

L'HÉTÉROGÉNÉITÉ EN ÉCONOMÉTRIE

Avant-propos

Du 12 au 14 juin 1989, la conférence internationale consacrée à L'HÉTÉROGÉNÉITÉ EN ÉCONOMÉTRIE¹ a réuni à Paris dans les locaux de l'ENSAE une cinquantaine de chercheurs. Vingt-six communications venant des centres de recherches les plus réputés en économétrie² y ont été présentées. Ce numéro spécial apporte au travers d'une sélection de onze articles un bilan des travaux les plus récents dans ce domaine.

*
* *

La modélisation économétrique prend appui sur la modélisation statistique semi-paramétrique. Celle-ci repose entre autre sur l'hypothèse d'invariance des paramètres, valeurs caractéristiques des lois probabilistes qui régissent le phénomène observé. Dans le cas des données individuelles d'entreprises ou de ménages, cette hypothèse se traduit par l'homogénéité des comportements qui suppose que les paramètres essentiels, élasticités, propensions, etc. sont applicables à la totalité de la population. Pour les séries temporelles, elle renvoie à la stabilité des relations d'une période à une autre, à la persistance de la loi au cours du temps.

Si en statistique mathématique cette stabilité est la règle, en réalité les observations sont rarement homogènes. Et, le cadre dressé par cette première doit être aménagé en conséquence. C'est bien d'un aménagement qu'il s'agit car le modèle probabiliste est assez souple pour permettre une adaptation des modèles économétriques classiques à une hétérogénéité contrôlée.

1. Cette conférence a bénéficié des aides matérielles et financières de l'ADRES, de l'ENSAE et du CEPREMAP.

2. Le programme complet de cette conférence est donné en annexe.

Le modèle linéaire est l'illustration de base où l'on peut détailler assez simplement tous les stades d'une adaptation à la situation d'hétérogénéité. Dans cette dernière circonstance, les paramètres sont donc variables. Ceci peut-être obtenu de deux manières :

i) Les paramètres varient avec des variables exogènes; on recourt généralement à un modèle linéaire pour les paramètres, si bien que le modèle adapté reste linéaire en fonction des nouveaux paramètres. Cette manière de procéder est classique dans le cas de l'étude des ménages où une série de variables sociodémographiques est ajoutée aux variables de bases pour tenir compte des différences de comportements individuels.

ii) Les paramètres sont aléatoires; on suppose alors qu'ils sont tirés dans une loi de probabilité (généralement caractérisée par ses deux premiers moments) qui dépend de nouveaux paramètres stables. La combinaison de ces deux niveaux de modélisation pour la variable endogène d'une part, pour les paramètres associés d'autre part, produit un modèle linéaire hétéroscédastique.

La première manière bien classique ne relève pas de la recherche, la seconde est plus complexe voire très complexe lorsque le modèle est non linéaire. C'est l'objet d'étude de la plupart des articles de ce numéro.

Ainsi, L. BAUWENS et M. LUBRANO – *Bayesian diagnostics for heterogeneity* – se placent dans un cadre bayésien pour analyser ce modèle combiné. Ils adoptent deux façons de détecter l'hétéroscédasticité. La première repose sur les résidus bayésiens produits par le modèle homoscédastique; elle est à relier aux tests de spécification de l'économétrie classique qui mettent à l'épreuve le modèle de base en étudiant les particularités de certaines de ses statistiques. La seconde adopte une approche plus standard où l'on part du modèle hétéroscédastique (donc hétérogène) pour tester les restrictions qui mènent au modèle homoscédastique (donc homogène). Les outils classiques de la statistique bayésienne sont ici utilisables, L. Bauwens et M. Lubrano suggèrent le test du « ratio de Bayes ».

Lorsqu'un phénomène économique porte les traces de l'hétérogénéité des comportements individuels, sa contrepartie macroéconomique ne relève généralement pas d'un décalque du modèle microéconomique. N. FORTIN – *Fonction de production et biais d'agrégation* – propose une méthodologie pour déterminer les biais qui résultent de

l'agrégation d'un modèle de production individuel dont les paramètres sont aléatoires. La méthode étend les situations favorables à l'agrégation et présente de nombreux avantages plus pratiques; elle permet entre autre de traiter le cas de relations optimisées, et surtout d'établir une correspondance entre les paramètres macroéconomiques et leurs microfondements. Ainsi dans le cas de la fonction de production, N. Fortin analyse grâce à ses outils les causes de l'instabilité des paramètres de la relation agrégée.

Ces paramètres aléatoires résultent en fait de variables inobservables, facteurs de l'hétérogénéité des comportements. C'est avec cette terminologie que C. GOURIÉROUX et A. MONFORT – *Simulation based inference in models with heterogeneity* – proposent une approche très générale qui permet de traiter les modèles non-linéaires. Ils supposent que la distribution des facteurs d'hétérogénéité n'est connue qu'à un paramètre près et qu'elle est reliée à un vecteur aléatoire dont la loi est connue. Ils montrent sur de nombreux exemples classiques que leur modélisation n'est pas restrictive. Dans ce cas non-linéaire la vraisemblance du modèle conditionnelle aux variables exogènes, n'est bien souvent pas calculable explicitement; c'est là que repose toute la difficulté. Mais ici la distribution des facteurs d'hétérogénéité peut être approchée par simulation et la vraisemblance approximée par moyennage empirique. Ainsi grâce à la loi des grands nombres, lorsque le nombre de simulations augmente indéfiniment, on obtient la vraisemblance. C'est l'analyse précise des propriétés des estimateurs déduits de la vraisemblance simulée que propose C. Gouriéroux et A. Monfort : biais, normalité asymptotique, etc.

Les facteurs inobservables sont à la base des modèles de panel. Ils interviennent sous la forme de perturbations individuelles indépendantes du temps. Celles-ci produisent une corrélation temporelle qui interfère avec la partie autorégressive des modèles dynamiques. Aussi ces modèles ne relèvent-ils pas de procédures standards d'estimation. Les particularités de ce modèle dynamique ont été très tôt répertoriées, mais les bonnes solutions au problème d'estimation ont mis longtemps à véritablement émerger. Depuis une bonne dizaine d'années, il est apparu que le rôle des valeurs initiales de la variable endogène, dont le nombre croît avec la taille de l'échantillon, doit parfaitement être maîtrisé. Fort de cet avertissement, R. BLUNDELL et R. SMITH

– *Conditions initiales et estimation efficace dans les modèles dynamiques sur données de panel* – apportent des solutions apparemment définitives et qui sont simples à mettre en application.

Pour ce modèle dynamique de panel, c'est l'hétérogénéité individuelle qui est en cause. La stabilité temporelle est pour sa part abordée dans ce numéro par deux articles. Le premier de R. LUUKKONEN et T. TERÄSVIRTA – *Testing linearity of economic time series against cyclical asymmetry* – se pose le problème de l'asymétrie des phénomènes temporels. La modélisation généralement adoptée fait intervenir une transition autorégressive, c'est-à-dire un saut dans la partie autorégressive définie à partir d'un seuil connu. Cette transition brutale s'accorde mal avec des tests asymptotiques fondés sur la vraisemblance, puisque ceux-ci n'ont plus de justification asymptotique dans ce cas. Aussi R. Luukkonen et T. Teräsvirta proposent-ils une approximation continue du saut qui permet de plus d'estimer le seuil et aussi les délais qui pourraient déclencher le changement de régime du modèle dynamique. Les tests d'asymétrie sont fondés sur la vraisemblance. Ils mesurent en fait une certaine non-linéarité du modèle; ils peuvent aussi détecter des phénomènes d'hétéroscédasticité conditionnelle (ARCH). Cette méthodologie est appliquée à des séries trimestrielles (chômage, production industrielle, etc.) de treize pays de l'OCDE. Elle apparaît alors peu robuste aux transformations sur les variables ou à la spécification des modèles dynamiques.

Le second article consacré à l'hétérogénéité temporelle est proposé par M. J. BIERENS – *Least squares estimation of linear and nonlinear ARMAX models under data heterogeneity* –. Il se place dans le cadre des modèles de série temporelle et il définit l'hétérogénéité sous une forme relativement générale grâce au concept de « v -stabilité » qui englobe les modèles VAR (autorégressifs vectoriels) à perturbations de type moyenne mobile, et à la notion « d'hétérogénéité propre » des processus qui permettent de garantir des propriétés de convergence des moyennes empiriques de certaines fonctions de variables aléatoires à la base des processus temporels. Ces résultats généraux sont appliqués aux processus ARMAX qui ne sont rien d'autres que des processus autorégressifs à retards échelonnés. L'hétérogénéité transite dans ce travail par l'intermédiaire des variables exogènes (la partie X de l'ARMAX) qui suivent des processus hétérogènes où les paramètres peuvent être

variables dans le temps. M. J. Bierens démontre des propriétés de convergence et de normalité asymptotique des estimateurs des moindres carrés pour les modèles ARMAX soumis à l'hétérogénéité des processus exogènes; c'est-à-dire vérifiant la v -stabilité et l'hétérogénéité propre.

Pour juger précisément de la validité d'un modèle, la forme initiale du test de la matrice d'information de WHITE avait une prétention quasi-universelle. En fait, l'hypothèse alternative implicite de ce test a bien souvent une interprétation claire et plus réduite. Ainsi dans le cadre du modèle de régression linéaire ou non-linéaire, il s'avère que cette hypothèse se rapporte aux coefficients aléatoires et donc à l'hétérogénéité. C'est en s'appuyant sur ce résultat que R. DAVIDSON et J. G. MACKINNON – *une nouvelle forme du test de la matrice d'information* – proposent de remplacer ce test de WHITE par le test d'une régression étendue dite « double-longueur ». Une telle approche est d'autant plus importante pour le praticien que les versions classiques du test de la matrice d'information rejettent trop souvent l'hypothèse de bonne spécification, c'est-à-dire l'homogénéité, même pour des échantillons de petite taille. Par des simulations, mais surtout (dans un cadre élémentaire) par des développements asymptotiques, R. G. Davidson et J. C. MacKinnon prouvent la supériorité statistique de leur test au moins pour les modèles linéaires.

F. LAISNEY, M. LECHNER et S. STRØM – *Lessons from specification test for a labour supply model* – mettent à l'épreuve les tests de spécification. Leur domaine de recherche est celui de l'offre de travail des femmes mariées à partir des données de l'enquête budget des familles françaises de 1979. Leurs premières évaluations de l'élasticité du salaire sont apparues trop grandes et nos trois auteurs ont recherché une spécification plus souple du modèle ainsi qu'une possible hétérogénéité des comportements. Cette dernière est approchée sous deux formes : l'hétéroscédasticité et la variabilité des coefficients. Quelle que soit la spécification retenue, les données recèlent des traces importantes d'hétérogénéité qui font conclure à F. Laisney, M. Lechner et S. Strøm au côté très approximatif des outils de modélisation dans ce domaine.

Quelque peu en dehors des outils méthodologiques élaborés pour l'analyse ou la recherche de l'hétérogénéité, les trois derniers articles de ce numéro spécial, sont des applications économétriques sur données individuelles où les divers auteurs modélisent des différences de comportement.

R. L. LUMSDAINE, J. M. STOCK et D. A. WISE — *Fenêtres et retraites* — proposent une modélisation aléatoire de la prise de retraite par les salariés, où ceux-ci décident de stopper leur activité lorsque leurs anticipations montrent qu'ils n'ont pas intérêt à continuer. Par ce modèle, ils ont cherché à mettre en évidence l'effet des plans dits à « fenêtre » mis en place aux États-Unis au début des années quatre-vingt. Ces plans ont pour but de faire prendre la retraite assez tôt autour de 55 ans pour soutenir l'emploi des jeunes et favoriser les mutations organisationnelles et technologiques de l'industrie moderne. Ils offrent des avantages à la retraite anticipée pour des groupes de salariés, définis par l'âge, le statut ou par l'appartenance à un atelier ou une partie de l'entreprise à condition que l'arrêt d'activité ait lieu avant une date déterminée. Le modèle montre par exemple que ces plans augmentent de 50% la proportion des salariés employés à 52 ans et qui se retirent avant 60 ans.

L'enquête emploi néerlandaise a collecté le salaire d'avril 1985 déclaré immédiatement en avril 1985 et un an et demi plus tard en octobre 1986. J. HARTOG et H. VAN OPHEM — *Wages and measurement errors* — tentent sur ces données de déterminer les causes des erreurs de mémoire des enquêtés. Il s'avère que celles-ci ne sont pas uniquement dues au hasard puisque les changements de position professionnelle expliquent une bonne part des différences observées dans la déclaration des salaires aux deux enquêtes.

Enfin A. MOREAU et M. VISSER — *Durée du chômage des jeunes en France* — estiment un modèle de recherche de travail (Job search) sur les données de l'enquête sur l'emploi de 1986 restreintes aux jeunes de moins de 26 ans. Pour ce modèle, la probabilité de sortie du chômage se décompose en un taux d'arrivée d'offre d'emploi et une probabilité d'acceptation de cette offre. Dans la population étudiée, il apparaît que la probabilité d'acceptation n'est pas la composante la plus discriminante, mais qu'au contraire les durées de chômage varient essentiellement en fonction de la fréquence des offres d'emploi. Ainsi, les jeunes femmes et les non-diplômés sont-ils très défavorisés sur ce point.

*
* *

Au terme de cette présentation, il apparaît que les économètres ont développé de nombreux instruments de modélisation et d'inférence adaptés à l'hétérogénéité des observations que celles-ci soient en coupe instantanée, en série temporelle

ou en panel. Les avancées sont donc importantes sur le plan de la représentation statistique de la variabilité des comportements et sur celui de la maîtrise globale du phénomène. Mais les avancées méthodologiques devront être jugées à l'avenir sur leur capacité à faire progresser la connaissance des comportements individuels et à mesurer leur impact sur les phénomènes macroéconomiques.

Alain TROGNON

PROGRAMME OF THE INTERNATIONAL
CONFERENCE ON

**HETEROGENEITY
IN
ECONOMETRICS**

PARIS, JUNE 12-14, 1989

PLACE OF THE CONFERENCE:

ÉCOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION
ÉCONOMIQUE, 3, avenue Pierre-Larousse, 92240 MALAKOFF,
FRANCE.

ROOM: AMPHI. 1.

SPONSORS: ADRES, CEPREMAP, ENSAE.

ORGANIZERS: C. GOURIÉROUX, A. TROGNON.

MONDAY, JUNE 12

9.00 a.m.- 9.30 a.m.: Registration (Room 22).

9.30 a.m.-10.15 a.m.: *Opening Session.*

Chair: MILLERON J. C. (Directeur Général de l'INSEE).

GRANDMONT J.-M. (CEPREMAP): "Aggregation Issues in Economic Theory".

10.15 a.m.-10.45 a.m.: Coffee Break.

10.45 a.m.-12.15 a.m.: *Estimation Methods.*

Chair: CHAMBERLAIN G. (Harvard Univ.).

HECKMAN J. and WALKER J. (Yale Univ., Univ. of Wisconsin): "Testing the Mixtures of Exponential Hypothesis and Estimating the Mixing Distribution by the Method of Moments".

GOURIEROUX C. and MONFORT A. (CEPREMAP-CREST, INSEE): "Simulation Based Inference in Models with Heterogeneity".

HAJIVASSILIOU V. A. and MCFADDEN D. (Yale Univ., MIT): "Country Heterogeneity and External Debt Crisis. Estimation by the Method of Simulated Moments".

12.15 p.m.-2.00 p.m.: Lunch.

2.00 p.m.-3.30 p.m.: *Wages and Labour Supply*.

Chair: DAVIDSON R. (Univ. d'Aix-Marseille).

STOCK J. and WISE D. (Harvard Univ.): "Windows and Retirement".

LAISNEY F., LECHNER M., STRØM S. and NIEDERGESASS A. (Heidelberg Univ., Oslo Univ.): "Female Labour Supply in France: Lessons from Specification Tests".

HARTOG J. and VAN OPHEM H. (Amsterdam Univ., Leyden Univ.): "Wages and Measurement Error".

3.30 p.m.-3.45 p.m.: Break.

3.45 p.m.-5.00 p.m.: *Testing for Neglected Heterogeneity*.

Chair: HECKMAN J. (Yale Univ.).

CHESHER A. (Univ. of Bristol): "Finite Sample Properties of Tests for Neglected Heterogeneity".

DAVIDSON R. and MCKINNON J. (Univ. d'Aix-Marseille, Queen's Univ.): "A New Form of the Information Matrix Test".

5.15 p.m.: Drinks party.

TUESDAY, JUNE 13

9.15 a.m.-10.30 a.m.: *Estimation of Hazard Models*.

Chair: LANCASTER T. (Brown Univ.).

MEYER B. (Northwestern Univ.): "Semi-parametric Estimation of Hazard Models".

RIDDER G. (Groningen Univ.): "Heterogeneity, Duration and Attrition in Multi-wave Panel Data".

10.30 a.m.-11.00 a.m.: Coffee Break.

11.00 a.m.-12.15 a.m.: *Bayesian Approaches*

Chair: MONFORT A. (INSEE).

LUBRANO M. and BAUWENS L. (Univ. d'Aix-Marseille): "Bayesian Diagnostics for Heteroskedasticity".

FLORENS J.-P., HENDRY D. and RICHARD J.-F. (Univ. de Toulouse, Nuffield College, Oxford, Duke Univ.): "Encompassing and Specificity".

12.15 p.m.-2.00 p.m.: Lunch.

2.00 p.m.-3.30 p.m.: *Application to Unemployment Spells*.

Chair: DAGENAIS M. (Univ. de Montréal).

LALLA M. (Modene Univ.): "Analysis of Unemployment Spells Generated from a Queuing Model".

MOREAU A. and VISSER M. (INSEE, CREST): "Duration of Youth Unemployment in France".

3.15 p.m.-4.45 p.m.: *Panel Models*.

Chair: GRILICHES Z. (Harvard Univ.).

BOND S. and MEGHIR C. (Oxford, University College): "Dynamic Investment Models and the Firm's Financial Policy".

BHARGAVA A. (Houston Univ.): "Malnutrition and Role of Individual Variation".

8.30 p.m.: Conference Dinner.

WEDNESDAY, JUNE 14

9.15 a.m.-10.30 p.m.: *Aggregation Biases*.

Chair: TERASVIRTA T. (ETLA, Helsinki).

MAIRESSE J. and GRILICHES Z. (ENSAE, Harvard Univ.): "Are there Stable Production Functions?".

FORTIN N. (Univ. de Montréal): "A Unified Theory of Aggregation: Similarity and Separability reconsidered".

10.30 a.m.-11.00 p.m.: Coffee break.

11.00 a.m.-12.15 p.m.: *Semi-Parametric Estimation Methods*.

Chair: BIERENS H. (Univ. of Amsterdam).

CHAMBERLAIN G. (Harvard Univ.): "Efficiency Bounds for Semi-Parametric Regression".

ANDREWS D. (Yale Univ.): "Semi-Parametric Econometric Models".

12.15 p.m.-2.00 p.m.: Lunch.

2.00 p.m.-3.15 p.m.: *Dynamics and Heterogeneity*.

Chair: RIDDER G. (Groningen Univ.).

LIPPI M. and FORNI M. (Modene Univ.): "On the Dynamic Specification of Aggregated Models".

BLUNDELL R. and SMITH R. (Univ. College, Univ. of Manchester): "Efficient Estimation and Testing in Dynamic Panel Data Models".

3.15 p.m.-3.30 p.m.: Break.

3.30 p.m.-4.45 p.m.: *Nonlinearity*.

Chair: TROGNON A. (INSEE).

BIERENS H. (Univ. of Amsterdam): "Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Heterogeneity".

LUUKKONEN R. and TERASVIRTA (ETLA, Helsinki): "Testing Linearity of Economic Time Series Against Cyclical Asymmetry".

4.45 p.m.-5.15 p.m.:

Chair: CHESHER A. (Univ. of Bristol).

LANCASTER T. (Brown Univ.): "Some Results on Identification and Efficient Estimation of Duration Models".