la consommation (IPC), il s'agit du PIB en terme réel, des masses monétaires M_1 , M_2 et M_3 , des crédits bancaires en Suisse, des indices de change nominal et réel pondérés en fonction du commerce extérieur, du taux d'intérêt à court et à long terme ainsi que la différence entre ces deux taux. Les résultats de tests de stationnarité montrent que les premières différences de ces variables peuvent être considérées comme stationnaires.

Des tests de causalité de Granger donnent une première impression de la valeur informative de ces variables quant à l'inflation future. Ils indiquent si l'adjonction d'une des dix variables améliore la prévision de l'inflation, qui se fonde uniquement sur l'évo-

11 Voir également Sims, Stock et Watson (1990).

12 Voir Jordan (1999).

lution passée de celle-ci. Si tel est le cas la variable en question comprend des informations sur l'inflation future. On dit alors qu'il s'agit d'une cause de l'inflation selon le concept de causalité de Granger. Les résultats des tests de causalité de Granger sont résumés au tableau 2.1.

Les résultats démontrent que la plupart des variables sont informatives sur l'évolution future de l'inflation et peuvent être utilisées pour des prévisions. Pris isolément, les cours du change et l'écart entre les taux d'intérêt ne causent pas l'inflation au sens de Granger. Toutefois, l'expérience indique que les cours de change et l'écart entre les taux d'intérêt combinés à d'autres variables peuvent améliorer les prévisions d'inflation.

Toute phase de prévision débute par un examen préalable des caractéristiques de prévision de tous les modèles VAR qui peuvent être formés à partir du groupe des variables informatives (5 variables par modèle au maximum). Cet examen préalable a lieu en deux étapes:

1. Les prévisions d'inflation annuelle sont calculées pour des horizons prévisionnels allant jusqu'à trois ans. Ces prévisions se rapportent au passé récent. Elles sont établies à l'extérieur de l'échantillon des données utilisées pour l'estimation des modèles VAR (*out-of-sample forecast*).
2. Les prévisions établies à l'extérieur de l'échantillon sont évaluées d'après le critère des écarts quadratiques moyens (RMSE), et les meilleurs modèles VAR sont sélectionnés.¹³

Quatre caractéristiques des prévisions à l'aide de modèles VAR¹⁴ se sont confirmées depuis que, au début de 2000, la nouvelle stratégie de politique monétaire est entrée en vigueur:

1. Le RMSE augmente avec la longueur de la période de prévision. Avec des données corrigées des variations saisonnières et pour un horizon prévisionnel de un an, les meilleures prévisions ont un RMSE de moins de 1,0 point de pourcentage du taux d'inflation. Pour une période de prévision de trois ans, les meilleurs modèles VAR ont un RMSE inférieur à 1,5 points de pourcentage.
2. Pour des prévisions à long terme, les modèles VAR à variables en niveau donnent de meilleurs résultats que ceux à variables en première différence.
3. Les meilleures prévisions sont réalisées avec des modèles VAR contenant au maximum de trois à quatre variables.¹⁵
4. Les variables qui sont le plus fréquemment sélectionnées dans les modèles VAR qui produisent les meilleures prévisions à long terme sont: les crédits, les taux d'intérêt et les agrégats monétaires.

La sélection des modèles qui seront employés pour les prévisions d'inflation se fait sur la base d'un examen préalable. Cet examen sert en outre à détecter d'éventuelles ruptures de structure. Comme la sélection des modèles est renouvelée chaque trimestre, les modèles ayant les meilleures qualités de prévision après une éventuelle rupture de structure devraient être retenus. Il a été constaté jusqu'à maintenant que le groupe des meilleurs modèles VAR sélectionnés ne change que rarement.

Test de causalité de Granger

Tableau 2.1

Variable	Retards = 4		Retards = 6	
	Statistique F	Valeur p	Statistique F	Valeur p
Produit intérieur brut réel	2,12	0,085*	1,87	0,096*
Masse monétaire M ₁	3,42	0,012**	2,08	0,064*
Masse monétaire M ₂	3,58	0,009***	1,98	0,078*
Masse monétaire M ₃	1,58	0,187	1,92	0,087*
Crédits bancaires en Suisse	3,51	0,010***	2,06	0,067*
Cours du change nominal pondéré en fonction du commerce extérieur	1,57	0,189	1,47	0,199
Cours du change réel pond. suivant le commerce extérieur	1,47	0,219	1,37	0,235
Taux court nominal	2,46	0,051*	2,46	0,030**
Taux long nominal	4,80	0,001***	3,41	0,005***
Ecart entre les taux long et court	0,98	0,425	1,43	0,211

On a recouru à des variables en première différence avec un retard maximum de 4 et de 6 trimestres. L'hypothèse testée est celle d'une non-causalité de Granger. Le rejet de l'hypothèse testée est marqué d'un astérisque. Une, deux ou trois astérisques signifient que l'hypothèse testée est rejetée avec un seuil de signification de respectivement 10%, 5% et 1%. Le test a lieu par couples de variables et sans tenir compte de relations de cointégration éventuelles.

13 Le RMSE est la racine carrée de l'erreur quadratique moyenne de prévision. Il n'est pas possible de comparer sans autre la qualité des prévisions extérieures à l'échantillon d'un modèle VAR avec celle d'un modèle économétrique traditionnel à l'aide du critère RMSE.

Les prévisions VAR étant complètement dynamiques, le RMSE d'un modèle économétrique traditionnel ne peut être comparé avec celui d'un modèle VAR que si les variables exogènes intégrées dans le modèle économétrique traditionnel font aussi l'objet de prévisions.

14 Voir Jordan (1999).

15 Le manque relatif de précision des prévisions réalisées à l'aide des modèles à cinq variables pourrait provenir du manque de degrés de liberté dans l'estimation de tels modèles.

2.3 Combinaison de prévisions à partir de divers modèles VAR

Après la sélection des meilleurs modèles VAR, les prévisions calculées grâce à ces modèles sont combinées en une seule. Supposons par exemple que, lors de l'examen préalable, les trois modèles VAR1, VAR2 et VAR3 aient été sélectionnées pour l'horizon prévisionnel i . Les modèles seront réestimés tout d'abord en incluant les données les plus récentes. Ensuite, les prévisions $\hat{\pi}_{VAR1, T+i}$, $\hat{\pi}_{VAR2, T+i}$ et $\hat{\pi}_{VAR3, T+i}$ seront calculées pour l'inflation i périodes dans le futur. La prévision combinée correspond alors à la moyenne pondérée de ces prévisions:

$$\hat{\pi}_{T+i} = w_1 \hat{\pi}_{VAR1, T+i} + w_2 \hat{\pi}_{VAR2, T+i} + w_3 \hat{\pi}_{VAR3, T+i}$$

dans laquelle w_1 , w_2 et w_3 sont les pondérations attribuées aux diverses prévisions. Il existe différentes méthodes pour pondérer les prévisions. La méthode de la moyenne simple attribue la même pondération à chaque prévision; le total des pondérations étant de un. Une méthode alternative consiste à estimer ces pondérations à l'aide des prévisions calculées lors de l'examen préalable. Dans la méthode des moindres carrés, les pondérations sont calculées grâce à une régression du taux d'inflation réalisé sur les diverses prévisions VAR. Actuellement, la BNS utilise la méthode de la moyenne simple. Ce processus est appliqué aux horizons prévisionnels $i = 2, 4, 6, 8, 10, 12$. Les prévisions des meilleurs modèles sont combinées pour chaque horizon i .

Le gain de précision atteint par les prévisions combinées découle de la diversification. Exception faite d'une corrélation parfaitement positive entre les différentes erreurs de prévision, une prévision faite sur la base d'une moyenne pondérée comporte une variance d'erreur inférieure à la moyenne de celles des diverses prévisions. Cet effet de diversification ne se manifeste que pour des prévisions dont l'espérance mathématique des erreurs de prévisions est nulle. Comme les modèles VAR n'ont été soumis à aucune restriction, cette propriété peut être admise. Le fait que l'erreur de prévision ait une variance inférieure signifie également que le RMSE des prévisions combinées est inférieur au RMSE moyen des diverses prévisions dont il résulte.¹⁶

Outre une précision accrue, les prévisions combinées ont les grands avantages suivants:

1. La combinaison des prévisions permet d'obtenir une *agrégation des informations*. Alors que les prévisions VAR ne tiennent compte que d'un petit nombre de variables informatives, les prévisions combinées peuvent assouplir cette restriction.
2. La combinaison de prévisions évite de se concentrer sur un seul modèle VAR. Ainsi, le problème de l'*insécurité quant à la spécification correcte du modèle* est atténué.
3. L'estimation des pondérations de la combinaison – au moyen de la méthode des moindres carrés par exemple – indique quelles prévisions VAR comportent des informations qui ne sont pas disponibles dans les autres. En outre, un changement de pondération signale une rupture de structure.

Jordan et Savioz (2001) montrent que le RMSE des prévisions d'inflation peut être réduit considérablement par des combinaisons. La méthode de la moyenne simple restreint de plus de 10% le RMSE des prévisions d'une durée d'une année. Pour les horizons prévisionnels de deux et de trois ans, la méthode la plus adéquate est celle des moindres carrés. La réduction du RMSE se chiffre de nouveau à environ 10%. Ces indications se rapportent à une comparaison du RMSE moyen des prévisions VAR avec le RMSE moyen des prévisions VAR combinées. En revanche, si l'on compare les meilleures prévisions VAR avec les meilleures prévisions VAR combinées, l'efficacité est encore accrue. Pour les prévisions d'inflation dont l'horizon s'étend à deux ou trois ans, le RMSE moyen des meilleures prévisions VAR combinées pondérées par la méthode des moindres carrés est inférieur de 30% au RMSE moyen des meilleurs modèles VAR.

Sur le graphique 2.1 figurent des prévisions de modèles VAR non structurels pour une période allant du deuxième trimestre 2000 au quatrième trimestre 2001. Seules les informations disponibles au moment de chaque prévision ont été utilisées. Les prévisions ont été faites au moyen du processus en deux étapes décrit ci-dessus.

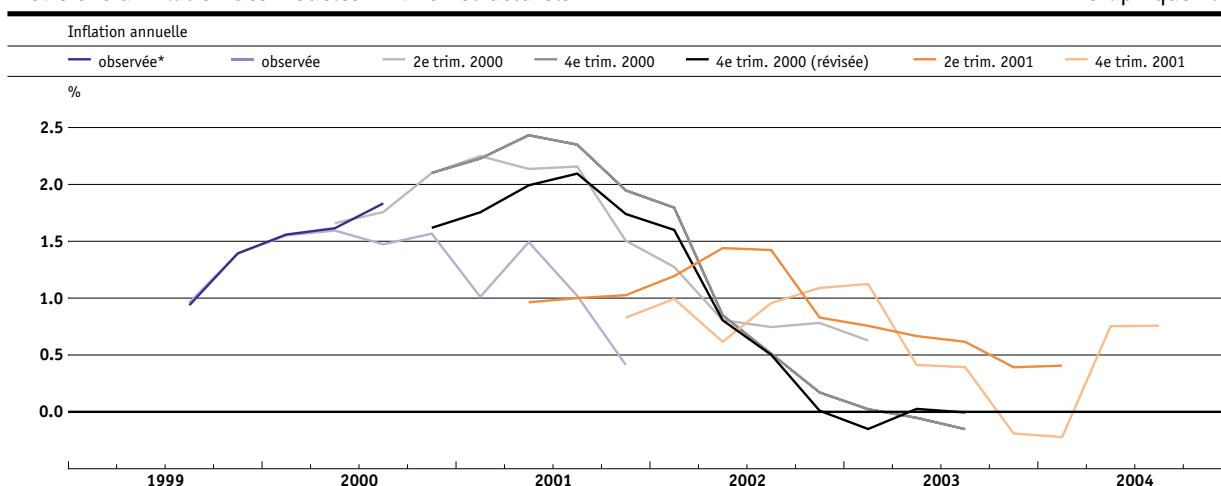
16 Le RMSE est la somme de la variance de l'erreur de prévision et du biais au carré de la prévision. Pour des prévisions sans biais, le RMSE est donc fonction de la variance de l'erreur de prévision.

Le graphique représente l'inflation annuelle calculée selon l'IPC désaisonnalisé. L'IPC ayant été corrigé d'une erreur de mesure au quatrième trimestre 2000, l'inflation annuelle est indiquée telle que calculée avant cette date. On constate sur le graphique 2.1 que les prévisions ont rapidement indiqué un net ralentissement de l'inflation. Le fait que les prévisions aient dépassé la limite des 2% au début de 2001, s'explique partiellement par l'erreur de mesure de l'IPC. Cela se constate clairement lorsque l'on compare la prévision révisée sur la base de l'IPC corrigé avec la prévision calculée avant la correction. Il est frappant que les prévisions VAR aient indiqué très tôt l'affaiblissement de l'inflation, que l'évolution de l'inflation observée n'indiquait pas avant le deuxième trimestre 2001.

Comme susmentionné, les prévisions des modèles VAR non structurels convergent vers une inflation basse si l'estimation résulte de données issues d'un régime de politique monétaire axé sur le maintien de la stabilité des prix. Par conséquent, l'information la plus importante fournie par ces prévisions est non pas la convergence vers un taux d'inflation faible, mais la rapidité et la manière dont l'adaptation au niveau historique de l'inflation se fait. Dans un horizon de trois ans, les prévisions du graphique 2.1 convergent légèrement au-dessous de 1%. Si en revanche sur l'horizon de prévision de trois ans, les prévisions d'inflation n'indiquaient aucune tendance à la convergence vers la stabilité des prix ou indiquaient même une évolution divergente la politique monétaire devrait dévier du comportement historique.

Prévisions d'inflation des modèles VAR non structurels

Graphique 2.1



* ancien indice IPC

3 Modèles VAR structurels (SVAR)

Dans ce chapitre, nous nous demanderons comment les modèles VAR permettent d'analyser différents scénarios de politique monétaire. Les prévisions de base ne peuvent visiblement pas être utilisées à cette fin, puisque celles-ci impliquent une politique économique correspondant à la moyenne historique. Toutefois, en recourant à quelques connaissances théoriques concernant les effets de la politique monétaire sur l'économie, cette question peut être examinée à l'aide des prévisions conditionnelles de modèles structurels VAR (SVAR). Dans les sections 3.1 et 3.2, nous décrivons les aspects fondamentaux de la méthode. Le modèle SVAR utilisé par la BNS et les différents types de prévisions conditionnelles obtenues avec ce modèle seront traités aux sections 3.3 et 3.4.

3.1 Simulations de politique monétaire

Cette section examine comment les modèles VAR ou plus précisément les modèles SVAR permettent d'effectuer des simulations de politique monétaire. Concrètement, il s'agit d'effectuer des prévisions sur les variables d'un modèle VAR qui soient subordonnées à certaines conditions. Une telle condition peut être l'exigence qu'un instrument contrôlé par la politique monétaire suive une trajectoire prédéterminée. Une autre condition envisageable est qu'une variable cible atteigne à un moment précis dans le futur une valeur prédéterminée. De telles prévisions conditionnelles peuvent être utilisées pour évaluer l'effet de décisions alternatives de politique monétaire, en examinant par exemple les conséquences dans le temps de la fixation à différents niveaux du taux d'intérêt sur les prévisions d'inflation, de la croissance de la production ou de la masse monétaire.

Lors de calculs de prévisions conditionnelles par des modèles VAR, il est nécessaire de tenir compte de la corrélation des variables entre elles à l'intérieur d'une même période. Nous allons l'illustrer au moyen du modèle VAR(1). La forme réduite de ce modèle correspond à

$$y_t = Dy_{t-1} + \varepsilon_t, \\ E(\varepsilon_t \varepsilon_t') = \Omega.$$

où $t=1,2,\dots,T$, le vecteur $y_t = (y_{1t}, y_{2t})'$ comprenant les variables à prévoir. Par estimation, nous connaissons

la matrice de coefficients D et la matrice de variance-covariance des résidus Ω . Les deux matrices ont la dimension 2×2 . Soulignons que les innovations ε_t sont généralement corrélées entre elles, c'est-à-dire que Ω n'est pas diagonale. Les éléments essentiels permettant de calculer la prévision conditionnelle peuvent être illustrés à l'aide d'une prévision en une étape. Partant du moment actuel T , la prévision non conditionnelle en une étape de y_{T+1} se formule dans le cadre de ce modèle:

$$\hat{y}_{T+1} = Dy_T$$

Par conséquent, l'erreur de prévision est de ε_{T+1} et correspond au vecteur des innovations au moment $T+1$. Dans la prévision *non conditionnelle* basée sur des modèles VAR, ε_{T+1} (et, le cas échéant, l'erreur de prévision pour des prévisions à plus long terme) est posée comme égale à la valeur de l'espérance mathématique, soit équivalente à zéro. Ce faisant, on part de l'idée qu'aucune information sur les innovations ne sera disponible au-delà du moment T .

Si la politique monétaire peut contrôler une des variables de y_t , tel le taux d'intérêt y_{2t} , on peut se demander comment la prévision de l'autre, à savoir y_{1t} (l'inflation par exemple), changera lorsque l'évolution future de la variable de politique monétaire sera prédéterminée. Pour répondre à cette question, il suffit, en partant du moment T , de modifier la politique monétaire de façon à ce que la variable de politique monétaire $y_{2, T+1}$ atteigne la valeur visée. Puis on examinera comment cela influe sur la prévision de l'autre variable. De telles prévisions conditionnelles permettent alors d'évaluer des scénarios alternatifs de politique monétaire.

On peut illustrer cette stratégie simplement à l'aide du modèle à deux variables pour le moment $T+1$:

$$\begin{bmatrix} y_{1, T+1} \\ y_{2, T+1} \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} y_{1, T} \\ y_{2, T} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1, T+1} \\ \varepsilon_{2, T+1} \end{bmatrix}.$$

Si $y_{2,t}$ est la variable de politique monétaire, il faut choisir $\varepsilon_{2, T+1}$ de façon que $y_{2, T+1}$ ($y_{1, T}$ et $y_{2, T}$ étant donnés) atteigne la valeur souhaitée $y_{2, T+1}^*$. Relevons cependant que $\varepsilon_{1, T+1}$ ne peut pas être posé simplement comme égal à zéro. Comme les éléments de ε_t sont généralement corrélés, il convient de tenir compte de la corrélation historique entre les innovations et de fixer $\varepsilon_{1, T+1}$ en conséquence. En d'autres termes, l'information passée quant à la corrélation simultanée des deux éléments de y_t peut être utilisée pour améliorer la prévision de $y_{1, T+1}$ si $y_{2, T+1}$ est connu.

Les innovations de la variable de politique monétaire $\varepsilon_{2,t}$ s'expliquent partiellement par des réactions systématiques de la politique monétaire aux variations inattendues des autres variables et partiellement par des chocs de politique monétaire. Or seuls les chocs de politique monétaire jouent un rôle pour le calcul d'une prévision conditionnelle, aucune information sur les variations inattendues des autres variables n'étant disponible au moment de la simulation. Le choc de politique monétaire sera alors déterminé de façon à ce que la variable de politique monétaire atteigne la valeur souhaitée par l'institut d'émission. A cet égard, il faut tenir compte du changement attendu de la variable de politique monétaire sur la base des valeurs passées de toutes les variables. Dans le cas des prévisions conditionnelles, il ne faut tenir compte que des effets des chocs de politique monétaire. La forme structurelle du modèle VAR tient explicitement compte de ces effets. Par conséquent, il convient de calculer directement les prévisions conditionnelles au moyen de la forme structurelle du modèle.

La forme structurelle du modèle VAR(1) s'exprime comme

$$B_0 y_t = B_1 y_{t-1} + u_t, \\ E(u_t u_t') = I$$

où $D = B_0^{-1} B_1$ und $\varepsilon_t = B_0^{-1} u_t$. Les innovations de la forme structurelle u_t , dont la variance est normalisée à un, peuvent dorénavant être interprétées comme des chocs structurels, puisqu'ils ne sont pas corrélés. Ainsi, les chocs structurels peuvent être attribués sans conteste – formellement, mais pas économiquement – à la fluctuation d'une variable. Les interdépendances entre les variables qui se manifestent durant une période sont capturées par la matrice B_0 .

Au moment $T+1$, le système peut être décrit comme suit:

$$y_{T+1} = D y_T + A_0 u_{T+1}, \\ A_0 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = B_0^{-1}.$$

On peut maintenant appliquer la méthode décrite ci-dessus à la fixation des chocs structurels: $u_{2,T+1}$ est déterminé de sorte que $y_{2,T+1}$ ait la valeur souhaitée sachant que $u_{1,T+1}$ est égal à zéro. Ce processus est légitime du fait que les chocs structurels ne sont pas corrélés. Ainsi, $u_{2,T+1}$ peut être considéré comme un choc de politique monétaire destiné à amener $y_{2,T+1}$ à la valeur souhaitée $y_{2,T+1}^*$. Ses effets

sur l'autre variable doivent également être examinés. Le choc de politique monétaire $u_{2,T+1}$ nécessaire correspond à

$$u_{2,T+1} = \frac{1}{a_{22}} (y_{2,T+1}^* - E_T y_{2,T+1}).$$

La prévision conditionnelle de $y_{1,T+1}$ se formule alors ainsi:

$$\hat{y}_{1,T+1} = d_{11} y_{1,T} + d_{12} y_{2,T} + a_{12} u_{2,T+1}.$$

La détermination de la forme structurelle du modèle VAR, c'est à dire la matrice B_0 et, ce faisant, la matrice A_0 constitue un pas décisif dans l'établissement de la prévision conditionnelle. Ce pas consiste à déterminer les chocs structurels à partir des chocs de la forme réduite et sera décrit à la section suivante.

Soulignons tout d'abord que la méthode décrite ci-dessus peut être appliquée récursivement si l'on s'intéresse à une prévision conditionnelle sur une longue période. Dans ce cas, la trajectoire dans le temps de la variable de politique monétaire est prédéterminée pour le nombre souhaité de périodes et on cherchera la prévision conditionnelle de l'autre variable. A cette fin, on recourra à la prévision venant d'être calculée pour y_{T+1} . A partir de

$$y_{T+2} = D \hat{y}_{T+1} + A_0 u_{T+2},$$

on calculera alors le choc de politique monétaire $u_{2,T+2}$ au moyen de la valeur prédéterminée $y_{2,T+2}^*$ et de la prévision conditionnelle de y_{T+1} . On posera $u_{1,T+2}$ égal à zéro pour les mêmes raisons que précédemment. Le calcul permettra ainsi d'établir la prévision conditionnelle de $y_{1,T+2}$. Cette procédure peut être répétée jusqu'à la date ultime de la période prévisionnelle.

3.2 L'identification de chocs structurels dans des modèles VAR

Dans la section précédente, nous avons souligné l'importance de la forme structurelle d'un modèle VAR pour le calcul de la prévision conditionnelle. Il convient maintenant d'identifier la forme structurelle d'un modèle VAR à partir de la forme réduite.

Nous avons déjà indiqué que seules les innovations de modèles VAR structurels pouvaient être considérées comme des chocs structurels. Les innovations de la forme réduite sont généralement corrélées, nous empêchant ainsi de les interpréter de façon structurelle. Seuls des chocs non corrélés constituent des sources indépendantes de changement, alors que des chocs corrélés comprennent toujours l'influence de diverses sources. Dans l'identification d'un modèle VAR, structurel le problème essentiel consiste à déterminer la matrice B_0 permettant de calculer la forme structurelle (non observable) à partir de la forme réduite (observable). La matrice B_0 doit remplir la condition $B_0\Omega B_0' = I$ pour assurer que les innovations u_t de la forme structurelle soient non corrélées et que la variance soit unitaire. Comme la matrice B_0 comporte n^2 éléments, mais que la condition ci-dessus ne fournit que $n(n+1)/2$ restrictions en raison de la symétrie de la matrice de variance-covariance Ω , il faut encore au moins $n(n-1)/2$ restrictions pour déterminer B_0 . Pour un modèle VAR à quatre variables par exemple, six restrictions sont nécessaires. Il serait judicieux de justifier ces restrictions supplémentaires, nommées d'habitude restrictions d'identification, en se basant sur la théorie économique. On s'assurerait ainsi que les innovations de la forme structurelle ont effectivement une interprétation économique structurelle.

La forme traditionnelle de restrictions d'identification, que nous avons déjà mentionnée à la section 1.2, se fonde sur le fait que la matrice B_0 ou A_0 est triangulaire (modèle VAR récursif). Il en résulte que les chocs structurels seront identifiés au moyen de leur effet sur les variables dans le temps. Seul le premier choc de u_t touche immédiatement la première variable de y_t les autres chocs n'agissant qu'avec retard. La deuxième variable ne sera touchée que par les deux premiers chocs, etc. On peut décider, par exemple, que la croissance de la production et l'inflation ne réagiront qu'avec un certain retard à des chocs de politique monétaire. Cette limitation temporelle de l'effet des chocs ne peut être judicieuse sur le

plan économique que si la fréquence d'observations des variables est relativement élevée. En outre, il n'est pas toujours possible de trouver des restrictions judicieuses sur le plan économique qui impliquent une structure triangulaire de B_0 .

La critique formulée par Cooley et LeRoy (1985) à l'égard de la forme traditionnelle des restrictions d'identification dans les modèles VAR a abouti à toute une série d'autres méthodes dépassant la triangularité de B_0 . D'une part, elles restreignent l'effet à court terme des divers chocs et, d'autre part, leur effet à long terme (ou la combinaison des deux).¹⁷ Dans ces conditions, on peut partir, par exemple, de l'idée que le choc de politique monétaire et celui au taux d'intérêt nominal influent simultanément sur le taux d'intérêt et la masse monétaire et que ces deux chocs n'exercent pas d'effet immédiat sur l'inflation et le PIB. Dans un modèle VAR à quatre variables (inflation, croissances de la production et de la masse monétaire, taux d'intérêt), il en résulte la structure non triangulaire suivante:

$$A_0 = \begin{bmatrix} x & x & 0 & 0 \\ x & x & 0 & 0 \\ x & x & x & x \\ x & x & x & x \end{bmatrix}$$

Le modèle n'est pas encore identifié par ces quatre restrictions à court terme. Nous obtiendrons les restrictions encore nécessaires par des restrictions supplémentaires à long terme. Une restriction envisageable serait de postuler, par exemple, que les chocs de politique monétaire n'exerceront aucun effet à long terme sur le PIB en raison de la neutralité de la monnaie. Notons qu'en ce qui concerne les restrictions d'identification à long terme des modèles VAR structurels, il faut tenir compte du fait que celles-ci ne s'appliquent qu'à des variables intégrées d'ordre un (ou davantage). Par conséquent, certains chocs exercent des effets permanents sur ces variables. Si tel est le cas, on pourra déterminer à partir de la forme structurelle d'un modèle VAR(1) en première différence

$$B_0\Delta y_t = B_1\Delta y_{t-1} + u_t,$$

la représentation «moyenne mobile» du modèle

$$\Delta y_t = A_0 u_t + A_1 u_{t-1} + A_2 u_{t-2} + \dots$$

Cette formulation, dénommée également fonction de réponse aux impulsions, représente l'effet dynamique que les chocs structurels contemporains et passés exercent sur la première différence des variables. Ces effets se manifestent dans les matrices

17 Bernanke (1986), Blanchard et Watson (1986) ainsi que Sims (1986) ont introduit des restrictions d'identification à court terme non fondées sur la triangularité de B_0 . Shapiro et Watson (1988), Blanchard et Quah (1989) ainsi que King et consorts (1992) sont des exemples précoces

d'identification de chocs au moyen de restrictions à long terme. Galí (1992), le premier, a combiné des restrictions à court et à long terme.

A_0, A_1, A_2, \dots . La fonction de réponse aux impulsions pour le niveau des variables se formule alors ainsi:

$$y_t = y_0 + A_0 u_t + (A_0 + A_1) u_{t-1} + (A_0 + A_1 + A_2) u_{t-2} + \dots + (A_0 + A_1 + \dots + A_{t-1}) u_1.$$

Par conséquent, il est évident que la matrice $A(1) = A_0 + A_1 + \dots + A_{t-1}$ représente les effets à long terme exercés par les chocs structurels sur le niveau des variables. On obtient les restrictions d'identification en posant certains éléments de la matrice $A(1)$ comme égaux à zéro. Une relation directe, quoique relativement compliquée, existe entre $A(1)$ et, respectivement, la matrice B_0 et A_0 . Par conséquent les hypothèses que la théorie économique nous offre concernant la forme de la matrice $A(1)$ sont transposées à la matrice B_0 , ce qui permet d'identifier la forme structurelle.

Relevons également que pour identifier un modèle VAR structurel il peut s'avérer adéquat du point de vue de la théorie économique d'appliquer plus de restrictions que nécessaires. On obtient alors un VAR suridentifié. La validité des restrictions supplémentaires peut être testée statistiquement.

En identifiant un modèle VAR, on cherche à obtenir des chocs structurels auxquels on peut attacher une interprétation économique. A des fins de contrôle, il est donc utile d'examiner les effets des chocs identifiés pour déceler d'éventuelles contradictions entre les effets des chocs structurels estimés empiriquement et les effets attendus par la théorie économique; et le cas échéant modifier le modèle. L'instrument servant à évaluer des modèles structurels VAR est la fonction de réponse aux impulsions. Celle-ci reflète la réaction dans le temps des variables aux chocs identifiés. D'une part l'examen de la fonction de réponse aux impulsions fournit des informations sur les conséquences dans le temps des chocs; ainsi, elles renseignent notamment sur l'importance des délais de réaction à la politique monétaire. Pour que les chocs identifiés puissent servir à la prévision conditionnelle leurs effets doivent, d'autre part, corroborer les hypothèses de la théorie économique en la matière. Il serait par exemple peu crédible d'un point de vue économique que les chocs utilisés dans la simulation de la politique monétaire soient basés sur l'idée qu'une politique monétaire axée sur l'expansion réduise l'inflation à long terme.

La désignation SVAR (VAR structurel) donne peut-être l'impression que cette catégorie de modèle est en étroite relation avec les modèles structurels macroéconométriques traditionnels. Les explications susmentionnées démontrent cependant que tel n'est le cas que de manière très limitée. Dans des modèles VAR structurels, seules sont fixées de rares restrictions à court ou long terme qui sont fondées sur la théorie économique. Par ailleurs, la relation entre les variables reste indéterminée et est définie implicitement sur la base de données empiriques, ce qui est un avantage indéniable. Dans des modèles structurels traditionnels, l'effet de la politique monétaire sur l'économie (mécanisme de transmission) est formulé, au contraire, par une description détaillée du processus de transmission des variations de taux d'intérêt et de change à de nombreuses composantes de la demande et aux déflateurs de celles-ci. En raison des nombreuses théories économiques, parfois contradictoires, quant au mécanisme de transmission, il s'agit là, certainement, d'un inconvénient.

3.3 Le modèle SVAR utilisé par la BNS

Dans cette section, nous présenterons le modèle SVAR utilisé par la BNS dans ses analyses de politique monétaire. Le modèle comprend quatre variables, à savoir le taux d'inflation calculé selon l'IPC, le taux de croissance du PIB réel, le taux de croissance de la masse monétaire M_1 et la première différence du Libor à trois mois. Le vecteur des variables incluses dans le modèle peut donc être défini comme

$$y'_t = (\Delta \log p_t, \Delta \log y_t, \Delta \log m_t, \Delta r_t),$$

les différences des logarithmes correspondant aux taux de croissance des variables.

Comment justifier le choix des variables? Il convient visiblement de tenir compte du taux d'inflation et du taux d'intérêt, puisqu'ils jouent un rôle l'un en tant qu'objectif final et l'autre en tant qu'objectif opérationnel dans la stratégie de politique monétaire de la BNS. Le taux de croissance du PIB est également indispensable, puisque cette donnée doit aussi être prise en compte dans la politique monétaire. Finalement, la masse monétaire M_1 a été incluse, car les opérations monétaires se manifestent très vite dans les variations des taux d'intérêt et des masses monétaires. En d'autres termes, la masse monétaire fournit des informations importantes pour l'identification d'un choc de politique monétaire. Toutes les variables sont introduites en première différence. Ceci s'explique, d'une part, par les résultats des tests de racine unitaire mentionnés et, d'autre part, par le fait que l'on ne puisse pas, rejeter l'hypothèse de non-cointégration des variables considérées. Ceci est en contradiction avec l'idée d'une fonction de demande de monnaie stable à long terme, c'est-à-dire d'une relation stable entre la masse monétaire, le niveau des prix, le revenu réel et le niveau des taux d'intérêt. Pour la Suisse comme pour de nombreux autres pays, une telle relation d'équilibre à long terme n'existe que dans le cas d'agrégats définis largement tels que M_3 , et non dans le cas de M_1 . Une fonction instable de la demande de M_1 signifie qu'il existe des chocs à la demande de monnaie aux effets permanents sur la détention réelle de monnaie. Cet état de fait permet une identification plus aisée des chocs de politique monétaire.

En examinant la liste des variables, beaucoup de lecteurs déploreront certainement l'absence du cours de change. Dans l'analyse de la politique monétaire d'une petite économie ouverte, ce facteur a une grande importance. Il n'entre pas ici en ligne de

compte pour deux raisons. Premièrement, le taux de change constitue une série chronologique très volatile, subissant des ruptures structurelles (introduction puis abrogation de l'objectif de change à la fin des années septante, introduction de l'euro à la fin des années nonante). Ainsi, il est difficile à représenter de manière adéquate dans un modèle VAR linéaire. Deuxièmement, les fonctions de réponse aux impulsions reflètent implicitement l'effet exercé sur l'inflation et la croissance réelle par des variables non prises directement en considération. Un avantage de la méthode SVAR est précisément le fait que les relations entre variables ne doivent pas être spécifiées de manière complète, mais qu'elles résultent de l'interaction dynamique de celles-ci. Pour étayer empiriquement cette opinion, on constatera, d'une part, que l'effet exercé par un choc de politique monétaire sur le niveau des prix et du PIB dans le modèle SVAR avec le cours de change ne se distingue pas, pour l'essentiel, de celui d'un modèle SVAR sans le taux de change. D'autre part, avec un modèle à cinq variables les réactions dynamiques de ces variables sont estimées avec plus d'imprécision.

Il importe de faire quelques remarques sur le choix de l'agrégat de la masse monétaire. A première vue, les avoirs en compte de virements, soit les réserves des banques à la BNS, constitue l'indicateur monétaire le plus approprié de la politique monétaire. Toutefois, cette variable est également très volatile et subit des ruptures structurelles. L'introduction du SIC et de nouvelles prescriptions en matière de liquidités au début de 1988 ont entraîné de profonds changements de la demande d'avoirs en comptes de virements. De plus, le fait que l'objectif opérationnel de politique monétaire ait passé, à la fin des années nonante, des avoirs en compte de virements au taux du Libor à trois mois a changé les propriétés de la série chronologique des avoirs en comptes de virements. Une autre rupture structurelle a résulté de l'introduction des pensions de titres intrajournalières, qui a encore réduit le besoin des banques en avoirs en comptes de virements. Les caractéristiques de ces avoirs en tant que série chronologique ont donc fortement changé au cours du temps, de sorte que cette série ne peut pas être utilisée dans le modèle. Nous recourons donc à la masse monétaire M_1 , qui constitue une bonne approximation du concept de caisse de transactions des modèles théoriques.

Nous devons maintenant nous occuper des hypothèses économiques servant à identifier le modèle, en particulier du choc de politique monétaire

indispensable au calcul des prévisions conditionnelles. Il faut ici indiquer que nous pouvons potentiellement identifier quatre chocs à partir de nos quatre variables: un choc de l'offre ou de productivité, un choc des prix, un choc de politique monétaire et, finalement, un choc du taux d'intérêt nominal. Pour identifier le choc monétaire, des restrictions à court et à long terme ont été utilisées:

– Durant le trimestre en cours, les chocs monétaires (politique monétaire et taux d'intérêt nominal) ne se répercutent pas sur l'inflation et la croissance réelle. Il en résulte formellement que la matrice A_0 et, partant, son inverse B_0 sont des matrices triangulaires par bloc (les quatre éléments en haut à droite sont nuls).

– A long terme, seuls les chocs de l'offre influenceront sur le PIB réel, alors que des chocs du taux d'intérêt nominal exerceront, en plus, des effets à long terme sur le taux d'intérêt. La masse monétaire M_1 et le niveau des prix seront influencés à long terme par les quatre chocs. Il en découle que cinq éléments de la matrice $A(l)$ représentant la somme des coefficients des fonctions de réponse aux impulsions sont égaux à zéro.

Ces hypothèses se justifient, d'une part, par l'existence de rigidités et frictions à court terme qui retardent l'effet de la politique monétaire sur la production et les prix. D'autre part, elles reflètent la neutralité à long terme de la monnaie, qui implique que des perturbations dans l'offre de monnaie se reflètent seulement à long terme sur le niveau des prix et la masse monétaire nominale.

Nous avons noté à la section 3.2 que six restrictions sont nécessaires au minimum pour identifier les chocs structurels d'un modèle VAR à quatre variables. Comme neuf restrictions ont été imposées, le modèle est donc suridentifié. Des tests statistiques quant à la validité des restrictions d'identification supplémentaires montrent que celles-ci ne sont pas en contradiction avec les données. Ces tests montrent même qu'une structure triangulaire des matrices B_0 (et A_0) est compatible avec les données.

L'estimation de ce modèle par des données trimestrielles allant du deuxième trimestre de 1974 à la même période de 2002 aboutit aux fonctions de réponse aux impulsions figurant dans le graphique 3.1. Elles montrent l'effet d'un choc de politique monétaire. Il est à noter que l'on représente les effets cumulés, à savoir les effets sur le niveau des variables, et non les premières différences. Les intervalles de confiance sont calculés à l'aide de la méthode «bootstrap». Un choc de politique monétaire de la taille d'un écart-type va immédiatement accroître la masse monétaire M_1 de

1,5% et réduire le taux du Libor à trois mois de presque 0,5 point de pourcentage (voir diagrammes en haut et en bas à droite). Au cours des quatre trimestres suivants, cet effet dit de liquidité se réduit à zéro en raison de l'augmentation des anticipations inflationnistes. Le taux d'intérêt monte alors au-dessus de son niveau initial (au maximum près de 0,1 point de pourcentage), pour converger à son niveau de départ près de quatre ans plus tard. Après une réaction initialement négative, le PIB réel réagit positivement avec un retard. Une augmentation cumulative du taux de croissance de 0,1 point de pourcentage est enregistrée après huit trimestres (voir diagramme en bas à gauche). Dès la troisième année, l'effet se réduit pratiquement à sa valeur de long terme de zéro. Le niveau des prix à la consommation ne réagit qu'avec un fort décalage à une politique monétaire expansive: plus de trois ans s'écoulent avant que l'effet de la politique monétaire expansive se soit répercuté intégralement sur le niveau des prix; la hausse du niveau des prix s'inscrit à quelque 0,4 pour cent. Après trois à quatre trimestres, la hausse des prix à la consommation ralentit; il faut l'attribuer à la baisse des loyers consécutive à celle des taux d'intérêt (voir diagramme en haut à gauche).

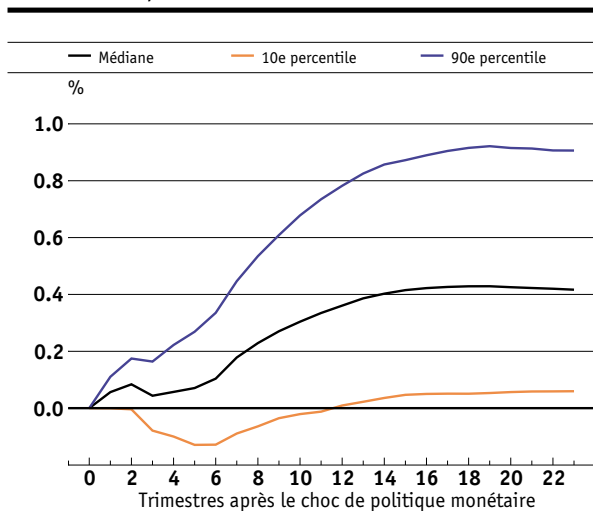
Les effets des chocs de politique monétaire paraissent plausibles. Ils prouvent que la banque centrale peut seulement influencer à court terme sur la masse monétaire et le niveau des taux d'intérêt. A moyen terme, c'est-à-dire de un à deux ans, la politique monétaire influe sur l'économie réelle; à long terme, c'est-à-dire après trois ans et davantage, seul subsiste l'effet sur le niveau des prix. L'effet négatif qu'une politique monétaire expansive exerce aux deuxième et troisième trimestres sur le PIB étonne au premier abord. Outre par l'imprécision de l'estimation (voir l'intervalle de confiance), il peut s'expliquer, par l'effet de la courbe j : la dépréciation du franc résultant de la politique monétaire expansive réduit à court terme la contribution au PIB réel des échanges réels avec l'étranger; les exportations et importations réelles ne réagissent en effet que faiblement à cette dépréciation.

3.4 Prévisions d'inflation conditionnelles

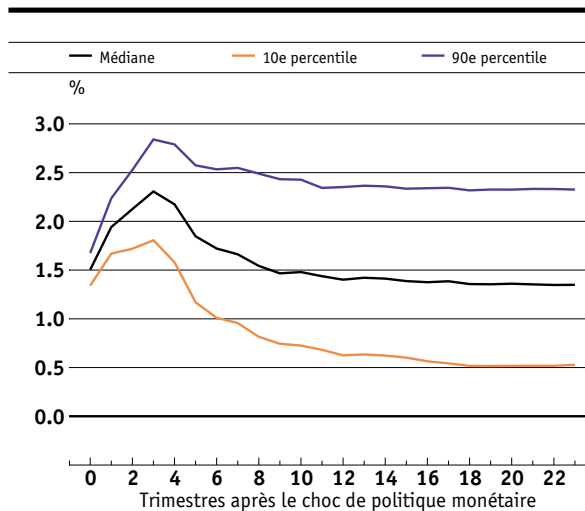
Dans cette section, nous décrivons l'emploi du modèle SVAR pour évaluer la politique monétaire. A cet égard, il faut tout d'abord tenir compte de la critique de Lucas aux évaluations économétriques d'une politique économique.¹⁸ Dans le cas d'un changement

18 Pour une plus ample discussion du bien-fondé de la critique de Lucas pour ce type d'évaluation à l'aide de modèles VAR, le lecteur se référera à Leeper et Zha (1999).

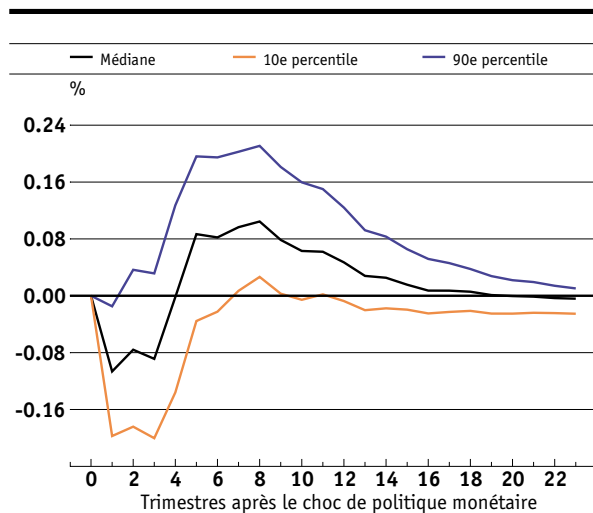
Réaction des prix à la consommation



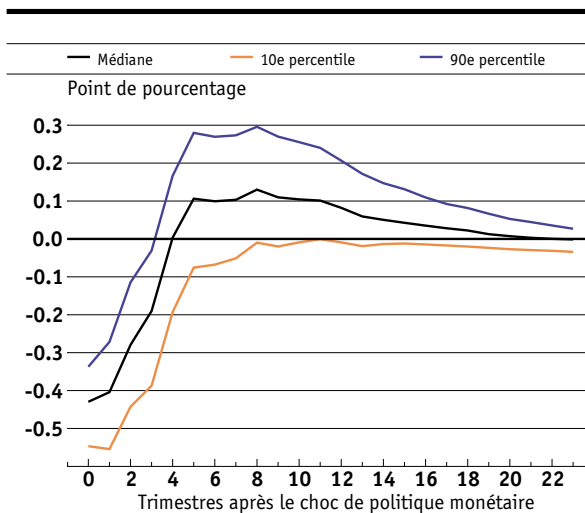
Réaction de la masse monétaire M1



Réaction du PIB



Réaction du Libor à trois mois



systématique de la politique se répercutant sur la formation des anticipations des agents économiques privés, Lucas affirme que les modèles économétriques, dont les paramètres ont été estimés dans un autre régime, perdent leur validité. Il nous faut donc nous demander si un changement systématique de la politique monétaire suisse a eu lieu depuis 1974. A notre avis, il faut répondre par la négative. Depuis le passage à des changes flexibles, la politique de la BNS a été axée sur un objectif final de stabilité des prix qui se caractérise par un taux d'inflation bas. Bien que les procédures opérationnelles et les objectifs intermédiaires visés aient changé et influé sur les caractéristiques de séries temporelles comme les réserves bancaires, les taux d'intérêt à court terme (taux de l'argent au jour le jour, taux des pensions de titres), on peut estimer que ces modifications n'ont toutefois pas exercé d'effets sur l'inflation, les croissances de M_1 et du PIB, et le taux du Libor à trois mois.

a) Prévisions conditionnelles avec taux d'intérêt constant

Après ces remarques préalables, nous examinerons tout d'abord les conséquences d'une fixation du taux d'intérêt à un niveau r^* durant les trois années suivantes. Lors d'une première étape, la prévision non conditionnelle pour la période $T+1$ est calculée au moyen du modèle SVAR. En règle générale, la valeur du taux d'intérêt obtenue différera de la valeur-objectif r^* . Le choc de politique monétaire de la période $T+1$ est ainsi déterminé de façon à ce que le taux d'intérêt atteigne le niveau souhaité. A l'aide du coefficient a_{43} de la fonction de réponse aux impulsions indiquant la réaction contemporaine du taux d'intérêt (quatrième variable du modèle) à un choc de politique monétaire (troisième choc du modèle), le choc monétaire requis s'écrit:

$$u_{3, T+1} = \frac{1}{a_{43}} (r^* - E_T r_{T+1}).$$

La prévision conditionnelle de $T+1$ se formule alors comme suit:

$$\hat{y}_{T+1} = E_T y_{T+1} + a_{33} u_{3, T+1},$$

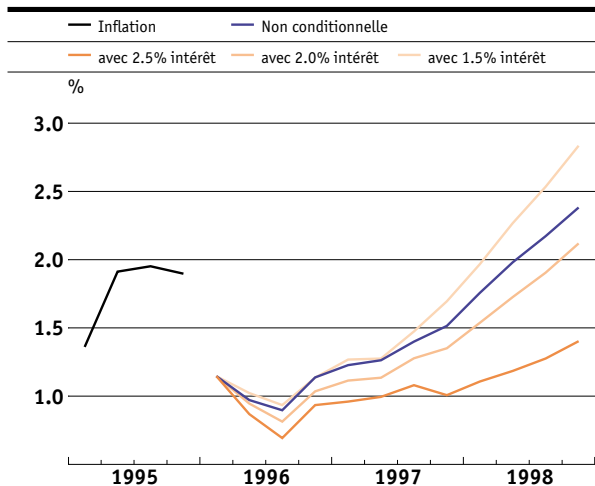
où le vecteur a_3 correspond à la troisième colonne de la matrice A_0 . Lors de la prochaine étape, nous partirons de la prévision conditionnelle de $T+1$ pour calculer une prévision non conditionnelle pour $T+2$ grâce au modèle SVAR. Nous adapterons ensuite cette prévision au taux d'intérêt prédéterminé, comme pour la période $T+1$. Cette procédure sera répétée jusqu'à ce que l'on atteigne la date ultime de la période prévisionnelle (c'est-à-dire douze trimestres).

Pour l'illustrer, nous considérerons les effets attendus sur l'économie de divers taux d'intérêt constants prédits par le modèle au quatrième trimestre de 1995 pour la période allant du premier trimestre 1996 au quatrième de 1998. Nous calculerons la prévision conditionnelle à la fin de l'année 1995, sous l'hypothèse du maintien, par la BNS, du taux d'intérêt constant les trois années suivantes à 1,5, 2,0 et 2,5%. Le graphique 3.2 indique l'évolution du taux d'inflation, ainsi que des taux de croissance du PIB réel et de la masse monétaire M_1 résultant de ces divers taux d'intérêt à court terme. Exception faite du taux d'intérêt, toutes les variables sont présentées comme des taux de croissance par rapport au trimestre précédent. Outre les prévisions conditionnelles, une prévision non conditionnelle y figure également. Elle ne contient aucun choc de politique monétaire et le taux d'intérêt y fait l'objet d'une prévision.

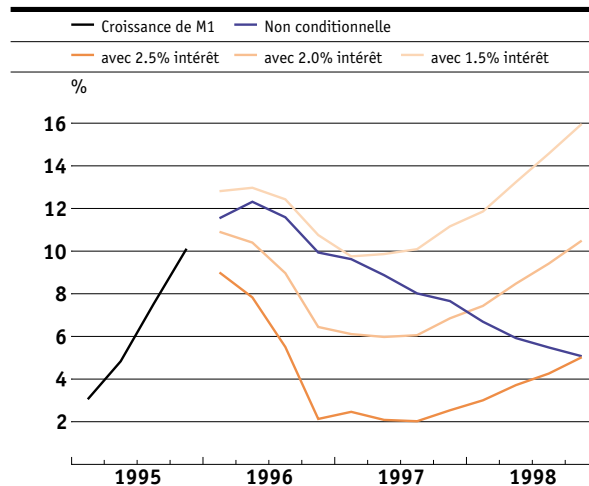
Lorsque le niveau du taux d'intérêt est fixé à 1,5% à partir du début de 1996, on peut voir sur le graphique que l'inflation prévue dépasse la barrière des 2% à partir de la seconde moitié de 1998. Si le niveau du taux d'intérêt est maintenu à 2%, ce dépassement ne se manifestera qu'à partir de la fin de 1998. Pour maintenir la prévision d'inflation nettement au-dessous de 2% durant l'horizon prévisionnel de trois ans, une politique monétaire plus restrictive s'impose, à savoir le relèvement et le maintien du taux d'intérêt à 2,5%.

La prévision conditionnelle du PIB réel est l'image inversée de celle de l'inflation: la prévision calculée en partant d'un taux d'intérêt de 1,5% dépasse de 0,2 point de pourcentage celle qui résulte d'un taux d'intérêt supposé de 2,0%; toutefois, cette différence découle d'une forte baisse du niveau du taux d'intérêt réel; en raison des effets différés sur la croissance réelle et l'inflation, les pleines conséquences de cette politique n'apparaissent qu'au delà de l'horizon prévisionnel considéré, c'est-à-dire dans les années 1999 et 2000. Les prévisions conditionnelles à taux d'intérêt constants aident à déterminer le degré d'activité nécessaire de la politique monétaire. Cependant, les résultats susmentionnés montrent aussi le problème soulevé par la fixation du taux d'intérêt nominal pendant trois ans consécutifs alors que l'inflation évolue. Il est évident que la BNS ne maintiendrait pas le taux d'intérêt au même niveau dans ces circonstances. Pour que le taux d'inflation reste à long terme entre 0% et 2%, soit la plage assimilée à la stabilité des prix, la politique monétaire devrait être durcie au cours des derniers trimestres de la période de prévision tant dans le scénario de 1,5% que dans celui des 2,0%.

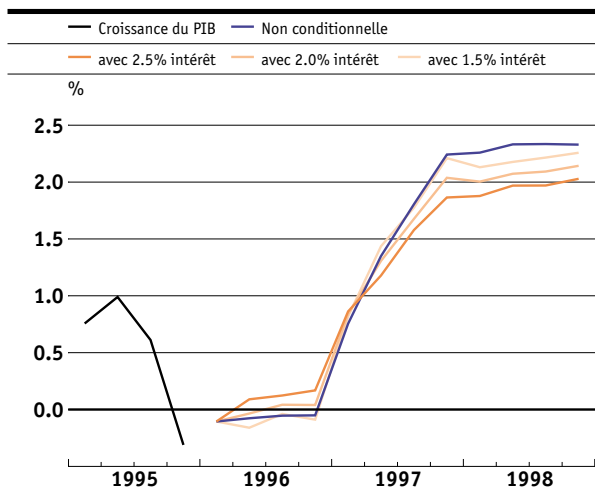
Prévisions d'inflation



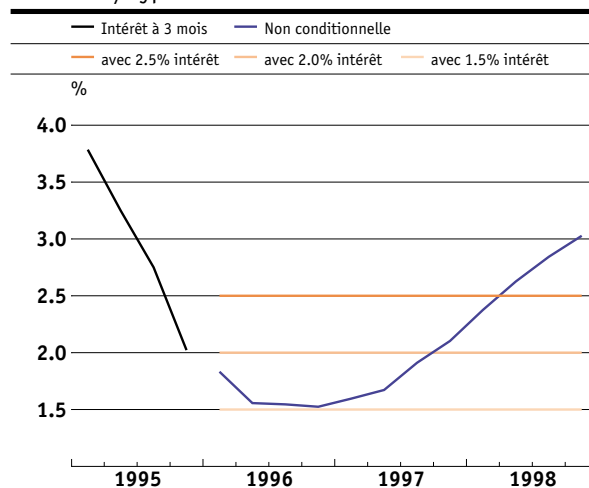
Prévisions de M1



Prévisions du PIB



Prévisions/hypothèses relatives au Libor à 3 mois



b) Prévisions conditionnelles à taux d'intérêt variable et objectif mobile d'inflation

Compte tenu d'un objectif mobile d'inflation le modèle SVAR peut également être utilisé pour en simuler les implications en termes de fixation du taux d'intérêt. Par exemple, à chaque période le taux d'intérêt peut être fixé de telle manière qu'un taux moyen d'inflation de 2% soit visé pour les trois prochaines années. Dans le cadre de cette simulation, il convient d'adapter la politique monétaire de façon à ce que le niveau prévu des prix augmente en moyenne de quelque 0,5% par trimestre, soit 2% par an au cours des trois prochaines années. Cette condition peut être incluse dans la prévision SVAR, comme lors de la procédure relative au taux d'intérêt prédéterminé.

A nouveau, le point de départ consiste en une prévision non conditionnelle du niveau des prix durant les douze prochains trimestres. Si cette prévision dépasse la valeur visée d'un taux d'inflation de 0,5% par trimestre ($\pi^*=0,5\%$), la politique monétaire doit être adaptée. Le montant du choc monétaire requis se calcule à l'aide des fonctions de réponse aux impulsions, représentées dans le graphique 3.1. Si nous désignons par $AA_{13}(12)$, la réponse du niveau des prix (première variable du modèle), après un délai de douze trimestres, à un choc de politique monétaire (troisième choc du modèle), le choc monétaire requis est exprimé:

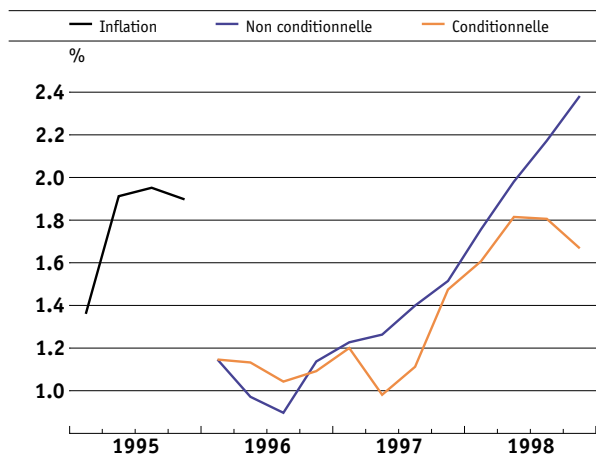
$$u_{3, T+1} = \frac{1}{AA_{13}(12)} ((\log p_{T+12} \pi^*) - E_T \log p_{T+12}).$$

Le choc de politique monétaire se calcule donc de façon que l'inflation trimestrielle moyenne visée π^* soit atteinte en $T+12$. Grâce au choc de politique monétaire ainsi calculé, on peut établir, comme dans le cas du taux d'intérêt constant, une prévision conditionnelle pour $T+1$. La procédure peut être répétée de manière analogue pour les périodes suivantes $T+1$, $T+2$, ..., $T+12$.

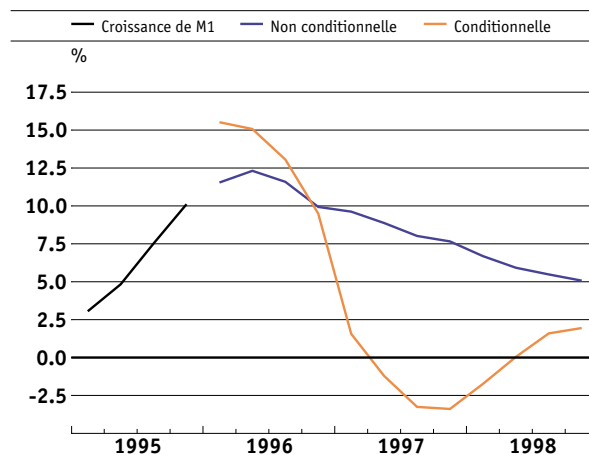
Pour illustrer la méthode, nous allons considérer les prévisions conditionnelles pour les années 1996 à 1998. Les résultats sont présentés dans le graphique 3.3 de la même manière que l'étaient ceux des prévisions conditionnelles dans le cas de taux d'intérêt constants. Pendant les premiers semestres de prévision, le taux d'intérêt est abaissé bien au-dessous de 2%. La raison en est que les prévisions non conditionnelles de l'inflation moyenne annuelle sont alors nettement inférieures à 2% pour les trois prochaines années. Les prévisions non conditionnelles montrent une hausse de l'inflation triennale à partir de 1997. De ce fait, la politique monétaire devient plus restrictive (taux d'intérêt relevé à 3%); la tendance inflationniste ascendante de la dernière année de la période de prévision est ainsi interrompue.

Dans la simulation décrite ci-dessus, on part de l'idée que la politique monétaire tient toujours compte de l'inflation moyenne attendue au cours des trois prochaines années au moment où le choc de politique monétaire est fixé. En règle générale, cela implique que la politique monétaire n'atteint pas l'inflation visée avec précision. Il est évident, par exemple, que le choc fixé au moment $T+3$ sur la base d'une prévision d'inflation moyenne des trimestres $T+3$ à $T+14$ exercera un effet sur l'inflation moyenne qui se réalisera de $T+1$ à $T+12$. Or il n'en est pas tenu compte au moment $T+1$ lorsque le choc est fixé sur la base de l'inflation moyenne attendue pour $T+1$ à $T+12$. Dans ce type de prévisions conditionnelles on néglige donc au moment où une décision est prise (par exemple en $T+1$), les décisions qui seront prises par la suite (par exemple en $T+3$). Cela explique pourquoi dans le graphique 3.3, le taux d'inflation reste au-dessous des 2% visés.

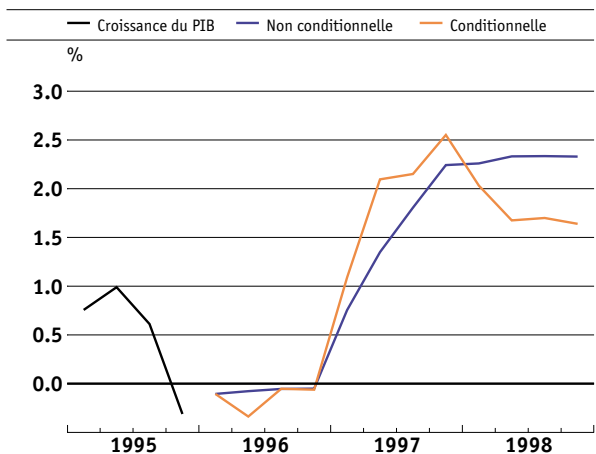
Prévisions d'inflation



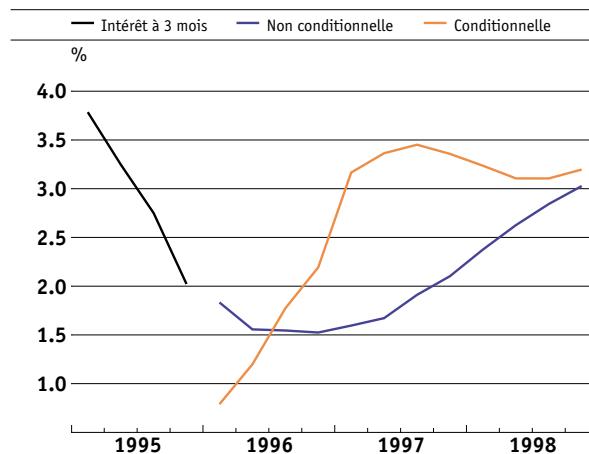
Prévisions de M1



Prévisions du PIB



Prévisions/hypothèses relatives au Libor à 3 mois



c) Prévisions conditionnelles à taux d'intérêt variable et objectif fixe d'inflation

Le modèle SVAR peut être également utilisé lorsque l'objectif d'inflation moyenne pour la période des trois ans à venir doit être exactement atteint. Ce cas n'a que valeur d'exemple, puisque la BNS ne vise pas une inflation moyenne déterminée, mais s'efforce de maintenir l'inflation au-dessous de 2%.

Dans ces circonstances, la procédure de simulation est légèrement différente. La politique monétaire a davantage de latitude, les chocs de politique monétaire pouvant être choisis sur tout l'horizon de prévision pour atteindre l'objectif visé. Comme point de départ, on utilise de nouveau la différence entre le niveau des prix attendus à la période $T+12$ et le niveau des prix visé avec un taux d'inflation moyen de π^* :

$$d = (\log p_{T+12} \pi^*) - E_T \log p_{T+12}.$$

Il convient maintenant de déterminer les chocs de politique monétaire de $T+1$ à $T+12$ de façon à ce que le niveau des prix attendu à la période $T+12$ corresponde au niveau des prix visé par la simulation. On y parviendra en minimisant le montant des chocs au carré. Mathématiquement, cela correspond au problème d'optimisation

$$\sum_{i=1}^{12} u_{3,T+i}^2 \rightarrow \min$$

comportant la contrainte suivante:

$$\sum_{i=1}^{12} AA_{13}(12-i) u_{3,T+i} = d.$$

La contrainte implique que la somme totale des chocs de politique monétaire pondérés par les coefficients appropriés de la fonction de réponse aux impulsions corresponde exactement à la différence entre le niveau des prix prévu non conditionnellement et celui visé. L'expression

$$u_{3,T+i} = \frac{AA_{13}(12-i)}{\sum_{j=0}^{11} [AA_{13}(j)]^2} d$$

représente la solution de cet exercice d'optimisation. Comme la réaction du niveau des prix au choc de politique monétaire a tendance à augmenter avec le temps (voir graphique 3.1), les chocs nécessaires de politique monétaire ont tendance à diminuer au cours de la période prévisionnelle.

Le graphique 3.4 illustre cette procédure durant la période allant du premier trimestre 1996 au quatrième de 1998. Il s'avère que la prévision non conditionnelle ne doit être modifiée que marginalement pour atteindre exactement l'objectif des 2% d'inflation triennale moyenne. L'inflation étant très faible durant les premiers trimestres de prévision, où elle

avoisine 1%, la politique monétaire est assouplie par rapport à la prévision non conditionnelle des huit premiers trimestres; il en résulte une hausse du taux d'inflation prévu à plus de 2,5%. Ce résultat est inévitable au vu de la position initiale du taux d'inflation nettement inférieur à 2%, qui rend nécessaire une hausse du taux d'inflation vers la fin de la période de prévision, ceci afin d'atteindre l'objectif fixé pour la moyenne triennale. Il est évident que si une telle croissance du taux d'inflation venait à se réaliser, il en résulterait progressivement une politique monétaire toujours plus restrictive.

d) Opérations courantes ou chocs de politique bénins

Comme mentionné, il faut veiller à ne pas simuler un changement systématique de politique monétaire lors de prévisions conditionnelles, sous peine de s'exposer à la critique de Lucas. En conséquence, les chocs de politique monétaire doivent être non systématiques et compatibles avec l'expérience historique. De tels chocs sont considérés comme *bénins* par Leeper et Zha (1999) («modest policy shocks») et, partant, admissibles, car *usuels* dans les opérations courantes de la banque centrale («business as usual policy»).

Cette exigence peut être vérifiée statistiquement. Premièrement, nous pouvons calculer la moyenne des chocs simulés pendant les 12 périodes, soit

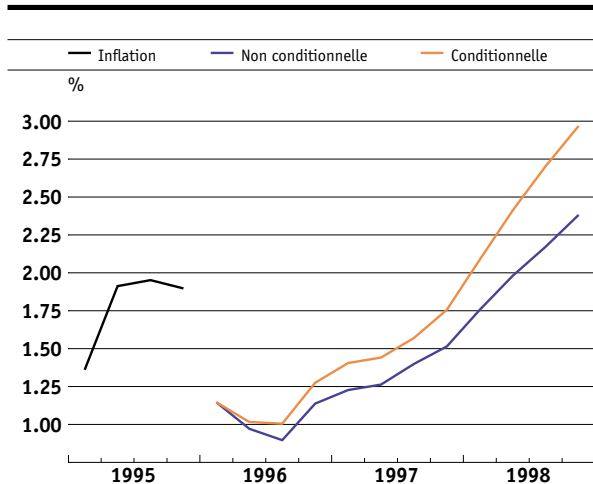
$$\eta(12, T) = \sum_{i=1}^{12} u_{3,T+i} / 12,$$

et examiner si cette statistique s'écarte fortement de zéro. Dans l'hypothèse de chocs distribués de manière identique et indépendante, cette statistique est distribuée de manière asymptotique selon une distribution normale avec une espérance mathématique de zéro et une variance de $1/T$. Ainsi, nous pouvons tester si les chocs de politique monétaire simulés subissent une distorsion positive ou négative systématique. Deuxièmement, nous pouvons examiner, en additionnant le carré des chocs simulés, si leur grandeur correspond à la moyenne historique. Dans l'hypothèse de chocs accidentels à variance de 1, la statistique suivante

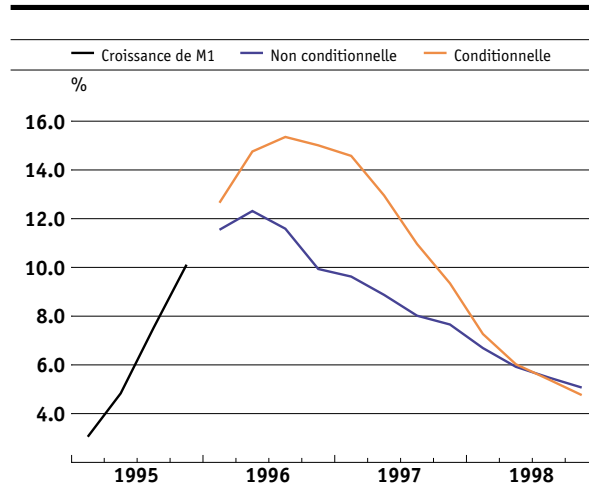
$$Q(12, T) = \sum_{i=1}^{12} u_{3,T+i}^2$$

est distribuée asymptotiquement selon la loi de χ^2 avec 12 degrés de liberté. Le rejet de l'hypothèse implique que la grandeur des chocs simulés est trop élevée. En l'espèce, ces statistiques montrent que les simulations effectuées à titre d'exemple sont usuelles et ne représentent pas une déviation systématique par rapport à la politique économique poursuivie dans le passé.

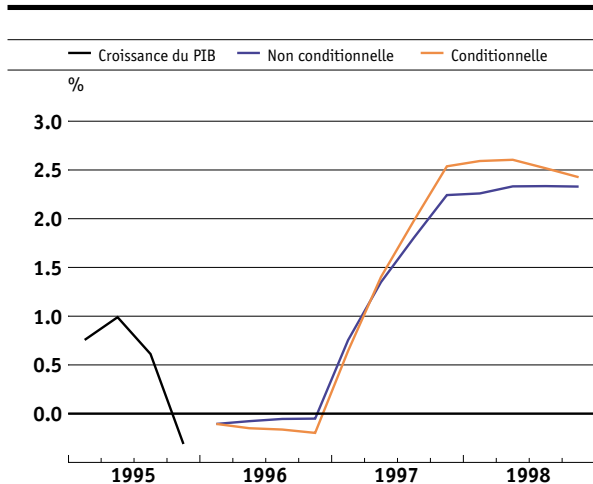
Prévision d'inflation



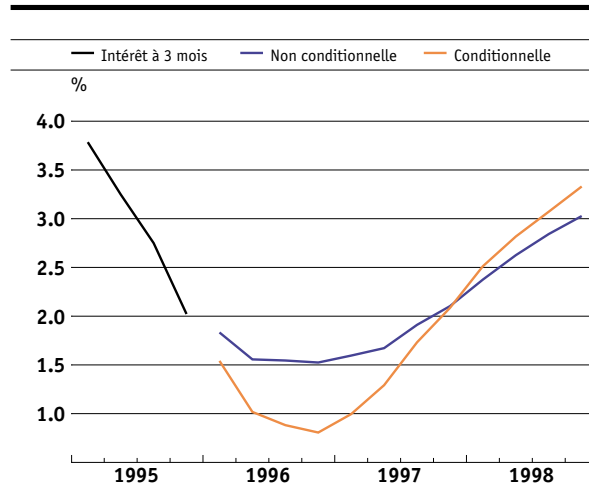
Prévisions de M1



Prévisions du PIB



Prévisions/hypothèses relatives au Libor à 3 mois



4 Conclusions

Dans ces conclusions, nous montrerons quelques limites des modèles VAR et esquisserons quelques pistes pour les surmonter. A cet égard, il faut distinguer entre les problèmes qui concernent la modélisation VAR en général et ceux qui se rapportent spécifiquement à l'identification de modèles VAR structurels. Généralement, deux propriétés spécifiques limitent les modèles VAR: ils sont linéaires et le nombre des variables qu'ils peuvent inclure est relativement limité. Ces caractéristiques influent naturellement sur les résultats atteints sur la base de ces modèles. Quant aux prévisions conditionnelles effectuées par des modèles VAR structurels, elles dépendent de manière cruciale des hypothèses faites pour identifier les modèles.

La formulation linéaire des modèles VAR ne permet pas de saisir des relations non linéaires éventuelles entre les variables, ainsi que des dépendances temporelles non linéaires des variables. Le second point, notamment, a fait l'objet d'une intense recherche économétrique. Il s'est avéré que de nombreuses variables économiques se caractérisent par une volatilité qui se modifie au fil du temps. De plus, cette volatilité a une structure temporelle particulière du fait qu'elle dépend de son propre passé. Les modèles GARCH, qui reflètent cette caractéristique des variables, ont connu un très grand succès, notamment dans la modélisation univariée du cours de change et de variables des marchés financiers. La combinaison de modèles VAR et de modèles GARCH pourrait ainsi permettre de saisir des dépendances non linéaires dans un contexte multivarié. Toutefois, l'estimation de ces modèles requiert toujours beaucoup d'observations; leur application se limite donc aux données financières. La fréquence relativement faible à laquelle les données macroéconomiques sont disponibles constitue de fait un obstacle général à l'application de modèles non linéaires.

Généralement, les modèles VAR ne comportent qu'un nombre de variables relativement restreint. Deux à six variables constituent la règle, les modèles à dix variables étant déjà beaucoup plus inhabituels. Cette limitation peut influencer sur la qualité des prévisions des modèles. Des changements importants de variables non-incluses dans le modèle conduisent, suivant les circonstances, à des erreurs importantes de prévision. Jusqu'à un certain degré, on peut éliminer cette source d'erreurs en accroissant le nombre de variables incluses dans le modèle. Cependant, il faut relever que le nombre des coefficients à estimer augmente au carré du nombre de variables et qu'il atteint déjà 150 avec un VAR à six variables et quatre retards. En considérant que seules 160 observations sont disponibles pour des données trimestrielles d'une durée de 40 ans, le problème qui en découle devient évident.

Une possibilité de le résoudre consiste à estimer des modèles VAR bayésiens. Pour appliquer cette méthode proposée par Litterman (1980), certaines suppositions sont faites quant aux relations entre les variables et à la dépendance des variables dans le temps. Chez Litterman, ces modèles sont restreints de manière à ce que chaque variable dépende plus fortement de son propre passé que de celui des autres variables. Il y est supposé que les observations perdent de l'importance avec le temps. De telles hypothèses permettent d'augmenter considérablement le nombre des coefficients qui peuvent être estimés et d'inclure un plus grand nombre de variables dans les modèles. De plus il s'est avéré que la qualité des prévisions non conditionnelles peut être améliorée par des modèles VAR bayésiens.

Outre ces réserves générales que l'on peut émettre envers les modèles VAR, il faut tenir compte, dans le cas des modèles SVAR, que les résultats dépendent des hypothèses faites pour identifier les chocs structurels. Si les hypothèses d'identification sont fausses, un choc structurel identifié comme choc de politique monétaire pourra inclure, de manière erronée, une composante systématique de la politique monétaire. Il est aussi possible que le choc monétaire identifié représente en fait une combinaison du véritable choc de politique monétaire et d'autres chocs. Bien que les résultats de tout modèle économétrique dépendent des hypothèses faites, il est utile de connaître l'importance de cette dépendance.

Pour identifier des chocs structurels, on recourt à des informations qui reflètent soit des données institutionnelles, soit des vues théoriques. Ainsi, des renseignements relatifs à l'instrument de politique monétaire et à ses effets à court terme, ou des informations quant à la neutralité d'un choc monétaire à long terme peuvent être utilisées pour l'identification des chocs. Une certaine incertitude quant à la validité de ces hypothèses d'identification étant inévitable, Faust (1998) et Uhlig (2001) ont proposé d'inclure cette incertitude dans l'identification des chocs. Leur idée consiste à d'abord identifier les chocs structurels à l'aide de différentes hypothèses, puis de confronter les résultats de ces différents chocs. Ainsi, on pourrait comparer diverses prévisions conditionnelles sur l'effet de chocs de politique monétaire qui reposent sur différentes hypothèses d'identification. Cette comparaison des prévisions fournit une mesure de l'incertitude due aux hypothèses d'identification.

Références bibliographiques

- Bernanke, B.S. 1986. Alternative Explanations of the Money-Income Correlation. *Carnegie Rochester Conference Series on Public Policy* 25: 49–100.
- Blanchard, O.J. et Quah, D. 1989. The Dynamic Effect of Aggregate Supply and Demand Disturbances. *American Economic Review* 79: 65–77.
- Blanchard, O.J. et Watson, M.W. 1986. Are Business Cycles All Alike? Dans *The American Business Cycle – Continuity and Change*, éd. R. Gordon. Chicago: University of Chicago Press.
- Canova, F. 1995. Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting. Dans *Handbook of Applied Econometrics, Macroeconometrics*, éd. M.H. Pesaran et M. Wickens. Cambridge MA: Blackwell.
- Cooley, T.F. et LeRoy, S.F. 1985. A Theoretical Macroeconomics: A Critique. *Journal of Monetary Economics* 16: 283–308.
- Dickey, D. et Fuller, W.A. 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* 49: 1057–72.
- Engle, R.E. et Granger C.W.J. 1987. Cointegration and Error-Correction: Representation, Estimation and Testing. *Econometrica* 55: 251–76.
- Faust, J. 1988. The Robustness of Identified VAR Conclusions About Money. *Carnegie-Rochester Conference Series on Public Policy* 49: 207–244.
- Friedman, M. 1968. The Role of Monetary Policy. *American Economic Review* 58: 1–17.
- Friedman, M. et Schwartz, A.J. 1963. *A Monetary History of the United States, 1867–1970*. Princeton: Princeton University Press.
- Gali, J. 1992. How Well Does the IS-LM Model Fit Postwar US Data? *Quarterly Journal of Economics* 107: 709–735.
- Johansen, S. 1988. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12: 231–254.
- Jordan, T.J. 1999. Inflationsprognosen mit VAR-Systemen. Miméo, Banque nationale suisse.
- Jordan, T.J. 1998. Monetary Policy Shocks in Switzerland: Is There a Liquidity Effect? Miméo, Banque nationale suisse.
- Jordan, T.J. et Peytrignet, M. 2001. La prévision d'inflation de la Banque nationale suisse. *Bulletin trimestriel de la Banque nationale suisse*, 19, 63–89.
- Jordan, T.J. et Savioz, M.R. 2001. Is there a Point in Combining Inflation Forecasts from VAR-Models? An Empirical Investigation for Switzerland. Miméo, Banque nationale suisse.
- King, R.G., Plosser, C.I., Stock, J.H. et Watson, M.W. 1991. Stochastic Trends and Economic Fluctuations. *American Economic Review* 81:819–840.
- Kugler, P. et Jordan, T.J. 2000. Vector Autoregressions and the Analysis of Monetary Policy Interventions: The Swiss Case. Miméo, Banque nationale suisse.
- Kirchgässner, G. et Savioz, M.R. 1997. Empirische Forschung in den Wirtschaftswissenschaften: Ein Überblick. *Homo oeconomicus* 16: 209–247.
- Koopmans, T.C. et Hood, W.C. 1953. The Estimation of Simultaneous Linear Economic Relationships. Dans *Studies in Econometric Method*, éd. W.C. Hood et T.C. Koopmans, 112–199. New York: John Wiley.
- Leeper, E.M. et Zha, T. 1999. Modest Policy Interventions. Federal Reserve Bank of Atlanta Working Paper 99–122.
- Litterman, R.B. 1980. A Bayesian Procedure for Forecasting with Vector Autoregression. Massachusetts Institute of Technology, Department of Economics Working Paper.
- Liu, T.C. 1960. Eteridentification, Structural Estimation, and Forecasting. *Econometrica* 28: 855–865.
- Lucas, R.E. 1976. Macroeconomic Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46.
- Phelps, E.S. 1967. Phillips Curves, Expectations of Inflation and Optimal Employment over Time. *Economica* 34: 254–281.
- Shapiro, M.D. et Watson, M.W. 1988. Sources of Business Cycle Fluctuations. *NBER Macroeconomics Annual* 3: 108–148.
- Sims, C.A. 1980. Macroeconomics and Reality. *Econometrica* 48: 1–48.
- Sims, C.A. 1986 Are Forecasting Models Usable for Policy Analysis. *Federal Reserve Bank of Minneapolis, Quarterly Review* 10: 2–16.
- Sims, C.A., Stock, J. et Watson, M. 1990. Inference in Linear Time Series Models with Some Unit Roots. *Econometrica* 58:113–44.
- Stalder, P. 2001. Un modèle macroéconométrique pour la Suisse, *Bulletin trimestriel de la Banque nationale suisse* 19, 63–89.
- Stock, J.H. et Watson, M.W. 2001. Vector Autoregressions. *Journal of Economic Perspectives* 1(4): 101–115.
- Uhlig, H. 2001. What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure. manuscrit, Humboldt Universität Berlin.
- Zha, T. 1999. Block Recursion and Structural Vector Autoregressions, *Journal of Econometrics* 90: 265–289.