

CONSTRUCTION, DÉVELOPPEMENT ET ENTRETIEN DU PANEL D'ENTREPRISES DU KOF/ETHZ

APERÇU DU PROJET

OBJECTIFS ET TRAVAUX EFFECTUÉS

Le projet « panel » avait pour objectifs: (1) construction d'un panel d'entreprises représentatif de l'économie suisse; (2) d'ajuster le panel régulièrement aux changements de la population (variable de référence étant le nombre d'employés équivalent plein temps du recensement fédéral des entreprises), et de l'entretenir suite aux mutations de toutes sortes; (3) de mener des enquêtes auprès des entreprises du panel; (4) de permettre une gestion optimale des données collectées et des variables supplémentaires (adresses, fusions, faillites, variables structurelles, facteurs de pondération, etc.); (5) d'évaluer et analyser les données le plus rigoureusement possible. On aperçoit aisément que le projet est à la fois théorique et pratique, le résultat final étant une base de données transversales et chronologiques, un lot de méthodes et de procédures statistiques, et des études relatives aux différentes enquêtes menées.

Dans une phase I du projet (*Donzé/Hollenstein 2000*), nous avons constitué un panel d'environ 7'200 entreprises et l'avons utilisé avec succès dans la conduite de trois enquêtes. Quelques outils statistiques et informatiques, notamment pour la constitution du panel (stratification optimale, échantillonnage des entreprises, conception de la base de données des adresses, etc.) et pour l'évaluation des enquêtes (traitement de la non-réponse partielle et globale, correction du biais de sélection, problème de l'inférence pour une enquête complexe, etc.) ont été développés.

Dans la phase II du projet (*Donzé/Hollenstein 2003*), nous avons poursuivi les travaux effectués. En premier lieu, ceux-ci consistaient à (1) l'entretien du panel suite aux mutations de toutes sortes (adresses, faillites, etc.); (2) l'ajustement du panel selon les résultats du recensement fédéral des entreprises de 1998; (3) la conduite de trois autres enquêtes; (4) l'approfondissement des investigations méthodologiques et statistiques, en particulier pour la correction de la non-réponse et ses conséquences (inférence, méthodes d'estimation de différents modèles, etc.); (5) l'organisation des données des différentes enquêtes en une base de données transversales et chronologiques afin de produire rapidement et aisément des analyses de type « panel » lors de nos enquêtes.

RÉSULTATS ET PERSPECTIVES

Au terme du projet « panel », le KOF ETH Zurich est en mesure d'effectuer des enquêtes complexes avec un grand nombre de participants ainsi que de les évaluer. Les méthodes statistiques développées dans ce projet permettent une extrapolation des résultats de l'échantillon à l'ensemble de la population. L'utilisation de la base de données transversales et chronologiques a créé des conditions uniques en Suisse pour réaliser des analyses empiriques au niveau des entreprises du type « panel ». Un rattrapage du retard par rapport à l'étranger a eu lieu dans ce domaine de recherche actuellement de pointe. De ce point de vue, cela signifie un sérieux renforcement de l'infrastructure de recherche helvétique – ce qui était l'un des objectifs du PP « Demain la Suisse ».

Le KOF ETH Zurich investira de nouveaux moyens substantiels dans le développement du panel et garantira ainsi la permanence de cet instrument pour la recherche empirique. Au premier plan se trouvent (1) l'entretien permanent du panel et son ajustement; (2) réalisation d'autres enquêtes; (3) l'intégration successive des réponses obtenues dans la base de données transversales et chronologiques; (4) l'amélioration continue des méthodes et procédures statistiques; (5) l'utilisation du panel pour des projets de recherches du KOF ETH Zurich. En outre la base de données transversales et chronologiques sera mise à la disposition de chercheurs externes – tout en garantissant la protection des données.

LES ÉLÉMENTS PRINCIPAUX DU PROJET

CONSTRUCTION DU PANEL EN 1996

Le recensement fédéral des entreprises de 1995 servant de population de base, le panel est constitué de trois échantillons d'entreprises, des secteurs de l'industrie, de la construction et des services (*Donzé 1998*). Chaque échantillon est stratifié selon les branches économiques et les tailles des entreprises. Pour chaque branche économique, les classes des trois tailles des entreprises (petite, moyenne et grande) ont été calculées de manière optimale, avec comme variable de référence le nombre d'employés équivalent plein temps. Étant donné un nombre fixe d'entreprises par secteur d'activité – 3000, 600 et 3000 entreprises respectivement pour les secteurs de l'industrie, de la construction et des services – et en fixant que toutes les grandes entreprises figurent dans l'échantillon, un nombre optimal d'entreprises par strate a été calculé. Enfin, le panel est constitué en tirant aléatoirement sans remise dans le registre des entreprises des recensements, pour chaque strate, les entreprises répondant à nos critères.

L'AJUSTEMENT DU PANEL EN 1999

L'ajustement du panel s'est effectué d'abord en désactivant un certain nombre d'entreprises qui, par exemple, ont soit eu certaines mutations (fusion, faillite, etc.) ou qui n'ont jamais répondu à nos enquêtes. Ensuite, selon les données du recensement fédéral des entreprises de 1998, on a recalculé, selon la démarche utilisée pour la construction du panel, la grandeur optimale des strates et du nombre d'entreprises par strate. On a placé ensuite dans la nouvelle structure, les entreprises actives du panel. On a complété le panel en tirant aléatoirement les entreprises manquantes du fichier des entreprises du recensement 1998.

L'ÉTAT ACTUEL DU PANEL

Grâce aux informations que nous collectons, notamment par la réalisation d'enquêtes, le fichier d'adresses du panel est mis à jour régulièrement. Du aux fermetures, aux fusions etc. l'échantillon s'est réduit depuis l'ajustement de 1999. Le tableau 1 résume l'état actuel du panel.

En ce moment, le panel est mis à jour d'après la procédure décrite en utilisant les données du recensement fédéral des entreprises 2001. L'ajustement du panel sera achevé début 2004.

Tableau 1: Nombre d'entreprises du panel par secteur économique et selon la taille

	Petites			Moyennes			Grandes			Total		
	N	I-%	C-%	N	I-%	C-%	N	I-%	C-%	N	I-%	C-%
Industrie	1579	52	47	1153	38	48	307	10	43	3039	100	47
Construction	241	39	7	226	37	9	144	24	20	611	100	9
Services	1573	55	46	1055	36	43	267	9	37	2895	100	44
Total	3393	52	100	2434	37	100	718	11	100	6545	100	100

Notes: N est le nombre d'entreprises; I-% et C-% sont respectivement les pourcentages par rapport aux lignes et colonnes.

LES ENQUÊTES

Le panel est en soi un échantillon d'entreprises suisses avec lequel différents types d'enquêtes sont menés. Depuis sa conception, les « panélistes » ont été interrogés plusieurs fois. Nous procédons par envoi postal, puis nous faisons une à deux actions de relance téléphonique. Les enquêtes qui ont été effectuées, touchant des thèmes variables, sont les suivantes (cf. www.kof.ethz.ch pour les questionnaires):

- Les qualifications professionnelles de l'avenir, 1998;
- L'internationalisation de l'économie suisse, 1998;
- Les activités d'innovation, 1999;
- Changement organisationnel et introduction de technologies d'information et de communication, 2000;
- Les activités d'innovation, 2002;
- L'utilisation des technologies de l'information et de la communication au sein des entreprises, 2002;
- Les enquêtes annuelles sur les investissements, 1998, 1999, 2000, 2001, 2002.

En outre, on peut y ajouter les enquêtes des années 1990, 1993 et 1996 sur les activités d'innovation bien que le panel n'existait pas encore en 1996; les entreprises interrogées à l'époque paraissent aujourd'hui dans le panel et sont intégrées dans la base de données transversales et chronologiques (cf. ci-dessous).

Les enquêtes mentionnées ont fait l'objet de nombreuses publications. Elles ont non seulement livré des informations descriptives sur les entreprises suisses et l'ensemble de l'économie, mais aussi une base de données facilement exploitable pour des analyses économétriques (estimation de modèles) et méthodologiques (cf. www.kof.ethz.ch pour une liste des publications basées sur l'analyse des données du panel).

PRISE EN COMPTE DU PLAN DE SONDAGE ET LA CORRECTION DE LA NON-RÉPONSE

Sur la base de la littérature statistique, nous avons peu à peu développé une méthodologie nous permettant de traiter nos données rigoureusement. En particulier, nous nous sommes attachés à tenir compte de la non-réponse, étant un aspect important dans l'évaluation d'une enquête complexe, et d'en corriger les effets le mieux possible.

La non-réponse globale (« item non-response ») demeure importante quelle que soit l'enquête, car les enquêtes mentionnées ci-dessus ont connu un taux de réponse qui varie entre 30% et 56%, ce qui est par ailleurs pour ce genre d'enquêtes tout à fait satisfaisant. Ce taux de retour dépend de plusieurs facteurs, notamment de la complexité de l'enquête et du nombre d'actions de relance téléphonique. Celui-ci a d'ailleurs été sensiblement augmenté depuis la première enquête. La correction de la non-réponse globale et d'un biais de sélection éventuel découlant de cette dernière, s'effectue à l'aide d'un facteur de pondération, d'une part via la probabilité de réponse estimée et, d'autre part, par l'échantillonnage du facteur de pondération selon les variables clefs de l'enquête (*Donzé 2001a et 2002*). Notons que le facteur de pondération final est un produit de plusieurs composantes, lesquelles tiennent compte du plan de sondage, de la probabilité de réponse, d'un ajustement selon le nombre d'employés ou le chiffre d'affaires par strate.

L'évaluation du biais de sélection est obtenue grâce aux résultats d'une enquête supplémentaire auprès d'un échantillon des non-répondants de l'enquête initiale. Cette seconde enquête s'effectue par téléphone et s'intéresse à des variables clefs de l'enquête. Les données de cette seconde enquête nous permettent d'estimer s'il y a ou non une différence significative pour chaque variable clef de l'enquête entre les répondants et les non-répondants. Si cette différence existe, on se servira des totaux estimés par branche et taille d'entreprise pour caler (calage sur marges), ou étalonner, notre facteur de pondération.

La non-réponse partielle (« item non-response ») est quant à elle très variable d'une question à l'autre – normalement de 5% à 20% mais quelques fois plus de 50%. Ce type de non-réponse est corrigé par un système d'imputations multiples (*Donzé 2001b, c et 2003*).

Notons finalement que l'emploi d'un facteur de pondération ou la méthode de l'imputation multiple nécessite, notamment dans le cas des statistiques descriptives, d'adapter les estimateurs de la variance des coefficients estimés.

BASE DE DONNÉES TRANSVERSALES ET CHRONOLOGIQUES

Le but ultime de ce projet panel était de constituer avec les informations et les données collectées lors des différentes enquêtes une base de données transversales et chronologiques qui soit aisément exploitable pour des analyses économétriques. Il s'agit bien évidemment d'un acquis indéniable car une telle base n'existe pas à ce jour en Suisse. De plus, les analyses de données de panel sont un champ d'investigations théoriques et empiriques très prometteur.

Les données de chaque enquête ont subi un intense travail de vérification (cohérence logique, plausibilité des valeurs, etc.) et de corrections éventuelles. Les variables structurelles des entreprises ont été ajoutées ainsi que les variables définissant le plan de sondage (variables de stratification, caractéristiques des strates, etc.). D'autres variables ont été générées, par exemple les facteurs de pondération. Nous avons tâché au fur et à mesure des enquêtes de définir ces nouvelles variables de manière identique et d'adopter une notation uniforme. Enfin, nous avons systématiquement imputé les valeurs manquantes.

L'exploitation de la dimension chronologique de ces données nécessitait, d'une part, l'harmonisation des variables au contenu presque similaire; malheureusement, certaines variables ont dû être abandonnées n'étant pas l'objet d'enquêtes sur toutes les périodes. D'autre part, il était indispensable de vérifier la série d'entreprises du panel, afin d'éviter des discontinuités, en tenant compte des modifications structurelles (fusions, scissions, etc.); de cette manière nous avons garanti que les données collectées soient attribuées aux bonnes unités (groupe industriel, partie d'une entreprise etc.).

La base de données transversales et chronologiques contient les informations des enquêtes sur les activités d'innovation des années 1990, 1993, 1996, 1999, 2002; les données d'autres enquêtes seront ajoutées successivement au cours de l'année 2004. À ce jour, seule une étude d'un modèle d'innovation (estimation du type « panel ») sur la base des données 1993, 1996 et 1999 a envisagé la dimension temporelle (*Arvanitis et al. 2001*). Cette analyse a notamment montré qu'un « unbalanced panel » est indispensable pour faire des estimations du type « panel », la non-réponse étant considérable.

RÉFÉRENCES

- Arvanitis, S., Bezzola, M., Donzé, L., Hollenstein, H. und Marmet, D. (2001): *Innovationsaktivitäten in der Schweizer Wirtschaft. Eine Analyse der Ergebnisse der Innovationserhebung 1999*, Studienreihe Strukturbericht-erstattung Nr. 5, hrsg. vom Staatssekretariat für Wirtschaft (Seco), Bern.
- Donzé, L. (1998): Développement et entretien du «Panel d'entreprises» du KOF/ETHZ. Une étude méthodologique. *Forschungsbericht*, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, septembre. (www.mysunrise.ch/users/ldonze)
- Donzé, L. (2001a): Echantillon et non-réponse. Note sur la construction des facteurs de pondération de l'enquête 1999 sur l'innovation, *Forschungsbericht*, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Avril.
- Donzé, L. (2001b): L'imputation des données manquantes, la technique de l'imputation multiple, les conséquences sur l'analyse des données: l'enquête 1999 KOF/ETHZ sur l'innovation, *Schweiz. Zeitschrift für Volkswirtschaft und Statistik*, 137(3), 301–317.
- Donzé, L. (2001c): Multiple Imputation and Modeling: Some Experiences from the KOF/ETHZ's Innovation Survey 1999, 53e session de l'Institut international de statistique (IIS), Séoul, Institut international de statistique (IIS), Technique d'imputation, *Bulletin de l'Institut International de Statistique*, Communications libres, Livraison 2, LIX (2/3), 427–428.
- Donzé, L. (2002): Methodology to correct the unit non-response bias in the case of the KOF ETH Zurich's survey 2000 on Organisation and Information Technologies, Paper prepared to «The International Conference on Improving Surveys 2002 (ICIS)», Copenhagen, August, 25–28, 2002, published as *KOF Working Paper*, No. 79, Zurich, August. (www.kof.ethz.ch).
- Donzé, L. (2003): Multiple imputation to correct item non-response for the KOF ETH Zurich's business surveys. A sensitivity analysis of different imputation procedures, Paper presented at the «Basle Workshop on Item-nonresponse and Data Quality in Large Social Surveys», Basle, Octobre, 9–12.
- Donzé, L. und Hollenstein, H. (2000): *Aufbau, Weiterentwicklung und Pflege des KOF/ETH-Unternehmenspanels. Hauptergebnisse und Schlussfolgerungen für die weitere Verbesserung des Panels, Schlussbericht z.Hd. des Schweizerischen Nationalfonds. Schwerpunktprogramm «Zukunft der Schweiz» (Phase I)*, Projekt 5004–47794, Konjunkturforschungsstelle der ETH Zürich (KOF), Zürich, Februar. (www.kof.ethz.ch)
- Donzé, L. et Hollenstein, H. (2003): *Construction, développement et entretien du panel d'entreprises du KOF/ETH, Rapport final à l'attention du Fonds national suisse de la recherche scientifique. Programme prioritaire «Demain la Suisse» (Phase II)*, Projet 5004–58445, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Novembre. (www.kof.ethz.ch)