

Manuscrit déposé en novembre 1990

Daniel COURGEAU

Institut National d'Etudes
Démographiques, Paris

Dimensions spatiales du cycle de vie: les apports des enquêtes rétrospectives

Si le démographe s'est longtemps attaché à analyser séparément chaque phénomène démographique à l'état pur (L. Henry, 1972), force nous est maintenant de constater la nécessité de dépasser cette approche. Les diverses étapes du cycle familial ne se produisent pas indépendamment de la vie professionnelle et sociale de l'individu, ni du milieu spatial dans lequel il évolue et se déplace. En fait, ces divers aspects de la vie personnelle sont en étroite interaction les uns avec les autres.

Il faut cependant voir que les sources statistiques classiques (recensements, données de l'état civil, enquêtes transversales, etc.) ne permettaient pas d'avancer bien avant cette étude d'interactions : les recensements et les enquêtes transversales, par leur incapacité à saisir le longitudinal; les données de l'état civil, par le peu d'informations qu'elles comportent sur les individus concernés. Une réflexion approfondie sur ces sources a permis la mise en place de méthodes satisfaisantes pour les analyser : pour les données transversales, les méthodes de régression se révèlent bien adaptées; pour les données d'état civil, les méthodes d'analyse longitudinale classi-

ques ont entièrement fait leurs preuves. Dès lors que l'on désire analyser l'interaction entre vie familiale, professionnelle et sociale, dans un espace-temps propre, il paraît nécessaire de reformuler les bases mêmes de la recherche.

Notons d'abord que cette interaction ne se produit pas dans un espace-temps abstrait mais prend place dans une structure sociale particulière, qui est un élément premier à prendre en compte. Les sociologues nous ont montré que l'on peut définir ces structures sociales par un certain nombre de systèmes de relations auxquels sont associées des conduites particulières, plus ou moins développées selon le groupe ou la société considéré (Kimball et Pearsall, 1954) : systèmes familial, économique, politique, religieux, d'éducation, associatif et informel. Notre approche ne considérant pas une société comme fermée, mais au contraire en perpétuelle évolution, de nouveaux systèmes peuvent se mettre en place à l'avenir.

Dès lors, ce sera l'implication d'un individu dans ces divers systèmes de relation et l'interaction entre ces divers types d'implication, qui vont engendrer un espace et un

temps propre à chaque situation. On voit donc que l'espace et le temps ne sont pas des éléments premiers mais qu'ils ont leur origine et acquièrent leur signification dans les systèmes de relations.

Voyons plus en détail comment formaliser cette approche et les hypothèses qui permettent cette formalisation.

Supposons que nous travaillons sur une population donnée, observée à un instant t et occupant un certain espace avec une densité connue à ce même instant. La première hypothèse que nous formulons est qu'il existe une probabilité (conditionnée par l'expérience et les localisations passées de l'individu ainsi que par des systèmes de relations existant dans cette société) pour qu'il connaisse un changement donné dans ses caractéristiques familiales, économiques, etc., ou/et dans sa localisation spatiale, pendant le très court intervalle de temps ($t, t+dt$). Cette hypothèse entraîne également une autre condition : le changement d'état ou de localisation de l'individu se fait suffisamment rapidement pour que l'on puisse travailler en temps continu (1).

Pour pouvoir estimer ces probabilités, il faut encore faire une hypothèse supplémentaire. En effet, nous avons défini jusqu'à présent de façon suffisamment vague pour pouvoir englober le plus grand nombre de situations, les caractéristiques et la localisation de l'individu. Il nous faut maintenant supposer que ces caractéristiques et ces localisations sont en nombre fini et ce nombre devra être assez faible pour permettre l'estimation des probabilités précédentes. Ainsi, par exemple, on pourra définir dans un pays une zone métropolitaine opposée à une zone non métropolitaine pour caractériser la localisation d'un individu par une simple variable binaire. De même, on pourra simplifier la vie matrimoniale d'un individu en ne considérant que deux états possibles: marié ou non marié, ce dernier état englobant les célibataires, les veufs, les divorcés.

Ayant ainsi défini les objectifs de cette analyse, il nous faut voir comment recueillir les données pour la réaliser et présenter rapidement les méthodes d'analyse de ces biographies.

A. ENQUETES RETROSPECTIVES

Le recueil habituel de données démographiques, par recensements ou par les registres d'état civil (2), ne permet pas de saisir l'ensemble des événements, avec leur date et leur localisation, ainsi que les diverses caractéristiques des individus et des lieux en vue d'une telle analyse.

Il faut donc réaliser des enquêtes spécifiques si l'on veut disposer de toute l'information permettant cette analyse. Ces enquêtes peuvent être soit rétrospectives soit prospectives. Ces dernières permettent de recueillir avec une grande précision les divers événements, leur localisation et leur date, à un instant généralement très proche

de l'instant où ils se produisent. Malheureusement, elles nécessitent un long suivi d'une population dont on risque de perdre de nombreux individus au cours du temps. C'est la raison pour laquelle on préfère le plus souvent réaliser des enquêtes rétrospectives. Comme on interroge rétrospectivement les individus qu'une seule fois, les risques de perte sont liés uniquement aux non réponses à l'enquête. En revanche, les problèmes de mémoire, en particulier pour des événements peu marquants ou survenus il y a bien longtemps, risquent fortement de fausser l'analyse.

Il importe donc de s'assurer qu'un tel

(1) Des modèles en temps discret peuvent cependant être utilisés de façon similaire. Voir, par exemple, Kalbfleish et Prentice (1980), pp. 35-38.

(2) Les rares pays qui possèdent des registres de population sont mieux armés pour réaliser une telle analyse, mais le nombre restreint d'informations que comportent ces registres sur d'autres caractéristiques des individus (profession, éducation, etc.) ne permettent pas de la pousser très avant.

recueil rétrospectif n'introduit pas de biais importants pour l'analyse. Les résultats d'une enquête que nous venons de réaliser en Belgique (M. Poulain, B. Riandey et J.M. Firdion, 1990; D. Courgeau, 1990), viennent répondre à cette question.

L'existence en Belgique de registres de population recueillant tous les événements familiaux et les migrations, permet d'effectuer ce test. Pour ce faire, nous avons interrogé d'abord séparément et ensuite simultanément chaque conjoint d'un couple, en vue de confronter leurs réponses à celles du registre. L'échantillon était de 445 couples. Cette confrontation montre une bonne mémoire des événements familiaux, meilleure chez les femmes que chez les hommes. Si l'on prend un écart de plus ou moins un an autour de la date donnée par le registre, les mariages sont bien datés dans 99,8 % des cas chez les femmes (98,9 % chez les hommes) et les naissances d'enfant sont bien datées dans 99,3 % des cas chez les femmes (98,4 % chez les hommes). Le couple réuni donne une réponse encore meilleure que celle de la femme. En revanche, pour les migrations, en prenant le même écart, elles sont bien datées dans 76,7 % des cas seulement chez les femmes (69,8 % chez les hommes).

Si la mémoire des dates de migration est incorrecte, on peut douter des résultats des analyses effectuées à partir de telles enquêtes. C'est la raison pour laquelle nous avons réalisé les divers types d'analyses présentés plus loin dans cet article, sur ces données de migration. L'analyse non-paramétrique des durées de séjour dans les divers logements occupés depuis le mariage donne des résultats non discernables quelle que soit la source utilisée. L'analyse

paramétrique de ces mêmes durées en fonction de diverses caractéristiques de l'individu en début de séjour conduit également à des résultats non significativement différents. La troisième analyse semi-paramétrique portait sur l'interaction entre première migration et première naissance après mariage. Les résultats restent encore proches quelle que soit la source utilisée, mais la mauvaise qualité des réponses des hommes rend l'effet de certaines caractéristiques non significatif, alors qu'il l'était pour les femmes, les conjoints réunis et le registre.

On peut donc conclure de cette analyse que si les erreurs de datation (des migrations, en particulier) peuvent être importantes, dans les enquêtes rétrospectives, il semble que ces erreurs ne modifient pas ou changent peu l'ordre logique dans lequel les événements de la vie de l'enquêté se produisent. Cet ordre est correctement mémorisé, les erreurs de datation jouant le rôle d'un bruit de fond d'où l'on peut extraire une information cohérente quelle que soit la source utilisée. La mémoire semble donc fiable là où l'analyse l'exige.

Le questionnaire d'une telle enquête rétrospective devra couvrir le maximum d'informations sur l'espace vécu par l'enquêté tout au long de son existence : ses lieux de résidence successifs avec les dates des migrations, ses lieux de travail successifs avec les dates de changement, ses lieux de résidence secondaire, etc. Le questionnaire de l'enquête «Triple biographie» (D. Courgeau, 1988, pp. 113-118) permet une bonne saisie de ces divers lieux, simultanément à la localisation des divers événements familiaux et au recueil d'information d'intérêt sur la vie de l'enquêté, de ses parents et de ses enfants.

B. METHODES D'ANALYSE

Une fois cette information saisie dans son plus grand détail, diverses méthodes d'analyse sont mises en place pour l'analyser. Il s'agit maintenant de suivre les divers événements tout au long de la vie des individus, de voir les interrelations qui

existent entre eux et l'effet de diverses caractéristiques sur les étapes successives de la vie des enquêtés.

Il est possible d'aborder des aspects très différents de la mobilité spatiale : ensemble des changements de logement, migra-

tions vers ou hors des métropoles, etc. Selon l'aspect considéré, les méthodes d'analyse les plus satisfaisantes peuvent être très différentes.

1°) Analyse paramétrique de l'ensemble des migrations.

Supposons que l'on s'intéresse d'abord à l'ensemble des migrations, définies comme changements de logement ou comme changements d'autres unités spatiales (départements, régions, etc.). On dispose alors de la durée de séjour dans chaque logement (département, région, etc.) occupé par l'enquêté, que l'on va essayer d'expliquer par ses diverses caractéristiques en début de séjour. L'unité analysée sera dans ce cas la durée de séjour.

Les méthodes d'analyse paramétrique des biographies vont se révéler ici très utiles. Il faut voir que ces méthodes constituent la généralisation au longitudinal des méthodes de régression utilisées pour les données transversales.

Ainsi, après avoir vérifié qu'un modèle de Gompertz s'ajustait bien pour décrire la probabilité de migrer au cours du temps et que les caractéristiques mesurées jouaient de façon multiplicative sur le quotient instantané de migration, nous avons pu écrire un modèle de migration sous la forme suivante :

$$h(t; Z) = \exp(Z\beta + \rho t)$$

où $h(t)$ est le quotient instantané de migration à la durée de séjour t , Z étant un vecteur ligne correspondant aux diverses caractéristiques en début de séjour (état matrimonial, état professionnel, statut d'occupation du logement, etc.), β un vecteur colonne à estimer mesurant l'effet de ces caractéristiques sur la durée de séjour, ρ un paramètre également à estimer (D. Courgeau, 1985, a et b). L'utilisation de la méthode du maximum de vraisemblance permet d'estimer les paramètres β et ρ , ainsi que leurs variances et covariances.

L'avantage de cette méthode est de pouvoir faire intervenir un très grand nombre de caractéristiques pour mesurer leur effet sur

la durée de séjour. Lors de l'analyse de l'enquête triple biographie, nous avons pu faire intervenir 35 caractéristiques différentes. Lorsqu'une caractéristique est mesurée par une variable binaire (par exemple : 1 = marié, 0 = non marié), on voit que le paramètre, $\exp \beta$, mesure directement l'effet multiplicatif de cette caractéristique sur la probabilité de migrer. Ainsi, par exemple, pour les hommes nés en 1921-1925 le fait d'être marié correspond à un paramètre β égal à -0,452, ce qui indique une diminution des taux de changements de logement de plus du tiers (D. Courgeau, 1985, a,b), une fois les individus mariés.

Cette méthode n'est cependant pas sans poser certains problèmes. En premier lieu, il faut trouver la distribution s'adaptant le mieux aux durées de séjour observées, en donnant la préférence à celles dont le quotient et la fonction de séjour ont des formes explicites simples. Cela permet en effet de faciliter l'estimation du maximum de vraisemblance, en particulier, lorsqu'il y a des observations tronquées. En second lieu, il est nécessaire de vérifier, sur chacune des caractéristiques prises séparément, que le modèle à risques proportionnels (utilisé ici) s'applique correctement pour mesurer leur effet. Si tel n'est pas le cas, il est utile de désagréger la population (cela a été nécessaire pour le sexe de l'enquêté), en sous-population pour lesquelles on fait intervenir les caractéristiques ayant un effet proportionnel. Une autre possibilité est d'utiliser d'autres types de modèles s'adaptant mieux aux observations (modèle linéaire, modèle à sorties accélérées, etc.). Notons cependant ici que l'utilisation d'un modèle linéaire pour le quotient instantané peut conduire à des estimations négatives, ce qui est ici un grave inconvénient car ce quotient est par définition positif.

Un dernier inconvénient important de ces modèles est lié à l'effet des caractéristiques non observées par l'enquêté sur les paramètres estimés. En effet, une telle enquête ne peut évidemment pas saisir toutes les caractéristiques ayant une influence sur la durée de séjour étudiée. On peut penser

que d'autres caractéristiques, plus difficiles à observer et à saisir dans une enquête, ou même des caractéristiques dont le chercheur ne pense pas qu'elles puissent influencer sur l'événement, ont un effet non négligeable sur les quotients. Cette hétérogénéité non observée risque d'affecter les paramètres que nous avons estimés. Bien que l'on n'ait aucune idée de l'effet de ces caractéristiques non observées, on peut introduire une distribution suffisamment souple de cet effet, dont on mesure le rôle apparent sur les paramètres estimés correspondant aux caractéristiques observées. Sous ces conditions, il apparaît que selon la distribution supposée de cette hétérogénéité non observée (Heckman et Singer, 1984) ou même selon la distribution paramétrique prise pour estimer l'effet des caractéristiques observées (Trussel et Richard, 1987), les paramètres estimés peuvent varier énormément et même être de signe contraire. Ce dernier cas, lorsqu'il est vérifié, invalide toute estimation paramétrique de l'effet de ces variables.

2°) Analyse non paramétrique et semi-paramétrique de migrations peu répétitives.

Si le changement de logement est une migration très répétitive, on peut s'intéresser à des migrations plus rares dans la vie des enquêtés, en ce qu'elles correspondent à un changement très important du milieu d'existence: migration internationale, migration vers ou hors des métropoles, etc.

Dans ce cas, on peut considérer uniquement la première occurrence de cette migration, sachant qu'une seconde occurrence a peu de chances de se produire (3). En revanche, on peut chercher à analyser les interactions entre cette migration et divers événements du cycle de vie familial, professionnel, etc.

C'est ainsi que dans «Constitution de la famille et urbanisation» (D. Courgeau,

1987), nous avons pu étudier les liens entre première émigration d'une zone non-métropolitaine (ou d'une zone métropolitaine) et les divers événements familiaux : mariage, naissance du premier enfant, etc. Pour ce faire, nous avons utilisé d'abord, des méthodes non paramétriques, ensuite, des méthodes semi-paramétriques.

Les méthodes non paramétriques utilisées constituent une généralisation de l'analyse démographique longitudinale, à des interactions entre phénomènes de divers types. Le principe de l'analyse est résumé dans le schéma 1 qui présente l'interaction entre mariage et migration vers les métropoles.

Ainsi par exemple, à l'instant t , $N_r^c(t)$ individus vivant toujours hors métropole et encore célibataires, vont avoir un quotient d'émigration égal à $h_{rv}^c(t)$ et un quotient de nuptialité égal à $h_r^{cm}(t)$. Au même instant $N_{rv}^c(t)$ individus ayant migré antérieurement vers les métropoles comme célibataires vont avoir un quotient de nuptialité égal à $h_{rv}^{cm}(t)$, tandis que $N_r^m(t)$ individus mariés antérieurement hors métropole vont avoir un quotient d'émigration égal à $h_{rv}^m(t)$.

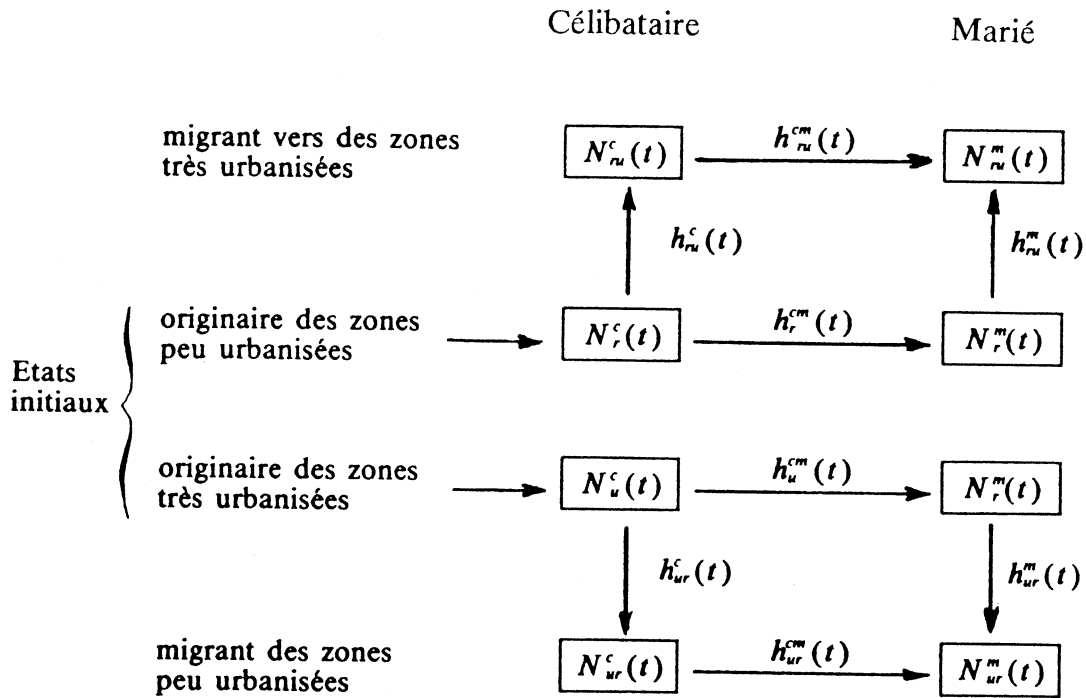
Il est possible d'estimer tous ces divers quotients en les supposant constants tout au long d'une année. On pourra alors comparer, par exemple, le quotient de nuptialité des individus toujours restés dans les zones peu urbanisées, $h_r^{cm}(t)$, avec celui des individus ayant effectué une migration vers les zones métropolitaines $h_{rv}^{cm}(t)$. Les tests pour ce faire ont été présentés dans un article antérieur (D. Courgeau et E. Lelièvre, 1986).

Pour poursuivre cette analyse, il est utile de voir l'effet de diverses caractéristiques des individus sur ces divers quotients. Les méthodes d'analyse semi-paramétrique permettent cette introduction. Plaçons-nous toujours dans le cas du mariage de l'individu selon qu'il a migré ou non vers les métropoles.

Pour cela, on suppose qu'il existe toujours

(3) On peut de toute façon étudier chacune des occurrences séparément, en utilisant les mêmes méthodes.

Schéma 1. - Divers états dans lesquels un individu peut se trouver et quotients instantanés de passage d'un état à un autre.



Source : D. Courgeau, 1987.

un quotient non-paramétrique sous-jacent s'appliquant aux migrants ou aux non migrants et que la migration le multiplie par une constante égale à : $\exp \beta$. L'effet des autres variables Z (intervenant avant le phénomène perturbateur) et Z' (intervenant après le phénomène perturbateur et pouvant donc être les mêmes que les précédentes, mais également différentes de celles-ci), peut s'exprimer dans le modèle suivant.

$$h^{cm}(t; z) = h_0^{cm}(t) \exp [Z \beta_1 + H(t-u) (\beta_0 + Z' \beta_2)] \quad (2)$$

où $h_0^{cm}(t)$ est le quotient sous-jacent, β_1 et β_2 sont les paramètres à estimer, $H(x)$ est la fonction de Heaviside (nulle si $x < 0$, égale à l'unité si $x > 0$), u la date de migration vers les métropoles. Ce modèle suppose encore que les diverses variables ont un effet multiplicatif sur le quotient estimé.

L'utilisation de ces modèles évite de choisir une forme paramétrique pour le quotient sous-jacent, ce qui est un gros avantage. Il a également été possible d'étudier de façon théorique comment l'omission de certaines caractéristiques indépendantes de celles observées, affectait les paramètres estimés (Bretagnolle et Huber-Carol, 1988). On a ainsi pu démontrer que cette omission n'affectait pas le signe des paramètres estimés. Cependant, elle entraîne une réduction de la valeur absolue de ces paramètres. De ce fait certaines caractéristiques qui semblaient n'avoir aucun effet significatif, peuvent devenir tout à fait significatives lorsqu'on introduit ces caractéristiques initialement non-observées.

Ces résultats nous amènent à privilégier l'approche semi-paramétrique qui permet d'être sûr du sens des effets observés.

CONCLUSION

Nous avons essayé de montrer ici certains apports des enquêtes rétrospectives lors-

que l'on désire analyser la dimension spatiale du cycle de vie.

L'approche paramétrique permet d'analyser toutes les migrations effectuées par une population tout au long de son existence, en fonction d'un très grand nombre de caractéristiques.

L'approche non paramétrique et semi-paramétrique permet d'analyser des migrations plus particulières et moins répétitives tout en faisant intervenir d'autres événements du cycle de vie en interaction avec elles. Une telle analyse permet en particulier, de mettre en évidence des types de dépendances nouveaux entre phénomènes démographiques. Ainsi, par exemple, la dépendance unilatérale indique que le fait d'avoir connu l'un des événements modifie la probabilité de connaître le second, mais qu'inversement, le fait d'avoir connu le second ne joue en rien sur la probabilité de connaître le premier. Une telle dépendance est apparue souvent lors des analyses de l'enquête «Triple biographie».

Il nous faut pour terminer dire dans quel sens poursuivre cette analyse. Nous avons

indiqué dans l'introduction un certain nombre d'hypothèses qui permettaient de justifier notre analyse.

Il nous en reste encore une très importante à poser pour la recherche à venir. En effet, nous avons indiqué que les probabilités de connaître les divers types d'événements existaient à un instant donné pour tout individu de la population sur laquelle on travaille.

Il importe pour l'avenir d'essayer de mettre en évidence les modifications au cours du temps de ces probabilités et d'émettre des hypothèses sur les raisons de ces modifications. Si l'analyse d'enquêtes biographiques rétrospectives permettra de mesurer la modification au cours du temps des probabilités des divers événements, en couvrant un très grand nombre de générations, d'autres types d'enquêtes seront, à mon avis, nécessaires pour mettre en évidence les raisons de ces modifications. Nous espérons cependant que l'avenir nous permettra de répondre à cette dernière question.

BIBLIOGRAPHIE

BRETAGNOLLE J. et HUBER-CAROL C. (1983), Effects of omitting covariates in Cox's model for Survival date, *Scand Journal of Statistics*, n° 15, pp. 125-138.

COURGEAU D. (1985a), Interaction between spatial mobility, family and career life-cycle : a French survey, *European Sociological Review*, vol.1, n° 2, pp. 139-162.

COURGEAU D. (1985b), Changements de logement, changements de département et cycle de vie, *L'Espace Géographique*, n° 4, pp. 289-306.

COURGEAU D. (1987), Constitution de la famille et urbanisation, *Population*, n° 1, pp. 57-82.

COURGEAU D. (1988), *Méthodes de mesure de la mobilité spatiale : migrations internes, mobilité temporaire, navettes*, Edition de l'INED, 301 pp.

COURGEAU D. (1990), Analyse des données biographiques erronées, *Population*, à paraître

COURGEAU D. et LELIÈVRE E. (1986), Nuptialité et agriculture, *Population*, n° 2, pp. 303-326.

COURGEAU D. et LELIÈVRE E. (1989), *Analyse démographique des biographies*, Edition de l'INED, 268 pp.

HECKMAN J. et SINGER B. (1984), A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica*, vol. 52, n° 2, pp. 271-320.

HENRY L. (1972), *Démographie : analyse et modèles*, Edition de l'INED (1984), 314 pp.

KALBFLEISH J. et PRENTICE R. (1980), *The statistical analysis of failure time data*, Wiley, New York, 321 pp.

KIMBALL T. et PEARSALL M. (1954), The nature of human groups, *The Talladega Story*, Un. of Alabama Press, pp. XIII-XXXII.

POULAIN M., RIANDEY B. et FIRDION J.M. (1990), Une expérimentation Franco-Belge sur la fiabilité des enquêtes rétrospectives : l'enquête 3B BIS, *Population*, à paraître.

TRUSSEL J. et RICHARD T. (1985), Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure, in *Sociological Methodology*, N. TUMA ed., Jossey Bass, pp. 242-276.