

L'élaboration des données de comptabilité annuelle et l'analyse économique

Stéphane Gregoir (stephane.gregoir@edhec.edu)

Juin 2008

Objectifs

Illustrer quelques uns des problèmes que l'on peut rencontrer lorsque l'on veut utiliser des données de comptabilité nationale dans un exercice d'analyse macro-économique quantitative.

- les finalités usuelles des travaux macroéconométriques, les limites en termes d'interprétation qui sont présentes à l'origine dans ces travaux.
- illustration des interactions entre les objectifs des études macroéconomiques (interprétation et prévision) et les hypothèses explicites ou implicites retenues
 1. dans les conventions d'élaboration des données comptables
 2. dans les traitements statistiques appliqués aux données
- une illustration empirique (mesures de causalité sur données CVS et sur données brutes)

Conclusion

Conclusion contradictoire: la modélisation macroéconomique est un art qui doit être rigoureux...la comptabilité nationale est un corpus conventionnel et pérenne qui doit continuellement évoluer

Souhait: différents produits pour différents usagers: coût et multiplicité de besoins

Les comparaisons des amplitudes des phénomènes considérés dans les différents exercices macroéconométriques ne sont pas informatives en termes absolus, mais relativement à un jeu de règles, de convention et de traitement. Si ces derniers altèrent le contenu informationnel des données ou leurs relations, il y a fragilité de l'analyse et de l'interprétation.

Objectifs principaux d'une analyse quantitative macro-économique

- estimer les relations entre plusieurs grandeurs macroéconomiques sur la base d'une théorie et ainsi évaluer la pertinence de mécanismes économiques,
- prévoir l'évolution d'un ensemble de variables économiques,
- simuler les conséquences d'une modification de l'environnement des décisions des acteurs suite à une réforme de la politique fiscale ou monétaire ou une modification de la situation internationale

Propriétés mathématiques : usage et interprétation des modèles pour chaque objectif (notions d'exogénéité)

L'interprétabilité des résultats

Repose sur la satisfaction de ces hypothèses techniques, mais aussi sur le contenu informationnel attaché aux mesures utilisées et aux analogies formelles que le modélisateur est amené à faire pour produire un discours susceptible d'entraîner l'adhésion des auditeurs.

1. Modélisation structurelle
2. Modélisation statistique

Modélisation structurelle

Interprétation des équations dites de comportement des agents économiques : argument de comportement microéconomique

Ensemble d'information et rationalité : usage rationnel de l'information disponible

Limites du cadre

- interprétation en termes de bien-être (Kirman (1993), Jerison (1984)),
- interprétation des simulations de politique
 - Critique de Lucas (1976)
 - Paramètres de préférence vs paramètres de comportement
 - Formation des anticipations (importance des chronologies)

La stabilité dans le temps des mécanismes et des mesures utilisées dans la modélisation peuvent permettre de proposer un discours en permettant de comparer des situations dans le temps (prévisions, comparaison de phases de cycle)

La modélisation statistique sans a priori

Critique de Sims (1980): identification des formes structurelles

- Description statistique avec contraintes progressivement appliquées aux modèles
- Identification de chocs structurels par leurs conséquences sur les variables
- Définition du choc: usage de données infra-annuelles
- Qualité de prévision mais difficultés de discours explicatif.

Les difficultés induites par les traitements des données de comptabilité nationale

Enjeux: stabilité et interprétation des phénomènes mesurés

1. Les difficultés d'ordre comptable
2. Les difficultés liées au traitement statistique

Les difficultés d'ordre comptable

1. Moment d'enregistrement
2. Différents systèmes de prix
3. Traitement des investissements directs à l'étranger

Moment d'enregistrement

- Difficultés plus présentes dans les données infra-annuelles plus fréquemment utilisées
- Cohérence plus grande entre le cadre des comptes annuels et les données de base (exercice comptable...)
- Mesures trimestrielles alignées sur des concepts annuels
- Le moment d'enregistrement et la description économique que l'on veut produire : Arbitrage entre conventions et interprétabilité économique

Enjeu : décrire les ensembles d'information disponible des agents et modéliser leurs règles de décision conditionnelles à ces ensembles et leur chronologie (analogie de discours).

Moment d'enregistrement : SEC95

datation des opérations : "accrual basis" traduit par "droits et obligations"

mise en oeuvre : en pratique recouvre, suivant les situations, les notions (i) de fait générateur, (ii) de droits constatés ou encore (iii) de moment de paiement

Conventions

- cas des impôts avec un décalage d'un an: inadapté pour la description des comportements des ménages et pour l'immédiateté du discours
- cas des productions à processus de production "long" à issue aléatoire
- les versements irréguliers de revenus aléatoires (ou susceptible de modification annoncée à certaines dates)

Processus de production "long"

Exemple: cas de la production agricole infra-annuelle:

Approche préconisée par le SEC95 : valoriser la production en cours d'une période en fonction de la somme des coûts induits corrigée d'un taux de marge

Confort d'usage mais peut rendre inexploitable les données pour une modélisation dans le but de simuler les conséquences d'une modification d'un politique, détection des inflexions conjoncturelles

Problèmes:

1. le taux de marge: valeur, évolution cyclique et tendancielle
2. le volume physique de production est aléatoire (suivant le caractère exceptionnel ou non de cet aléa, le SEC95 propose des traitements différents)

Versements irréguliers de revenus

Exemple: les versements irréguliers de revenus aléatoire (ou susceptible de modification annoncée à certaines dates)

Le cas des paiements de primes de fin d'année fonction de l'activité de l'entreprise pendant l'année ou des décisions politiques relatives à une mesure exceptionnelle touchant les prestations ou les impôts .

Différentes stratégies : base caisse difficile à lire mais base d'une estimation d'une équation de comportement des ménages (avec information sur le calendrier des annonces)

Modification des données (imputation en cours d'année et révision) =
Hypothèse sur le comportement des agents face à l'incertain et leur comportement d'épargne.

Les difficultés liées au traitement statistique

- Données lacunaires
- Saisonnalité
- Jours ouvrés
- Points aberrants

Données lacunaires

Confrontation aux limites des données disponibles

- Saisonnalité artificielle : consommation des ménages en produits manufacturés (échantillonnage)
- Absence de saisonnalité artificielle : consommation d'électricité des ménages, consommation intermédiaire en énergie

Corrections des variations saisonnières

$$y_t = T_t + C_t + S_t + \varepsilon_t$$

Approche descriptive (non structurelle)

$$y_{CVS,t} = T_t + C_t + \varepsilon_t$$

correction par moyennes mobiles centrées et modélisation ARIMA ou extraction de signal et modélisation SARIMA

Filtres différents pour des séries différentes, instabilité des modèles utilisés

Corrections des variations saisonnières

- Politique de révisions: à quel niveau réaliser la désaisonnalisation et à quelle fréquence doit-on réestimer les modèles ? Pas de réponse simple, dépend des données: risque d'instabilité.
- Déformation des liens temporels qui existent entre plusieurs variables macroéconomiques.
méthodes univariées + emploi de filtres différents pour chaque série
= biais dans l'estimation des relations dynamiques

L'analyse conjoncturelle exploite ces liens pour construire des interprétations et faire des comparaisons dans le temps ou entre pays.

Diffusion de données brutes.

Points aberrants

- influence sur les traitements de CVS, usage d'information exogène
- analyse de la causalité (comouvements identifiés sur des événements extrêmes)
- Traitement en aveugle de manière séparée des points aberrants, résultat artificiel en termes d'ampleur et de chronologie.
L'estimation d'une équation reliant les variables traitées peut être biaisée.

Analyse de la non-causalité au sens de Granger (Prédictabilité)

Cadre de travail (Dufour et Renault (1998), Dufour, Pelletier et Renault (2006))

$$Y_t = \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \\ Y_{3,t} \end{pmatrix}$$

Idée : tester la significativité jointe des coefficients associés à une variable de $Y_{1,t}$ dans la régression de $Y_{2,t}$: mesure de la causalité (prédictabilité) de $Y_{1,t}$ vers $Y_{2,t}$ à $Y_{3,t}$ donné

Analyse de la non-causalité au sens de Granger (Prédictabilité)

$$Y_t = \mu + \sum_{j=1}^p \phi_j Y_{t-j} + \varepsilon_t$$

$$Y_{t+h} = \mu^{(h)} + \sum_{j=1}^p \phi_j^{(h)} Y_{t+1-j} + \sum_{i=0}^{h-1} \psi_i \varepsilon_{t+h-i}$$

avec

$$\phi_k^{(h+1)} = \phi_{k+h} + \sum_{l=1}^h \phi_{h-l+1} \phi_k^{(h)}$$

$$\mu^{(h)} = \sum_{l=0}^{h-1} \phi_1^{(l)} \mu$$

Analyse de la non-causalité au sens de Granger (Prédictabilité)

Deux difficultés :

1. estimation d'une forme VARMA: OLS plutôt que GLS et correction de la matrice de variance des estimateurs (pour assurer des matrices estimées symétriques définies positives: noyau de Bartlett, sélection du paramètre de troncature (Newey-West (1994)))
2. variables éventuellement non stationnaires : Régression avec augmentation du nombre de retard (Toda et Yamamoto (1995)) pour garder des statistiques de Wald asymptotiquement distribués comme des χ_2

Remarque : ceci implique des pertes d'efficacité

Extension au cas de variables non stationnaires saisonnières (Gregoir (2007))

Analyse de la non-causalité au sens de Granger (Prédictabilité)

Choix des variables :

- r taux d'intérêt directeur mensuel Federal Reserve
- $M2$ agrégat monétaire M2 mensuel
- IPI indice de production industriel (regroupement) mensuel

Disponibles désaisonnalisés et non désaisonnalisés (limite sur la construction de l'IPI des Etats-Unis, en partie usage de mesure de consommation d'électricité)

Objectif: tester le schéma de causalité sur les données CVS et sur les données brutes pour le système trivarié ($r, M2, IPI$)

Analyse de la non-causalité au sens de Granger (Prédictabilité)

Analyse des données brutes

| | 0 | $\frac{\pi}{6}$ | $\frac{\pi}{3}$ | $\frac{\pi}{2}$ | $\frac{2\pi}{3}$ | $\frac{5\pi}{6}$ | π |
|------|-------|-----------------|-----------------|-----------------|------------------|------------------|-------|
| ipi | 130,0 | 1,73 | 2,16 | 4,26 | 0,95 | 1,09 | 0,33 |
| à 5% | oui | | | oui | | | |
| M2 | 239,6 | 0,85 | 0,61 | 1,13 | 6,03 | 0,69 | 0,28 |
| à 5% | oui | | | | oui | | |
| r | 1,19 | 0,66 | 0,77 | 0,53 | 0,55 | 1,06 | 0,47 |
| à 5% | oui | | | | | | |

Modèles VAR ajustés

choix du nombre de retards (p) : Critère Hannan-Quinn

Données CVS: VAR(3)+ constante

Données brutes: VAR(14)+dummies saisonnières

Analyse de la non-causalité au sens de Granger (Prédictabilité)

'*': rejet de la non-causalité à 5%

| horizon | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---------------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| $r \rightarrow IPI$ cvs | | | | | | | | | | | * | * |
| $r \rightarrow IPI$ brute | * | * | * | * | * | * | * | * | * | * | * | * |
| horizon | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| $r \rightarrow IPI$ cvs | * | * | * | * | * | * | * | * | * | * | * | * |
| $r \rightarrow IPI$ brute | * | * | * | * | * | * | * | * | * | * | * | * |
| horizon | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $r \rightarrow M2$ cvs | * | * | * | * | * | * | * | * | * | * | * | * |
| $r \rightarrow M2$ brute | * | * | * | * | * | * | * | | | | | |
| horizon | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| $r \rightarrow M2$ cvs | | | | | | | | | | | | * |
| $r \rightarrow M2$ brute | | | * | * | * | * | * | * | * | * | * | * |

Analyse de la non-causalité au sens de Granger (Prédictabilité)

| horizon | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|--------------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| $M2 \rightarrow r$ cvs | | | | | | | * | * | * | * | * | * |
| $M2 \rightarrow r$ brute | | | | | * | * | * | * | * | * | * | * |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| $M2 \rightarrow r$ cvs | * | * | * | * | * | * | * | * | * | * | * | * |
| $M2 \rightarrow r$ brute | * | * | * | * | * | * | * | * | * | * | * | * |

| horizon | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----------------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| $M2 \rightarrow IPI$ cvs | * | * | * | * | * | * | * | * | * | * | * | * |
| $M2 \rightarrow IPI$ brute | * | * | * | * | * | * | * | * | * | * | * | * |
| | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| $M2 \rightarrow IPI$ cvs | * | * | * | * | * | * | * | * | * | * | * | * |
| $M2 \rightarrow IPI$ brute | * | * | * | * | * | * | * | * | * | * | * | * |

Que fait-on lorsque l'on réalise une étude quantitative sur la base de données de comptabilité nationale ?

Cette question couvre deux dimensions, d'une part technique et d'autre part relatif au domaine de l'interprétation et du discours.

Modèles structurels ou de modèles moins contraints a priori, travail sur des données brutes : modèle plus compliqué, moins parcimonieux, produit un discours peu immédiat.

Travail sur données CVS : relations déformées et risque d'estimation et d'analyse instables

Modélisation structurelle

Le contenu informationnel des données de comptabilité nationale n'est pas neutre en termes de comportements économiques.

Usage fragile pour évaluer les conséquences d'un changement de politique économique

Il serait nécessaire pour aller plus loin d'apprécier les conséquences empiriques de ces hypothèses de construction des données. Cela semble en pratique difficile. Recours croissant à des données individuelles microéconomiques.

L'analogie microéconomique pour interpréter les équations dites de comportement peu adaptée, les données ne semblent pas fournir l'information pour décrire l'équation de décision d'un agent représentatif.

Modélisation statistique

La spécification d'un modèle = choix des variables et description des liens entre les variables.

Identification partielle : la crédibilité du discours s'assoit sur la capacité à décrire avec pertinence dans le passé des situations qui ont été analysées par ailleurs.

Nécessité d'une grande stabilité des méthodes d'élaboration et des traitements des données.

Crédibilité des données conditionne leur usage et l'utilité des résultats obtenus dans des exercices de modélisation macroéconomique.

Crédibilité des données et missions de la comptabilité nationale

Trois objectifs contradictoires,

1. la pertinence des descriptions proposées, un monde en évolution
2. une politique de révisions, comparabilité des exercices successifs et accumulation des savoirs
3. une permanence des méthodes et des concepts, comparabilité temporelle.

MERCI