

1977–1997

Changements et continuités en économétrie des données de panel

Patrick SEVESTRE *

Il y a un peu plus de 20 ans paraissait le numéro 30-31 d'*Annales de l'Insee* consacré à l'économétrie des données de panel. Ce numéro, dont plusieurs articles font encore référence, reprenait l'ensemble des contributions présentées à la conférence organisée à l'Insee en août 1977 par Pascal MAZODIER, Jacques MAIRESSE et Alain TROGNON. Une dizaine d'années plus tard, une autre conférence fut organisée, sur le même thème, à l'Université Paris XII-Val-de-Marne, à l'initiative des chercheurs de l'ERUDITE¹. Si celle-ci n'attira pas une participation aussi prestigieuse ni aussi nombreuse que la conférence de 1977, elle inaugura toutefois une série qui se poursuit aujourd'hui et qui, grâce à l'aide de Pietro BALESTRA, Badi BALTAGI, Jacques MAIRESSE et Alain TROGNON entre autres, est passée du statut de « *Journées d'étude* » à celui de « *Conférence internationale* » et a pris place parmi les lieux de rencontre privilégiés des spécialistes du domaine.

C'est donc à l'occasion d'un double anniversaire, celui des conférences de 1977 et de 1987, que l'ERUDITE a organisé, en 1997, la septième Conférence Internationale sur l'Économétrie des Données de Panel. De très nombreux participants à la conférence de 1977 y furent conviés à présenter de nouvelles contributions. Parmi ces « *pionniers* », je mentionnerai particulièrement Robert EISNER, Zvi GRILICHES et G.S. MADDALA, qui nous ont malheureusement quittés depuis. Tous les trois ont contribué de façon différente, mais très remarquable, à l'économétrie des panels².

Vingt années de développement de l'économétrie des données de panel...

Quels éléments de continuité, quels changements, la comparaison du présent numéro avec le n° 30-31 d'*Annales de l'Insee* permet-elle de repérer dans l'évolution de l'économétrie des données de panel au cours de ces vingt années ? La lecture des deux sommaires nous donne un premier élément de

* P. SEVESTRE : ERUDITE, Université Paris XII-Val-de-Marne.

1. Équipe de Recherches sur l'Utilisation des Données Individuelles-Temporelles en Économie.

2. Pour une publication reprenant les travaux de ces trois auteurs dans ce domaine et dans d'autres, voir EISNER [1998], GRILICHES [1998a,b] et MADDALA [1994].

réponse. Même si l'on ne peut pas vraiment exclure l'existence d'un « léger » biais de sélection, la fréquence des noms qui se retrouvent dans les deux participe de l'idée d'une certaine continuité. Cette constatation n'est pas totalement anecdotique. Elle montre que les pionniers de l'économétrie des panels, qui fondaient beaucoup d'espoir sur les possibilités d'amélioration de notre compréhension des mécanismes économiques qu'offraient ces données et les méthodes permettant de les exploiter, ont jugé ce nouveau champ suffisamment fertile pour continuer à le cultiver.

Un second élément de continuité peut être trouvé dans la forte imbrication des questionnements de nature économique avec les développements qu'ils induisent en matière de méthodes économétriques. Cette imbrication, initiée par l'article fondateur de Pietro BALESTRA et Marc NERLOVE [1966], se poursuit plus que jamais. La plupart des articles de méthode regroupés ici trouvent leur origine dans une question économique et comportent une application qui dépasse la seule illustration. Réciproquement, les articles présentant des travaux d'économie appliquée accordent une large place aux questions de méthode.

La thématique des recherches témoigne également d'une assez forte continuité. Ainsi, on trouve au sommaire des deux numéros des travaux appliqués portant sur les processus productifs des entreprises, sur leur comportement de demande de facteurs ou encore sur les déterminants des salaires et revenus individuels. Ces deux sommaires comportent aussi un nombre significatif d'articles méthodologiques relatifs à l'estimation des modèles dynamiques ou aux problèmes d'attrition et de sélection endogène des échantillons.

Quel(s) changement(s) peut-on, à l'inverse, déceler à la lecture comparée de ces deux numéros ? Le plus évident concerne la place prise récemment par la macroéconométrie sur panels de pays, et en particulier par les travaux relatifs à la croissance et à la convergence éventuelle des économies. Si l'on tient compte des articles méthodologiques fortement orientés dans cette direction, presque la moitié des articles de ce numéro y sont consacrés. Il en résulte des changements de perspective intéressants et, du point de vue des méthodes, la nécessité de développer de nouvelles techniques. En effet, l'essentiel des méthodes économétriques actuelles a été développé en relation avec les caractéristiques des panels microéconomiques utilisés jusqu'ici. Ces échantillons ayant le plus souvent une dimension temporelle limitée mais comportant un nombre d'individus élevé, les « bonnes » propriétés asymptotiques des méthodes actuelles supposent un grand nombre d'observations individuelles. Or, les panels de pays comportent généralement un nombre « moyen » de périodes et d'individus, et les méthodes « usuelles » ne sont donc pas toujours les mieux adaptées à ces données. De plus, comme le souligne G.S. MADDALA dans l'article qu'il consacre ici à l'économétrie de ces panels de pays, l'hétérogénéité marquée des comportements et l'existence de corrélations inter-individuelles sont deux éléments dont il importe aussi de tenir compte.

Finalement, que peut-on dire, d'une part, des apports de l'économétrie des données de panel à la connaissance économique, et, d'autre part, de la progression de la discipline elle-même ?

Lors de la conférence de 1997, Zvi GRILICHES, dans sa discussion de la communication présentée par Pascal MAZODIER et Alain TROGNON [1997] sur le développement de l'économétrie des panels au cours des vingt

dernières années, s'interrogeait sur le premier point. Dans l'article publié ici, G.S. MADDALA partage son interrogation quant à la pertinence de certains résultats économétriques visant à expliquer la croissance économique. Par contre, il considère que, du point de vue des méthodes, les progrès réalisés ont été indéniables. Venant de personnes comptant parmi les « *pères fondateurs* » de l'économétrie des données de panel, ces jugements ne peuvent être sous-estimés. Toutefois, on peut penser que ces pionniers attendaient peut-être trop, ou trop vite, de ces données et des méthodes visant à les exploiter au mieux. Si l'économétrie des données de panel s'est révélée moins fructueuse qu'escompté, elle a toutefois permis des avancées notables dans notre compréhension des comportements économiques. À titre d'exemple, la mise en évidence empirique de l'importance (sinon de la nécessité) de la prise en compte de l'hétérogénéité (observée et *inobservée*) des agents économiques dans l'analyse de leurs comportements peut être portée au crédit de l'économétrie des panels. Rechercher et évaluer précisément l'ensemble de ces apports à l'amélioration de notre connaissance du fonctionnement des économies supposerait une recension dépassant malheureusement le cadre de cette introduction.

Il reste que la perception accrue de l'intérêt de ces méthodes est indéniable. Il suffit pour s'en convaincre d'observer le développement considérable du recours aux données de panel en économie appliquée. Ainsi, le nombre de présences des mots « *panel data* » dans le titre ou les mots-clés des articles publiés dans les revues référencées dans EconLit a été multiplié par 25 entre 1975 et 1995 ! (Cf. P. MAZODIER et A. TROGNON [1997]). Nul doute que cela traduit, au moins dans une certaine mesure, l'intérêt des données de panel pour l'analyse des comportements économiques.

Ainsi, dans le premier article de ce numéro, Jacques MAIRESSE, Bronwyn HALL et Benoît MULKAY, concluent que, dans le domaine de l'analyse du comportement d'investissement des entreprises, les développements de l'économétrie des panels ont permis une meilleure compréhension de ce comportement, conduisant à des améliorations dans la modélisation et l'interprétation, et à une meilleure compréhension de ce qui peut, et ne peut pas, être mesuré.

... ont montré l'importance des biais d'endogénéité et de sélection...

L'objectif de cet article de J. MAIRESSE, B. HALL et B. MULKAY est double. D'une part, il cherche donc à répondre à la question de savoir si les progrès de l'économétrie des données de panel (tant du point de vue théorique que de la pratique) ont permis d'avoir une meilleure représentation, et une meilleure compréhension, du comportement d'investissement des entreprises. D'autre part, les auteurs cherchent à identifier d'éventuelles différences de comportement d'investissement, tant entre les entreprises françaises et américaines, qu'entre les deux périodes analysées : les années 70 et les années 80. Pour ce faire, les auteurs ont réalisé un travail considérable, consistant à analyser et à comparer les estimations de deux modèles (un modèle accélérateur-profit « *traditionnel* » et un modèle à correction d'erreur) par différentes méthodes (estimateurs « *standards* », inter- et intra-individuels, et méthode des moments généralisés) sur quatre échantillons : deux échantillons de firmes américaines et deux échantillons d'entreprises françaises (couvrant les années 1968–1979 et 1979–1993).

La comparaison des estimations obtenues par la méthode des moments généralisés avec celles associées à l'estimateur Intra-individuel montre au total assez peu de différences. Si le recours à la méthode des moments généralisés est supposé limiter les biais d'endogénéité et/ou d'erreur de mesure notamment (ce que ne permet pas l'estimateur Intra-individuel), ceci se fait au prix d'une moindre précision due à la faible corrélation des instruments avec les régresseurs endogènes. Par ailleurs, du point de vue des comportements d'investissement, il semble que les profits, qui jouaient un rôle significatif dans la décision d'investissement dans les années 70, tant en France qu'aux États-Unis, n'ont plus un tel effet dans les années 80 ; ce qui peut s'expliquer par la libéralisation des marchés financiers au cours de cette deuxième période, notamment en France.

L'article de N. COTE-COLISSON et F. LEGENDRE est en quelque sorte une illustration des pistes de recherche évoquées par J. MAIRESSE, B. HALL et B. MULKAY dans leur conclusion. Il vise à estimer l'élasticité de substitution entre le capital et l'emploi en distinguant les firmes à forte élasticité et celles à plus faible élasticité. Pour cela, un modèle à changement de régime est estimé, permettant une classification endogène des entreprises. L'équation déterminant le « *choix* » du régime est ici simplifiée, puisqu'elle se compose d'un effet individuel et d'une perturbation « *classique* ». L'hypothèse retenue est que les entreprises se trouvent toujours dans le même régime, hypothèse qui est tenable dès lors que l'on s'intéresse à un paramètre structurel de la technologie de production. Le modèle estimé comporte deux autres équations : l'une expliquant le ratio capital/travail, l'autre étant une forme réduite d'un modèle à deux équations, une fonction de production et une équation de détermination des prix. L'argument invoqué pour justifier cette spécification est qu'elle permet d'obtenir de meilleures estimations de l'élasticité de substitution que celles résultant de l'estimation de la seule fonction de production. Les estimations, réalisées à partir d'un échantillon d'environ 800 firmes industrielles, suivies sur la période 1981-1987, conduisent à partager les entreprises en deux groupes : le premier correspondant à une élasticité de substitution de 0.6 environ et le second à une élasticité de 0.9. Les auteurs s'attachent ensuite à rechercher les déterminants de l'élasticité de substitution de chacun des groupes. Celle-ci est d'autant plus élevée que les entreprises sont de grande taille, ont réduit précédemment leur ratio capital/travail et ont une main-d'œuvre peu qualifiée. On retrouve ici le résultat de Z. GRILICHES [1969], selon lequel le capital serait assez facilement substituable avec la main-d'œuvre non qualifiée mais plutôt complémentaire avec la main-d'œuvre qualifiée.

Les difficultés inhérentes à la sélection endogène des échantillons sont également au cœur de l'article de M. LECHNER consacré à l'analyse des effets de la formation continue en entreprise sur la rémunération et la probabilité d'emploi des salariés de l'ex-République Démocratique Allemande. L'intérêt de cette question n'est pas à démontrer tant les changements drastiques imposés à l'économie de l'ex-RDA par la chute du mur de Berlin ont nécessité et nécessitent encore une adaptation importante de sa main-d'œuvre dont, à l'évidence, la formation continue est un des vecteurs essentiels. Pour évaluer l'impact de celle-ci sur les salaires et la probabilité d'emploi des salariés de l'ex-RDA, M. LECHNER utilise des données concernant environ un millier de ces salariés, données tirées du panel socio-économique allemand

GSOEP (*German Socio-Economic Panel*) et couvrant la période 1990-1994. La question cruciale, dans ce type d'étude, est évidemment celle de l'évaluation correcte, *i.e.* toutes choses égales par ailleurs, des effets de la formation continue. On sait que la simple estimation, sur un échantillon de salariés, d'une équation expliquant le salaire par diverses variables, dont la formation continue, conduit à des résultats biaisés. En effet, la probabilité d'être concerné par une action de formation continue est endogène, notamment du fait de sa corrélation probable avec certaines caractéristiques inobservables des individus (leur adaptabilité par exemple), caractéristiques qui affectent également leur salaire. De plus, l'observation du salaire perçu suppose que l'individu ait conservé son emploi. Le traitement de cette question de la sélection est ici réalisé à l'aide de méthodes non paramétriques déjà développées par l'auteur (LECHNER [1999]). La conclusion principale qui peut être tirée des estimations réalisées est que la formation continue a un effet favorable sur la rémunération des salariés mais ne semble pas en avoir sur leur probabilité de garder leur emploi. Plusieurs explications sont avancées pour ce dernier résultat : soit la formation reçue au moment de la transition a perdu très rapidement de sa valeur, soit l'augmentation du salaire qui en a résulté a accru le risque d'être licencié, soit enfin les entreprises se sont vues contraintes de licencier massivement, sans tenir compte alors du capital humain de leurs salariés.

Le problème rencontré par M. LECHNER dans son évaluation de l'impact de la formation continue sur les revenus et le risque de chômage présente certaines similitudes avec celui analysé par R. MOFFITT, J. FITZGERALD et P. GOTTSCHALK. En effet, ces derniers s'intéressent à l'estimation d'un modèle économétrique lorsque la sélection se trouve être corrélée avec un régresseur *endogène* de l'équation d'intérêt. L'exemple le plus évident de ce type de situation est celui où l'on souhaite estimer un modèle autorégressif à partir d'un échantillon d'observations sélectionnées en fonction de la valeur de la variable endogène retardée, *i.e.* d'un échantillon sujet à un phénomène d'attrition. R. MOFFITT, J. FITZGERALD et P. GOTTSCHALK parlent dans ce cas de sélection fondée sur des variables observables, par opposition avec la situation généralement envisagée de sélection basée sur des variables inobservables. Ils proposent un estimateur convergent pour ce type de modèle, à savoir un estimateur des Moindres Carrés Pondérés dans lequel les pondérations sont calculées à partir d'estimations convergentes de l'équation de sélection. L'application de cette méthode à l'estimation d'une équation de salaire sur un échantillon de ménages extrait du PSID (*Panel Study of Income Dynamics*) montre que le biais d'attrition envisagé existe bien, même s'il concerne plutôt la constante et assez peu les coefficients des « véritables » variables explicatives.

M. ROCHINA-BARRACHINA s'est attelée pour sa part au problème plus « traditionnel », mais non nécessairement plus facile, de l'estimation d'un modèle économétrique avec sélection à partir de variables inobservables. La difficulté d'estimation de ce type de modèle est bien connue : la log-vraisemblance à maximiser est extrêmement complexe, de même que le sont les calculs nécessaires à l'obtention des régresseurs supplémentaires associés à la généralisation de l'estimateur d'HECKMAN en deux étapes au contexte des panels (*cf.* NIJMAN et VERBEEK [1995]). M. ROCHINA-BARRACHINA propose un estimateur relevant de cette approche, mais reposant sur des hypothèses différentes de celles considérées jusqu'ici dans la littérature (WOOLDRIDGE [1995]),

KYRIAZIDOU [1997]). La méthode proposée consiste dans un premier temps à estimer le modèle sur deux vagues d'observations consécutives en le différenciant pour éliminer les effets spécifiques et en conditionnant par la présence à ces deux vagues. Cela conduit à inclure deux termes correctifs dont la forme précise dépend des hypothèses sur le processus de sélection et la distribution jointe des inobservables. Dans un second temps, les différentes estimations correspondant à chaque couple de vagues successives sont combinées optimalement par un estimateur à distance minimale. Deux versions de cet estimateur sont présentées, correspondant respectivement à un estimateur paramétrique (resp. non paramétrique) de l'équation de sélection. Les propriétés à distance finie de ces estimateurs sont comparées à celles des estimateurs de KYRIAZIDOU [1997] et WOOLDRIDGE [1995] par le biais d'une étude de simulation. Il apparaît qu'ils jouissent d'une relative robustesse à divers types d'erreurs de spécification, notamment si on les compare aux estimateurs de KYRIAZIDOU et WOOLDRIDGE. Néanmoins, leur convergence repose sur l'exogénéité stricte des régresseurs, tant dans l'équation d'intérêt que dans l'équation de sélection. Si l'exogénéité stricte des régresseurs de l'équation d'intérêt peut être relâchée en recourant à la méthode des variables instrumentales, elle doit impérativement être vérifiée pour ce qui est de l'équation de sélection, ce qui peut être une hypothèse assez forte.

Dans l'article suivant, L. LILLARD cherche à évaluer l'impact des changements d'emploi et de la durée des emplois occupés sur la rémunération des salariés américains. Comme précédemment, l'endogénéité de ces déterminants doit être prise en compte pour espérer obtenir une évaluation correcte de leur influence. De fait, il est difficile d'exclure que ces variables soient corrélées avec certaines caractéristiques inobservables des salariés et de leurs emplois. Une première originalité de cet article réside dans l'utilisation qui est faite des données du NLSY (*National Longitudinal Survey of Youth*), données qui comportent des informations sur les individus (notamment leur calendrier d'emploi et les salaires correspondants) et sur l'emploi qu'ils occupent, ce qui permet d'identifier, dans les évolutions salariales, les contributions respectives des facteurs propres au salarié de ceux propres à l'emploi occupé. Du côté des salariés, l'objectif est de mesurer précisément l'effet des changements d'emploi et de la durée passée dans ces emplois sur les évolutions de salaire. Une seconde originalité tient au fait que ces effets sont croisés avec d'autres variables, permettant ainsi de mieux rendre compte de l'hétérogénéité des trajectoires salariales. Les estimations réalisées montrent une très forte hétérogénéité des salaires de départ, tant au niveau des salariés qu'à celui des emplois. De même, les impacts respectifs de l'expérience (au niveau individuel) et de l'ancienneté (au niveau des emplois occupés) sur les salaires sont très variables.

... et la nécessité de développer des méthodes nouvelles

L'article de G. CHAMBERLAIN et K. HIRANO est également relatif à l'évolution des salaires, mais il se situe dans une perspective quelque peu différente. Ils proposent une méthode permettant de déterminer la distribution conditionnelle des revenus futurs d'un individu, à partir d'informations le concernant personnellement et de données statistiques relatives à un échantillon d'autres individus. Une originalité du modèle proposé, par rapport à d'autres travaux

antérieurs, est de permettre une hétérogénéité au niveau de la variabilité des revenus futurs. Pourquoi chercher à déterminer cette distribution des revenus futurs ? Un exemple évident d'application de ces résultats est celui de l'analyse empirique des comportements de consommation. En effet, celle-ci suppose, dès lors que l'on se place dans un cadre inter-temporel, que l'on se donne *a priori* une distribution conditionnelle des revenus futurs, au même titre que l'on choisit une forme particulière de fonction d'utilité. Pour illustrer empiriquement la méthode proposée, G. CHAMBERLAIN et K. HIRANO réalisent des estimations à partir d'un extrait du PSID comprenant 813 hommes âgés de 24 à 33 ans. Les résultats obtenus montrent l'importance empirique de l'hétérogénéité des variances des revenus futurs, hétérogénéité qui accroît bien sûr celle des distributions conditionnelles de ces revenus.

Dans leur article, J. ABREVIAYA et J. HAUSMAN s'intéressent à un problème d'une tout autre nature. La question posée est celle de l'estimation de modèles économétriques non linéaires (modèles à variable dépendante qualitative, modèles de durée, modèles avec variables censurées...) lorsque la variable endogène est mesurée avec erreur. Comme ils le font remarquer justement, le fait que, dans les modèles linéaires, une erreur de mesure sur la variable expliquée n'entraîne pas de biais a conduit les économètres à ignorer cette difficulté dans le cadre plus général des modèles non linéaires. J. ABREVIAYA et J. HAUSMAN montrent que la convergence des méthodes paramétriques permettant d'estimer ce type de modèles nécessite une modélisation correcte du problème de mesure et que, de plus, leur mise en œuvre devient extrêmement complexe dès lors que le modèle à estimer n'est pas un modèle à variable dépendante qualitative. Ils proposent comme alternative de recourir à un estimateur non paramétrique, l'estimateur à rang monotone proposé par CAVANAGH et SHERMAN [1998] dont la convergence suppose seulement une condition de dominance stochastique, condition qui peut s'appliquer à de nombreuses formes d'erreur de mesure. Cet estimateur présente par contre, dans certains cas, l'inconvénient de ne permettre l'estimation des coefficients qu'à un facteur d'échelle près. C'est le cas notamment des modèles de durée sur lesquels néanmoins un accent particulier est mis dans l'article et dont une application empirique est présentée. Celle-ci concerne la durée du chômage, dont divers facteurs explicatifs sont évalués. Les estimations obtenues par l'estimateur à rang monotone conduisent à la conclusion que les allocations chômage n'ont aucun impact sur la durée du chômage, à la différence de ce qu'indiquent les résultats associés à d'autres méthodes d'estimation plus traditionnelles.

L'article de B. BALTAGI montre quant à lui l'intérêt que présente le principe des « *régressions augmentées* » (ou régressions « *artificielles* ») pour tester un assez vaste ensemble d'hypothèses relatives aux modèles de panel linéaires et non linéaires. Le principe général de ces méthodes consiste à ajouter au modèle économétrique d'intérêt une (des) variable(s) supplémentaire(s) (et une équation supplémentaire, pour les « *régressions doublées* ») dont le but est de simplifier la procédure d'estimation et/ou la mise en œuvre de tests d'hypothèses ³. Ces méthodes généralisent des résultats antérieurs tels que

3. Une présentation de ces méthodes, qui englobent les « régressions de GAUSS-NEWTON », les « *régressions doublées* » (« *double-length* ») et les « *régressions binaires* » (*binary response model regression*) peut être trouvée dans DAVIDSON et MACKINNON [1993].

celui, bien connu, d'HECKMAN [1976] pour le traitement des biais de sélection ou encore celui de GOURIEROUX et TROGNON [1998] qui ont montré comment l'adjonction d'un terme supplémentaire à divers types de modèles économétriques permet d'obtenir des estimations convergentes et asymptotiquement efficaces. De même, ces méthodes recouvrent, comme cas particulier, la suggestion de MUNDLAK [1978] consistant, pour tester l'existence d'effets corrélés, à tester la nullité des coefficients des régresseurs supplémentaires constitués par les moyennes individuelles des régresseurs du modèle d'intérêt. B. BALTAGI adapte ces méthodes aux modèles de panel et montre comment tester, à l'aide de régressions « *augmentées* », l'hypothèse d'absence d'effets individuels dans un modèle linéaire, dans un modèle à variable dépendante qualitative ou encore comment tester un modèle linéaire contre un modèle log-linéaire. Il montre également que le test d'HAUSMAN, l'approche de CHAMBERLAIN, les tests d'absence de biais de sélection proposés par WOOLDRIDGE [1995] constituent aussi des cas particuliers de ces méthodes de régression augmentée.

Par analogie avec les difficultés que présente l'estimation sur séries temporelles de modèles mettant en relation des variables exhibant une tendance, C. GRANGER et N. HYUNG s'intéressent au problème de l'estimation d'un modèle sur données de panel lorsque les variables sont fortement corrélées à la taille des individus. Ils montrent que si la taille n'est pas prise en compte dans la régression, on aboutit alors à une régression fallacieuse. Ils montrent en particulier que la mise en œuvre de tests de racine unitaire tels que celui proposé par LEVIN et LIN [1992] risque alors de conduire à accepter l'hypothèse de non stationnarité à tort. De plus, si les variables de taille ne sont pas fixes dans le temps mais varient très lentement (comme par exemple, la population, voire le revenu), ou encore si elles ont des queues de distribution épaisses, les prédictions hors-échantillon tirées des estimations sont biaisées. Une étude de simulation permet de quantifier les biais dans diverses configurations.

C. MEGHIR et F. WINDMEIJER s'intéressent quant à eux à une généralisation du modèle dynamique à erreurs composées « *standard* » dans laquelle la variance conditionnelle dépend de façon multiplicative d'effets spécifiques individuels. Ils exhibent des conditions d'orthogonalité permettant d'obtenir des estimations convergentes des paramètres dans plusieurs configurations, notamment lorsque les perturbations suivent un processus de moyenne mobile et lorsque les effets individuels de l'équation d'intérêt interagissent de façon multiplicative avec des effets fixes temporels « *à la HOLTZ-EAKIN, NEWEY et ROSEN [1988]* ». Les résultats de l'étude de simulation menée par les auteurs montrent que si l'estimation des paramètres des modèles considérés est matériellement possible, l'obtention de bonnes estimations de ces paramètres suppose que l'on dispose d'un très grand nombre d'observations dans la dimension individuelle.

Un développement récent de la macroéconométrie sur données de panel

L'article de T. J. WANSBEEK et T. KNAAP est, lui aussi, consacré à l'estimation de modèles dynamiques sur données de panel. Le modèle envisagé correspond au type de régressions rencontrées dans l'analyse de la croissance et de la convergence éventuelle des économies : il comporte des effets indivi-

duels spécifiques ainsi qu'un *trend* dont le coefficient diffère d'un individu à l'autre. Considérant ces paramètres comme des paramètres de nuisance, une double différenciation du modèle permet alors de les éliminer. Bien évidemment, l'application des MCO à ce modèle différencié conduit à des estimations non convergentes. Plusieurs solutions sont alors envisagées : la première est un estimateur des « *Moindres Carrés Ordinaires Modifiés* ». Celui-ci repose sur le calcul du biais asymptotique de l'estimateur des MCO sous l'hypothèse de stationnarité du processus ayant engendré les données. L'estimateur des MCO modifiés s'exprime alors comme une fonction de la seule estimation des MCO. Malheureusement, l'ensemble des valeurs possibles de cet estimateur se situe entre $-2/3$ et $-1/2$ (pour une valeur du vrai paramètre comprise entre 0 et 1) ; ce qui disqualifie complètement cette méthode. Les auteurs proposent donc de recourir à d'autres méthodes, de type Variables Instrumentales et Moments Généralisés. Ils présentent également un estimateur du « *maximum de vraisemblance à information limitée* » dont l'étude de simulation réalisée montre qu'il a un meilleur comportement dans des échantillons de taille finie ($N \leq 200$, $T \leq 15$) que les autres estimateurs considérés. Malheureusement, la distribution asymptotique de cet estimateur n'est pas connue, ce qui rend quelque peu délicate la mise en œuvre de tests. Toutefois, un estimateur des moments généralisés « *sous-optimal* », *i.e.* n'exploitant pas toutes les conditions d'orthogonalité (estimateur appelé IV3 par les auteurs), et dont la variance asymptotique est connue, apparaît avoir un comportement tout à fait satisfaisant.

R. CERMEÑO s'intéresse également à l'estimation de modèles dynamiques sur données de panel. Il montre comment l'estimateur basé sur la médiane, proposé par ANDREWS [1993] pour l'estimation de modèles autorégressifs sur séries temporelles, peut être étendu au cas des modèles autorégressifs à effets fixes. L'idée est voisine de celle de l'estimateur des MCO modifiés considéré par T. J. WANSBEEK et T. KNAAP : on applique une correction à l'estimateur intra-individuel (*i.e.* l'estimateur des MCO du modèle à effets fixes) visant à faire coïncider la médiane de la distribution de l'estimateur (obtenue par simulation) et l'estimation obtenue. La valeur du paramètre pour laquelle ces deux grandeurs coïncident est appelée estimateur sans biais basé sur la médiane. À l'aide d'une étude de simulation, R. CERMEÑO étudie les performances de cet estimateur. Il montre en particulier qu'il est robuste vis-à-vis de l'hétéroscédasticité et de l'autocorrélation dans la dimension individuelle. Ces propriétés sont importantes dans la mesure où l'un des champs privilégiés d'application de cette méthode est l'analyse de la convergence : or les données de pays utilisées pour ce faire conduisent fréquemment à ce genre de problème. Il faut, toutefois, noter une limite importante de cette méthode : elle n'est justifiée rigoureusement que dans le cas des modèles autorégressifs purs. Dans la pratique, la correction du biais qu'elle implique peut donc être trop importante puisque l'on sait que la présence de régresseurs exogènes conduit, si le modèle est correctement spécifié, à réduire le biais de l'estimateur intra-individuel. Une application de cette méthode à l'étude de la convergence des économies conduit à la conclusion que l'hypothèse de convergence ne peut être retenue que pour les États américains et pour ceux de l'OCDE.

L'article de M. NERLOVE est également relatif aux problèmes d'estimation de modèles dynamiques sur données de panel, mais il se situe dans une pers-

pective assez différente des précédents. Il développe une argumentation convaincante pour montrer que la seule analyse des estimations du maximum de vraisemblance peut donner une vision exagérément optimiste de la précision de ces estimations. M. NERLOVE milite en faveur de l'étude plus complète de la vraisemblance des modèles économétriques estimés. En effet, il montre que si la vraisemblance est assez plate dans certaines directions, *i.e.* pour certains paramètres, la précision des estimations évaluée au maximum de la vraisemblance donne une image trompeuse de leur fiabilité. Pour analyser le profil de la vraisemblance dans les différentes directions correspondant aux paramètres du modèle, M. NERLOVE propose de la concentrer par rapport à un sous-ensemble de paramètres et d'étudier les « *tranches* » ainsi obtenues. Un autre aspect intéressant, et important, de cet article concerne le traitement économétrique des observations initiales dans les modèles dynamiques de panel. M. NERLOVE développe les vraisemblances conditionnelles et non conditionnelles associées au modèle dynamique à erreurs composées « *standard* » et à un modèle incluant des *trends* spécifiques à chaque individu. Il montre comment, par différenciation, l'estimation de ce dernier peut se ramener à celle du premier. Il analyse ensuite les estimations du maximum de vraisemblance d'une équation de convergence basée sur le modèle de croissance de SOLOW-SWAN. Une des conclusions importantes est que bien que la vraisemblance soit assez plate dans la direction du paramètre correspondant à la part de la variance de l'effet individuel dans la variance totale des perturbations, sa précision estimée, au maximum de la vraisemblance, est assez bonne, donnant ainsi une fausse idée de la précision de cette estimation. Par ailleurs, les estimations du modèle avec *trend* spécifique conduisent à une évaluation de la vitesse de convergence des économies sensiblement plus élevée que celle tirée des estimations du modèle dynamique à erreurs composées.

C'est également à cette question de la convergence que s'attachent G. GAULIER, C. HURLIN et P. JEAN-PIERRE dans leur article. Ils proposent une généralisation de la méthode proposée par EVANS et KARRAS [1996] pour tester cette hypothèse. Rappelons que le test d'EVANS et KARRAS est un test de racine unitaire, dans lequel l'hypothèse nulle stipule l'existence d'une racine unitaire pour tous les pays, *i.e.* l'absence de convergence de tous ces pays. La généralisation proposée par les auteurs consiste à permettre une autocorrélation des perturbations différente pour chaque pays. La mise en œuvre de leur procédure de test sur plusieurs échantillons conduit à accepter l'hypothèse de convergence en ce qui concerne les pays de l'OCDE mais à la rejeter pour un ensemble de pays plus hétérogène, *i.e.* comprenant notamment des pays moins avancés. L'analyse qu'ils effectuent des estimations des effets individuels propres à chaque pays conduit à la conclusion que l'investissement, tant public que privé, explique les disparités observées de PIB par tête.

Les articles précédents conduisent à une double constatation : d'une part, les années récentes ont connu un développement important de l'économétrie sur des panels de pays, notamment en relation avec la question de la convergence des économies ; d'autre part, les difficultés rencontrées pour tester correctement cette hypothèse sont nombreuses. Cette constatation montre tout l'intérêt de l'article de G.S. MADDALA, qui procède à une évaluation critique de cette littérature. Pour illustrer son propos, il compare les travaux récents réalisés à partir de régressions inter-pays à « *la recherche d'un chat noir dans*

une chambre noire ». Selon lui, la première difficulté rencontrée dans ces travaux est relative à la qualité des données statistiques. Pour de nombreux pays, notamment les pays en développement, les données disponibles sont peu fiables et rendent de ce fait assez délicate la mise en œuvre sans précaution des techniques de l'économétrie des panels à des échantillons contenant ces pays. Par ailleurs, la question de l'hétérogénéité des trajectoires de croissance, comme celle des corrélations entre les perturbations du modèle pour différents pays sont le plus souvent ignorées, alors qu'elles peuvent affecter les résultats de manière significative. De fait, les résultats obtenus sont extrêmement divers et conduisent à des conclusions de politique économique parfois surréalistes. Comme C. GRANGER et N. HYUNG, G.S. MADDALA est quelque peu circonspect quant à la robustesse des conclusions que l'on peut tirer des tests de racine unitaire réalisés sur données de panel. Il insiste en particulier sur l'erreur, couramment faite, consistant à considérer à tort que le rejet de l'hypothèse nulle (de non convergence) permet de conclure à la convergence de tous les pays. Cette conclusion est erronée puisqu'il suffit que certains pays convergent, d'autres non, pour refuser l'hypothèse nulle. G.S. MADDALA conclut néanmoins que les problèmes soulevés par l'économétrie des panels de pays ont conduit, sur le plan économétrique, à certains progrès, dont plusieurs des articles qui précèdent sont une illustration. Si la recherche du « *chat noir* » n'a pas abouti en tant que telle, elle a quand même permis de trouver quelque chose !

L'article de P. HULTBERG, M. NADIRI, et R. SICKLES constitue en quelque sorte une réponse à l'une des critiques émises par G.S. MADDALA à l'encontre des travaux économétriques relatifs à la croissance des économies. Il vise, en effet, à rendre compte de certaines des interactions qui peuvent exister entre pays, à savoir, ici, les interactions fondées sur des effets de diffusion technologique. Les auteurs proposent, en effet, un modèle de croissance néo-classique dans lequel le rattrapage de son retard technologique peut, par le biais de flux technologique, conduire un pays à connaître une croissance plus rapide. Toutefois, cette possibilité de « *rattrapage* » peut être limitée par l'existence de freins, tant institutionnels que sociologiques ou politiques. Le modèle proposé permet ainsi l'estimation d'un « *taux d'adoption technologique* », mesurant la capacité des différents pays à s'approprier la technologie du pays considéré comme le plus avancé, les États-Unis. Différentes méthodes d'estimation sont mises en œuvre sur trois échantillons de pays couvrant respectivement l'Europe, l'Amérique Latine et l'Asie du Sud-Est. Les estimations du taux d'adoption technologique montrent que ce dernier est le plus élevé pour les pays d'Amérique Latine. Toutefois, l'insuffisance de l'investissement dans ces pays explique la persistance de leur retard. À l'inverse, les pays d'Asie du Sud-Est ont assez fortement bénéficié de ces effets de diffusion technologique. L'Europe, pour sa part, ne semble pas avoir tiré fortement profit de tels effets, ceci s'expliquant sans doute par la forte proximité des technologies européenne et américaine.

Enfin, dans le dernier article de ce numéro spécial, Y. MUNDLAK, D. LARSON et R. BUTZER proposent une analyse des différentiels de productivité agricole entre pays qui présente quelques points communs avec celle proposée par P. HULTBERG, M. NADIRI et R. SICKLES. En effet, la question posée dans cet article est celle de l'évaluation correcte de la contribution du capital et de l'emploi à la production, tenant compte du fait que les niveaux technolo-

giques diffèrent d'un pays à l'autre et sont endogènes, mais également des facteurs structurels propres à chaque pays. L'évaluation proposée repose sur la décomposition de la variabilité totale de la production. La variabilité inter-temporelle est ainsi supposée rendre compte du progrès technologique, la variabilité inter-individuelle traduisant l'influence des variables structurelles, à technologie donnée, et la variabilité intra-individuelle-temporelle retraçant pour sa part l'effet des variations dans les *outputs*, les *inputs* et les autres variables d'état. Les auteurs défendent donc l'idée que les estimations des élasticités de la production par rapport au travail et au capital obtenues par régression dans la dimension intra-individuelle-temporelle sont les plus pertinentes, car elles isolent les contributions propres de ces facteurs, contrairement aux estimations réalisées dans les autres dimensions.

Cet article montre qu'une analyse attentive de résultats produits par la mise en œuvre de méthodes d'estimation assez simples permet une bonne appréhension de la contribution des divers facteurs (*inputs*, technologie, facteurs structurels – institutionnels, sociologiques... –) aux différentiels de productivité observés d'un pays à l'autre. Il permet donc de clore cette introduction sur une vision assez optimiste de l'apport de l'économétrie des données de panel à la connaissance économique. ■

• Références bibliographiques

- ANDREWS D.W.K. (1993). – « Exactly Median-Unbiased Estimation of First Order Autoregressive / Unit Root Models », *Econometrica*, 61, pp. 139-165.
- BALESTRA P., NERLOVE M. (1966). – « Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Economic Model: The Demand for Natural Gas », *Econometrica*, 34, pp. 585-612.
- CAVANAGH C., SHERMAN R.P. (1998). – « Rank Estimators for Monotonic Index Models », *Journal of Econometrics*, 84, pp. 351-381.
- DAVIDSON R., MACKINNON J.G. (1993). – « Estimation and Inference in Econometrics », *Oxford University Press*.
- EISNER R. (1998). – « The Keynesian Revolution, Then and Now » (Vol. 1), and « Investment, National Income and Economic Policy » (Vol. 2), Series: Economists of the Twentieth Century, Edward Elgar Publishing Company.
- EVANS P., KARRAS G. (1996). – « Convergence revisited », *Journal of Monetary Economics*, 37, pp. 249-265.
- GOURIEROUX C., TROGNON A. (1988). – « Une note sur l'efficacité des procédures en deux étapes », in *Mélanges Économiques*, Essais en l'honneur d'Edmond Malinvaud, *Economica*.
- GRILICHES Z. (1969). – « Capital-Skill Complementarity », *Review of Economics and Statistics*, 56, pp. 465-468.
- GRILICHES Z. (1998a). – « R&D and Productivity: The Econometric Evidence », University of Chicago Press.
- GRILICHES Z. (1998b). – « Practicing Econometrics: Essays in Methods and Applications », Series: Economists of the Twentieth Century, Edward Elgar Publishing Company.
- HECKMAN J. (1976). – « The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models », *Annals of Economic and Social Measurement*, 5, pp. 475-492.
- HOLTZ-EAKIN D., NEWEY W.K., ROSEN H.S. (1988). – « Estimating Vector Autoregressions with Panel Data », *Econometrica*, 56, pp. 1371-1395.
- KYRIAZIDOU E. (1997). – « Estimation of a Panel Data Sample Selection Model », *Econometrica*, 65, pp. 1335-1364.
- LECHNER M. (1999). – « Earnings and Employment Effects in Continuous Off-the-Job Training in East Germany after Unification », à paraître dans *Journal of Business and Economic Statistics*.
- LEVIN A., LIN C.F. (1992). – « Unit Root Test in Panel Data: Asymptotic and Finite Samples Properties », UCSD *working paper* n° 92-23.
- MADDALA G.S. (1994). – « Econometrics Methods and Applications » (Vol. 1 and 2), Series: Economists of the Twentieth Century, Edward Elgar Publishing Company.
- MAZODIER P., TROGNON A. (1997). – « 20 Years of Publications: Panel Data on Panel Data? », *Mimeo*, communication à la Septième Conférence Internationale sur l'Utilisation des Données de Panel.
- MUNDLAK Y. (1978). – « On the Pooling of Time-Series and Cross-Section Data » *Econometrica*, 46, pp. 69-85.
- NIJMAN T., VERBEEK M. (1996). – « Incomplete Panels and Selection Bias », chapter 18, in « *The Econometrics of Panel Data* », A Handbook of the Theory with Application, second edition, Kluwer Academic Publishers.
- WOOLDRIDGE J.M. (1995). – « Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions », *Journal of Econometrics*, 68, pp. 115-132.