

Construction, développement et entretien du panel d'entreprises du KOF/ETHZ

Introduction générale

Le projet «panel» a pour objectifs (1) de mettre sur pied un panel d'entreprises représentatif de l'économie suisse ; (2) de mener des enquêtes auprès de ces entreprises ; (3) de permettre une gestion optimale des données collectées ; (4) d'évaluer et analyser ces données le plus rigoureusement possible. On perçoit aisément que le projet est à la fois théorique et pratique, le résultat final étant une base de données (adresses et données collectées), un lot de méthodes et de procédures statistiques, et des études relatives aux différentes enquêtes menées.

Dans une phase I du projet, nous avons constitué le panel d'entreprises et l'avons utilisé avec succès dans la conduite de plusieurs enquêtes. Quelques instruments statistiques et informatiques, notamment dans la constitution du panel (stratification optimale, échantillonnage des entreprises, conception de la base de données des adresses, etc.) et dans 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.).

La phase II du projet poursuit les travaux effectués, notamment en mettant à jour le panel, en développant quelques outils statistiques et informatiques supplémentaires d'analyse et d'estimation, en approfondissant certaines notions théoriques, mais surtout en organisant les données des différentes enquêtes en une base de données transversales et chronologiques de manière à produire rapidement et aisément des analyses de type «panel». Au terme du projet «panel», nous disposons d'un instrument unique, en Suisse en tout cas, permettant de mener des investigations dans un domaine de recherche actuellement de pointe.

Les étapes de la construction du panel

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. 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

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

L'état du panel

Grâce aux informations que nous collectons, notamment par la conduite d'enquêtes, le fichier d'adresses du panel peut être mis à jour régulièrement. Le tableau N°1 résume l'état actuel de notre panel.

Tableau N°1 :

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

	Petite			Moyenne			Grosse			Total		
	N	l-%	c-%	N	l-%	c-%	N	l-%	c-%	N	l-%	c-%
Industrie	1649	52	46	1186	38	48	320	10	43	3155	100	46
Construction	254	40	7	233	37	9	147	23	20	634	100	9
Services	1672	56	47	1076	36	43	271	9	37	3019	100	45
Total	3575	52	100	2495	37	100	738	11	100	6808	100	100

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

Les enquêtes

Le panel constitue un outil exceptionnel pour mener de façon optimale différents types d'enquêtes auprès des entreprises suisses. Depuis sa conception, nous l'avons utilisé avec succès dans cette tâche. Nous procédons par envoi postal puis nous faisons une à deux actions de relance téléphonique. Les enquêtes qui ont été effectuées sont les suivantes :

- ◆ 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 enquêtes sur les investissements, 1998, 1999, 2000, 2001

On peut en outre ajouter à cette masse énorme de données, les enquêtes 1993 et 1996 sur les activités d'innovation bien que le panel n'était en 1996 pas encore mis sur pied car les entreprises interrogées se retrouvent néanmoins dans le panel.

Toutes ces enquêtes ont fait l'objet de nombreuses publications. Elles ont non seulement livré d'intéressantes et importantes informations sur les entreprises suisses et l'économie suisse, mais également ont été le support à des investigations théoriques pointues.

Références

- Donzé, L. (1998): «Développement et entretien du «Panel d'entreprises» du KOF/ETHZ. Une étude méthodologique. Programme prioritaire «Demain la Suisse»», Projet 5004-47794, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Septembre 1998, 18.
- Donzé, L. (1999): «Enquête auprès des non-répondants de l'enquête KOF/ETHZ sur l'internationalisation de l'économie suisse. Une étude méthodologique.» Rapport intermédiaire à l'attention du Fonds national suisse de la recherche scientifique. Programme prioritaire «Demain la Suisse», Projet 5004-47794, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Janvier 1999, 8.
- Donzé, L. (2001a): «Echantillon et non-réponse. Note sur la construction des facteurs de pondération de l'enquête 1999 sur l'innovation», Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Avril 2001, 4.
- Donzé, L. (2001b): «Multiple Imputation and Modeling : Some Experiences from the KOF/ETHZ's Innovation Survey 1999», 53e session de l'Institut international de statistique, 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. (2001c): «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. et Hollenstein, H. (2000): «Aufbau, Weiterentwicklung und Pflege des KOF/ETH-Unternehmenspanels. Hauptergebnisse und Schlussfolgerungen für die weitere Verbesserung des Panels», Schlussbericht 2.Hd. des Schweizerischen Nationalfonds zum «Modul Arbeitswelt» des Schwerpunktprogramms «Zukunft der Schweiz» (Phase I), Projekt 5004-47794, Konjunkturforschungsstelle der ETH Zürich (KOF), Zurich, Februar 2000, 12.

La 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 corriger tant la non-réponse partielle que globale. En effet, celles-ci demeurent importantes quelle que soit l'enquête. Par exemple, l'enquête sur les «changements organisationnels et les technologies d'information et de communication», qui fut adressée à quelque 6'845 entreprises actives du panel des secteurs de l'industrie, de la construction et des services, connut un taux de réponse de 39,38%, qui est par ailleurs pour ce genre d'enquête relativement bon.

La non-réponse, qu'elle soit partielle ou totale, étant un aspect important dans l'évaluation d'une enquête complexe, nous nous devons d'en tenir compte et d'en corriger les effets le mieux possible. Pour ce faire, nous avons adopté la méthodologie développée et utilisée dans la phase I du projet.

La correction de la non-réponse globale («unit nonresponse») 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. 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 (caler sur marges), ou étalonner, notre facteur de pondération.

La non-réponse partielle («item nonresponse») est quant à elle très variable d'une question à l'autre (de 5-6% à plus de 50%). Ce type de non-réponse est corrigé par un système d'imputations multiples.

Notons finalement que l'emploi d'un facteur de pondération ou la méthode de l'imputation multiple nécessitent, 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 est 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'agira 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.

Nous disposons actuellement des données de plusieurs enquêtes. Celles-ci 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 ou d'autres variables relatives à l'une ou l'autre des questions de l'enquête. 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 de la plupart des variables.

L'exploitation de la dimension chronologique ou temporelle de ces données nécessite une étape supplémentaire, et non des moindres: l'harmonisation des variables au contenu presque similaire et l'organisation des données. Des hypothèses doivent être faites qui peuvent ne pas s'avérer adéquates par la suite. En outre, des variables doivent malheureusement être abandonnées pour ne pas avoir été l'objet d'enquêtes sur toutes les périodes.

A ce jour, seule une étude pilote d'un modèle d'innovation sur la base des données des enquêtes KOF 1993, 1996 et 1999 sur l'innovation a envisagé cette nouvelle dimension du panel.