

Un recensement 2010 basé sur les registres: le projet du Conseil fédéral était attendu depuis longtemps

Le Conseil des États a approuvé le 22 mars 2007 la loi sur le recensement fédéral de la population. Il s'est ainsi rangé à l'avis du Conseil fédéral, contre celui des cantons. Le recensement fédéral de la population (RFP) de 2010 devrait mettre en place de nombreux changements espérés depuis longtemps. Il se basera sur les registres des administrations et non plus sur un questionnaire distribué à l'ensemble des ménages. Il s'agit là d'une décision à la fois courageuse et sensée du point de vue technique. Le Conseil national se prononcera en juin sur la question.



Un des problèmes des recensements exhaustifs est que l'on ne peut pas estimer exactement l'importance des erreurs, au contraire des enquêtes par échantillonnage. Lors du RFP 1990, un questionnaire sur cinq a dû être complété. En 2000, c'était encore pire.

Photo: Keystone

La variante du Conseil fédéral actuellement débattue est une solution de compromis. À l'origine, l'Office fédéral de la statistique (OFS) prévoyait d'effectuer le recensement uniquement à partir des registres. Au fil des discussions, parfois houleuses, avec la Conférence suisse des offices régionaux de statistique (Corstat), il a proposé de compléter le recensement tel que d'abord prévu par une enquête auprès d'un grand échantillon de la population, pour collecter les données qui ne figurent pas dans ces registres. Le Conseil fédéral accède ainsi à la demande des services cantonaux de statistique, qui souhaitaient que l'on ne fasse pas l'impasse sur ces informations qui avaient toujours été relevées auparavant.



Pr Hans Wolfgang Brächinger
Président de la Commission de la statistique fédérale, Département d'économie quantitative, université de Fribourg

Une idée préconçue des «vrais» chiffres

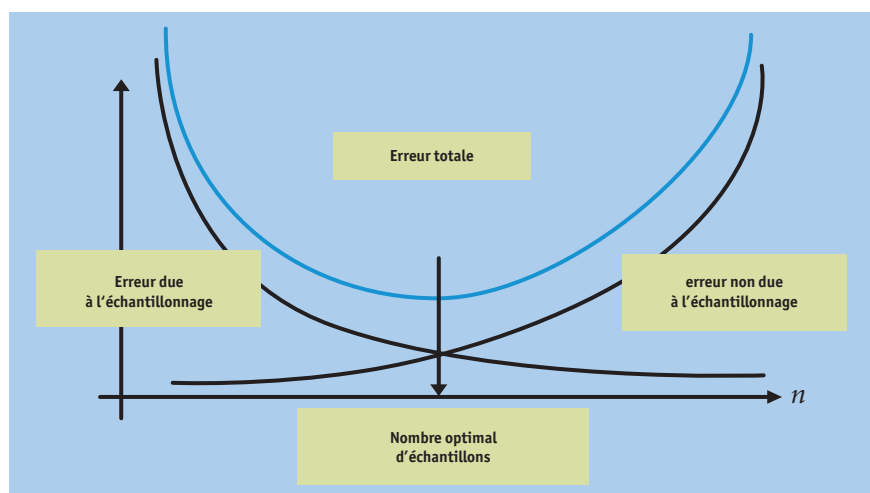
On peut aisément comprendre que la variante du Conseil fédéral soit contestée. Elle bat en brèche une conception du travail statistique vieille de plus d'un siècle: à toute grandeur correspond toujours une vraie valeur, qui peut être obtenue à l'aide d'un relevé exhaus-

tif. Il suffirait donc de compter suffisamment longtemps et avec suffisamment de précision! Le coût de l'opération étant déterminé par le prix de revient du questionnaire, il augmente linéairement avec le nombre des questionnaires.

Quelques exemples simples mettent en évidence la naïveté de cette conception. Prenons ainsi la population résidente de Zurich à la fin de l'année 2010, date de référence du prochain RFP. Outre le fait qu'il faut commencer par préciser qui devra être recensé, il est impossible, par l'expérience, de cerner directement cette valeur, qui peut paraître élémentaire à première vue. Personne n'est tout simplement en mesure de compter les habitants de Zurich et, par conséquent, de vérifier les chiffres de la statistique publique. Le dénombrement des derniers dix mille habitants, dont font partie tous ceux qui ne veulent pas être recensés, occasionne par ailleurs bien plus de travail que celui des dix premiers. Or, à peine les derniers habitants sont-ils enfin dénombrés que les premiers résultats du recensement ne sont déjà plus actuels.

Graphique 1

Nombre optimal d'échantillons représentatifs dans les relevés statistiques au regard de l'erreur totale



Source: Brachinger/La Vie économique

Des recensements non exempts d'erreurs

Les conditions pratiques dans lesquelles s'effectue un recensement traditionnel montrent qu'il faut cesser de croire qu'il s'agit d'une statistique « précise ». En général, les résultats d'un tel dénombrement sont tout sauf parfaits. On sait depuis longtemps que les données sont d'autant plus imprécises que le nombre d'individus à recenser est grand. Lors du RFP 1990, un questionnaire sur cinq a dû être « complété » ou « corrigé ». En 2000, les données brutes étaient encore plus lacunaires que celle de la précédente décennie. Quelque 30% des personnes de plus de 15 ans n'ont pas répondu à la question sur le travail familial et domestique, introduite cette année-là. En Suisse romande et en Suisse italienne, la part de non-réponse à cette question a même frôlé les 40% chez les personnes de nationalité étrangère. La proportion de la population qui a rencontré des difficultés à comprendre le questionnaire et à le remplir a augmenté par rapport à 1990. Il est impossible de quantifier l'ampleur de l'erreur statistique que cela entraîne.

Quant à l'idée que les coûts d'un recensement augmentent de manière linéaire, elle ne tient pas non plus. Lors des relevés précédents, le personnel des administrations communales, aidé d'auxiliaires engagés pour l'occasion, a dû fournir un travail important pour compléter les questionnaires lacunaires. Les coûts occasionnés n'ont jamais été clairement établis, sans parler de la qualité des ajouts. Il n'y a, dès lors, rien d'étonnant que l'opposition des associations de communes à un recensement traditionnel soit grande. Lors du RFP 2000, celles-ci ont tout simplement cessé de demander les informations manquantes une fois un certain plafond de coûts atteint.

Il n'y a pas de données parfaites

Les milieux de la statistique publique ont tendance à ne parler qu'à mots couverts de ces imperfections. La qualité parfois contestable des données publiques est devenue chose taboue au XIX^e siècle et au début du XX^e. Avec le développement des services statistiques, l'idée s'est répandue que les données relevées par une autorité n'avaient pas à être remises en question, et surtout pas par ladite autorité. Un service de statistique ne dénombrait, en outre, que ce qui existait; autrement dit, les résultats d'un recensement ne pouvaient être, par définition, que le reflet exact de la réalité. On pensait ne pouvoir corriger la mauvaise image de la statistique qu'en insistant sur l'exactitude des statistiques publiques.

Cette façon de voir explique aussi les réserves que certains services statistiques émettent à l'égard des estimations faites à partir d'échantillons ou des analyses fondées sur des modélisations. Il est, pourtant, évident que les données sont d'autant plus précises que le nombre d'unités à recenser est faible. Elles sont aussi disponibles plus rapidement. Bien évidemment, les résultats d'enquêtes menées auprès d'échantillons de la population ne sont pas parfaits: il existe une erreur d'échantillonnage due à la sélection au hasard des personnes interrogées. Cette erreur peut, toutefois, se calculer, à la différence de celle évoquée plus haut et qui n'est pas due à l'échantillonnage. Enfin, les coûts des enquêtes par échantillonnage progressent effectivement de manière linéaire.

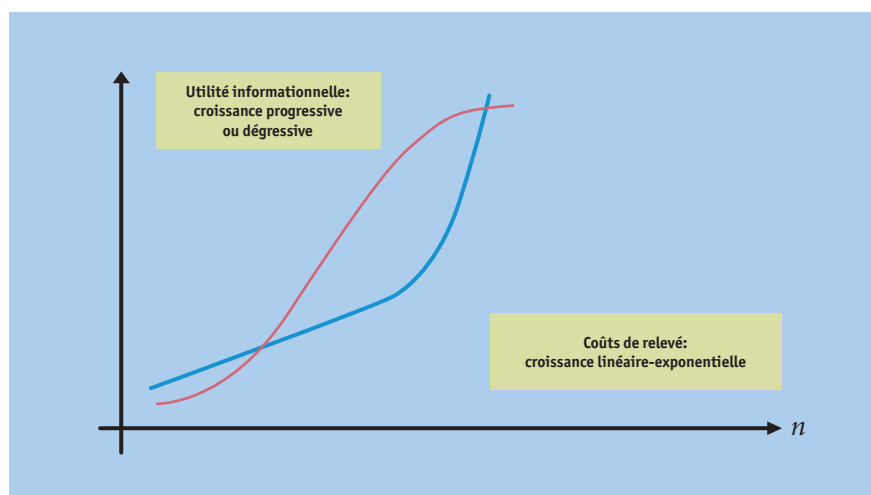
Une chose est sûre: il n'y a pas de données statistiques parfaites! Toute enquête est plus ou moins entachée d'erreurs. La qualité des résultats dépend de l'erreur totale, qui se compose elle-même des erreurs dues ou non à l'échantillonnage (voir *graphique 1*). La première se compose principalement de l'erreur aléatoire qui résulte de la sélection au hasard des personnes interrogées. L'erreur non due à l'échantillonnage comprend toutes celles qui proviennent de réponses inexactes et d'une saisie ou d'un traitement erroné. Les deux types d'erreurs évoluent en sens contraire: les premières augmentent avec la taille de l'échantillon, tandis que les secondes diminuent. Même si l'on ne saura jamais quelle est la taille optimale d'un échantillon n , une chose est claire: si l'on considère l'erreur totale, la dimension idéale d'une enquête par échantillonnage se situe bien en dessous de celle d'un relevé exhaustif.

Une erreur d'échantillonnage réduite

Si l'on doit choisir entre une enquête par échantillonnage, comme le Conseil fédéral le

Graphique 2

Évolution de l'utilité informationnelle et des coûts de relevé suivant la taille de l'échantillon



Source: Brachinger/La Vie économique

prévoit pour le relevé des variables ne figurant pas dans les registres, et une enquête exhaustive, comme le demandent les cantons, il faut s'interroger sur l'ampleur de l'erreur d'échantillonnage de la première, sachant qu'une enquête exhaustive implique de toute manière des inexactitudes de grande ampleur étrangères à l'échantillon choisi. Les simulations faites par l'OFS sur la base du dernier recensement montrent qu'il est possible de procéder à des estimations d'une grande précision à partir de 200 000 observations. Les enquêtes structurelles qu'il est prévu de réaliser chaque année auprès d'un tel échantillon fourniront des résultats représentatifs pour toutes les communes d'une certaine taille et même pour les quartiers des grandes villes.

Un échantillon de cette taille peut déjà permettre de s'exprimer avec suffisamment d'exactitude sur des éléments comme la langue, l'appartenance religieuse ou le statut sur le marché du travail des groupes de 15 000 personnes, soit environ 0,2% de la population suisse. La mise en «pool» des données collectées pendant plusieurs années d'affilée accroîtra encore la précision. Le groupage des données sur trois ans permet de ramener la taille des groupes à 5000 individus; à 3000 après cinq ans.

Une autre manière de mesurer le degré de précision consiste à se concentrer sur l'estimation et la découverte de petits groupes. Les études de l'OFS ont montré que, pour un échantillon net de 200 000 personnes, la taille minimale d'un groupe pouvant être mis en évidence sans trop d'erreur est de 140 personnes. Au bout de cinq ans, on disposera de données qui, mises en «pool», permettront même d'estimer des groupes de 30 personnes. On peut dès lors tabler sur le fait que l'erreur

totale découlant de la proposition du Conseil fédéral sera beaucoup plus faible que celle résultant de la variante des cantons.

Un autre élément, mis en avant par la Commission économique pour l'Europe des Nations Unies dans ses recommandations pour les recensements de 2010, plaide en faveur d'enquêtes par échantillonnage: c'est l'actualité des données, qui contribue largement à leur caractère informatif. L'intervalle qui sépare la collecte des données de leur publication devrait être aussi court que possible. Pour rappel, le rapport final du recensement de 2000 a été publié cinq ans plus tard. Les résultats des enquêtes structurelles prévues par le Conseil fédéral devraient en revanche être disponibles un an après la collecte des données.

L'utilité informationnelle

Quelle est à présent l'utilité informationnelle de l'une et l'autre variante? Dans le cas d'enquêtes statistiques, elle dépend bien évidemment du nombre d'unités étudiées. Si la taille de l'échantillon est très petite, cette utilité est faible. Exempte d'erreur, elle augmenterait linéairement avec la taille de l'échantillon. Toutefois, l'utilité d'un échantillon doit être diminuée de l'erreur totale qui lui est inhérente. Comme celle-ci commence par diminuer fortement, l'utilité de petits échantillons s'accroît à mesure que les rendements marginaux augmentent. Puis, une fois que l'échantillon a atteint une taille qui implique l'erreur totale la plus faible, cette dernière croît à nouveau. En d'autres termes, l'utilité informationnelle d'une enquête continue, certes, de progresser, mais les rendements marginaux diminuent (voir *graphique 2*). Les enquêtes statistiques obéissent aussi à la loi du rendement marginal, que l'on connaît de la théorie de l'entreprise.

Dans le cas du recensement de la population, c'est surtout le rapport coût-utilité qui intéresse les milieux politiques. La démonstration a été faite plus haut que les coûts d'une enquête commencent par augmenter de manière linéaire avant de suivre une courbe exponentielle. Si l'utilité informationnelle s'accroît d'abord progressivement avant que sa croissance ne ralentisse après avoir atteint un certain seuil, il en résulte que le coût de chaque unité d'enquête supplémentaire commence par diminuer, puis augmente, et ce de plus en plus rapidement. En d'autres termes, l'utilité supplémentaire produite par chaque franc investi dans une enquête commence par s'accroître avant de décroître de plus en plus vite. Ce raisonnement ne nous dit toujours pas où se situe la taille optimale de l'échantillon, mais une chose est sûre: si l'on tient aussi compte du

rapport coût-utilité, on constate que la solution du Conseil fédéral est préférable à celle des cantons.

L'harmonisation des registres

L'OFS et la Corstat s'accordent à dire qu'il faudra utiliser les registres existants lors du recensement de 2010. Il est dès lors dans l'intérêt de tous que l'harmonisation des registres cantonaux et communaux des habitants se fasse le plus rapidement possible. La loi sur l'harmonisation des registres (LHR), entrée en vigueur le 1^{er} novembre 2006, en fournit la base légale. Les cantons devront très rapidement arrêter des dispositions d'exécution pour assurer pleinement la sécurité du droit. Cette harmonisation des registres consiste pour l'essentiel à faire figurer au minimum les données correspondant aux identificateurs et aux caractères énumérés à l'art. 6 LHR, dans tous les registres des habitants.

Entre la date d'entrée en vigueur de la LHR et celle du recensement de la population à partir des registres, à la fin de 2010, il reste quatre ans. L'inscription du nouveau numéro d'assurance sociale (NAS), de l'identificateur de bâtiment (Egid) et de l'identificateur de logement (Ewid) pose le plus de problèmes. Il sera possible d'inscrire le NAS dans les registres de personnes à partir du 1^{er} janvier 2008, lorsque les dispositions correspondantes de la LHR seront entrées en vigueur. Alors qu'il reste à régler certains détails techniques pour reprendre le NAS dans les registres des habitants, on peut s'attendre à ce que ce soit chose faite d'ici à 2010.

En 2006, l'OFS a effectué un sondage auprès des communes pour évaluer l'avancement de l'harmonisation. Il en est ressorti qu'un assez grand nombre de registres des habitants contenait déjà l'Egid. Il est donc possible de l'intégrer d'ici 2010. Ce sondage a, toutefois, montré que les travaux ne sont pas aussi avancés en ce qui concerne l'Ewid. Compte tenu de la réalité communale, il ne faut guère s'attendre à voir cet identificateur figurer dans tous les registres d'ici à 2010.

Un identificateur de logement qui pose problème

Qu'est-ce que cela implique? L'Ewid sert à reconstituer les ménages: toutes les personnes dotées du même Ewid en forme un. Si l'on ne dispose pas de tous les Ewid, il est difficile de mener à bien une telle reconstitution. Le rapport final de l'OFS nous apprend que celle-ci ne s'est pas faite de manière satisfaisante lors du RFP 2000 et que les objectifs qualitatifs n'ont pas été atteints. Il précise encore que, pour les prochains recensements, il faudra re-

courir à des identificateurs de bâtiment et de logement qui devront figurer dans les registres des habitants et être tenus à jour. De toute évidence, les progrès restent insuffisants sur ce plan si l'on pense à l'échéance 2010. On peut, toutefois, faire remarquer qu'il est aussi possible de reconstituer les ménages en procédant à des contrôles de plausibilité.

C'est donc en toute connaissance de cause que le législateur a prévu, à l'article 19, alinéa 2, de la LHR, la possibilité pour le Conseil fédéral de prolonger le délai d'introduction de l'Ewid au-delà du recensement de 2010. Cependant, les acteurs impliqués, notamment les villes, sont bien décidés à mener à bien cette harmonisation. Divers statisticiens des services cantonaux ou communaux ont déjà fait savoir que, pour eux, 2010 était une échéance réaliste. Ce ne sera, toutefois, pas une catastrophe si les ménages ne sont pas reconstitués à cent pour cent en 2010; c'était déjà le cas en 2000.

Un prix trop élevé à payer

Un projet de recensement de la population doit être évalué de manière pragmatique, en tenant compte de ses différents aspects qui incluent: déranger le moins possible les personnes interrogées, fournir des données aussi actuelles et informatives que possibles, à un niveau géographique suffisamment détaillé, tout en garantissant l'erreur totale la plus faible possible, le rapport coût-utilité le plus avantageux et la meilleure faisabilité technique réalisable. Pour ce qui du degré de détail géographique, la variante des cantons est indubitablement supérieure, puisqu'elle implique une enquête exhaustive. Celle du Conseil fédéral l'emporte au regard de tous les autres critères.

On ne privilégiera la variante des cantons que si l'on accorde une importance extrême au degré de détail géographique. Mais il faudra alors compter avec un rapport coût-utilité nettement plus défavorable, une erreur totale impossible à chiffrer précisément et un travail bien plus important pour la population de notre pays. Le Conseil fédéral juge, non sans raison, que dans ce cas le prix à payer serait trop élevé. ■