

J.-P. RIVET

L'exploitation du recensement de l'industrie sur matériel électronique

Journal de la société statistique de Paris, tome 108 (1967), p. 166-170

http://www.numdam.org/item?id=JSFS_1967__108__166_0

© Société de statistique de Paris, 1967, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

III

L'EXPLOITATION DU RECENSEMENT DE L'INDUSTRIE SUR MATÉRIEL ÉLECTRONIQUE

C'est en mars 1965 qu'il fut décidé de vérifier et d'exploiter les questionnaires du Recensement industriel de 1963 sur matériel électronique. Cette décision tardive, intervenant deux ans après le lancement de l'enquête peut surprendre. Il faut en fait reconnaître que la charge de vérification des questionnaires renvoyés par les entreprises avait au début été sous-estimée, tant dans son ampleur que dans sa complexité. Il est rapidement apparu que les méthodes classiques de vérification (critique des questionnaires par des experts en technique industrielle et en comptabilité) bien que valables dans le fond, étaient inopérantes dans la forme. Les questionnaires les plus complexes ne comportaient pas moins de 300 postes liés entre eux par plusieurs dizaines de relations. Les vérifications opérées manuellement ne pouvaient donc être qu'imparfaites compte tenu de l'ampleur du travail à effectuer. D'autre part, la structure des questionnaires entraînait après vérification des opérations de chiffrement et de perforation passablement complexes où le risque d'erreur était loin d'être négligeable. C'est la nécessité de procéder à une vérification ultime de ces opérations délicates et le manque de puissance de la mécanographie classique qui entraînèrent finalement la décision d'utiliser le matériel électronique de l'I. N. S. E. E.

BREF RAPPEL ⁽¹⁾

Il n'est pas sans intérêt de rappeler ici la structure des questionnaires du recensement. Deux séries de questions étaient posées : questions de structure dites encore « générales » parce que concernant toutes les unités interrogées quelles que fussent les activités exercées (statistiques de secteur), questions de production dites encore « spécifiques » parce que dépendant des activités exercées (statistiques de branche) ⁽²⁾. D'autre part, les unités statistiques étaient de deux sortes : entreprise et établissement. Ces deux unités se confondent évidemment dans le cas d'entreprises à établissement unique. Rappelons enfin que les questionnaires étaient différenciés suivant la taille des unités interrogées, le critère de taille retenu étant le nombre de salariés de l'entreprise.

LES PRINCIPES DE LA VÉRIFICATION

Tous les contrôles s'exercent unité par unité et c'est sans doute là le point fondamental du schéma de vérification retenu : en effet, la cohérence étant assurée au niveau de chaque unité statistique (entreprise ou établissement), les tableaux mécanographiques finalement obtenus sont par définition cohérents et prêts à l'emploi; sans qu'il soit besoin de procéder à de nouveaux ajustements. Ceci n'empêche évidemment pas d'ultimes vérifications, en particulier par comparaison avec les données exogènes par ailleurs connues.

1. On se reportera utilement à l'article de M. Henri Palangié, inspecteur général à l'I. N. S. E. E., paru dans le bulletin.

2. On rappelle que la branche correspond à une activité exercée quelconque alors que le secteur correspond à l'activité principale de l'unité statistique considérée.

D'autre part, on a dissocié la vérification des questionnaires généraux et la vérification des questionnaires spécifiques, tout au moins théoriquement : les structures très différentes de ces deux séries de questionnaires justifiaient à elles seules cette dissociation. De plus, il est rapidement apparu que la complexité même des questionnaires généraux jouait en leur faveur : le nombre de contrôles possibles était tel qu'il devait finalement en résulter des questionnaires cohérents de façon à peu près parfaite. Il n'en allait pas tout à fait de même pour les questionnaires spécifiques pour lesquels il était beaucoup plus difficile d'en rendre cohérentes les différentes données de production, de consommation et de valeur des livraisons. Il était donc finalement assez naturel de chercher d'abord à obtenir des questionnaires généraux les meilleurs possibles, et de s'en servir ensuite pour vérifier les questionnaires spécifiques. En fait, sur le plan concret, les opérations se sont déroulées un peu différemment. Lorsqu'au moment du contrôle des questionnaires généraux était décelée une anomalie, c'est évidemment tout le dossier de l'unité statistique qui était remis en cause, aussi bien questionnaire général que questionnaire spécifique. On voit donc que, bien heureusement, cette dissociation toute théorique des deux schémas de vérification, était beaucoup moins rigide qu'on pourrait le craindre.

LA VÉRIFICATION DES QUESTIONNAIRES GÉNÉRAUX

Ces questionnaires se caractérisent par le fait qu'à un type de questionnaire donné correspondent un certain nombre de cartes mécanographiques bien déterminées. L'identité d'une carte est repérée par un nombre à deux chiffres. Ainsi, la carte 33 porte par exemple les heures ouvrées et les salaires et on sait qu'au questionnaire ET 132 correspondent les cartes 19 à 43.

La vérification se fait au moyen d'une succession de programmes, un programme ne pouvant être mis en route que lorsque le fichier est bon vis-à-vis du programme précédent.

Le premier programme dit *d'introduction* se contente de vérifier que les cartes que l'on se propose de traiter ne contiennent pas d'erreurs grossières qui rendraient l'exploitation future impossible. C'est ainsi que l'on contrôle que les types de carte décelés sont « possibles », que les numéros d'identité des unités statistiques sont également « possibles » (il s'agit du numéro d'identité bien connu à 12 chiffres) c'est-à-dire en particulier que les numéros de département sont bien compris entre 01 (Ain) et 90 (Territoire de Belfort) ⁽¹⁾, enfin que les secteurs d'activité lus dans la carte sont « possibles », c'est-à-dire qu'il s'agit bien d'un numéro à 3 chiffres existant dans la nomenclature des activités économiques, et que ce numéro entre bien dans le champ du recensement, champ entreprise ou champ établissement.

Le deuxième programme dit de *vérification du type de questionnaire*, se propose de faire, pour chaque unité, l'inventaire des cartes correspondantes, c'est-à-dire de s'assurer qu'à telle unité qui avait reçu tel questionnaire (le type de questionnaire était évidemment codé sur les cartes) correspond bien l'ensemble des cartes adéquates sans omission ni double emploi. Un nombre d'erreurs non négligeable a été relevé à cette occasion. On notera qu'à l'issue de ce deuxième programme, aucun contrôle ne s'est encore exercé sur les réponses mêmes des entreprises et des établissements. A ce stade, la structure du fichier est simplement cohérente quant au nombre et à la nature des cartes qui le composent, mais aucune donnée n'a encore été traitée.

1. En 1962, les départements 91 à 95 n'existaient pas.

C'est l'objet de la troisième série de programmes, dite de *codification* qui se propose donc de traiter les questionnaires à proprement parler.

Trois cas d'erreur (ou de possibilité d'erreur) peuvent se présenter en ce qui concerne les données enregistrées :

— il existe des données « non déclarées » (aucune réponse n'a été fournie) qu'il faut tant bien que mal redresser, c'est-à-dire en définitive réinventer compte tenu de ce que l'on sait par ailleurs. Ce procédé n'est évidemment justifié que lorsque les « non-déclarés » sont peu nombreux ;

Il existe des relations mathématiques (égalités ou inégalités) qui ne sont pas vérifiées. Par exemple, l'actif du bilan n'est pas égal au passif pour une raison quelconque. Il y a alors certainement une erreur qu'il faut déceler et corriger ;

Certains ratios ont une valeur suspecte. Par exemple le rapport charges sociales sur salaires versés semble faible. Il y a alors présomption d'erreur qui doit finalement se traduire soit par une correction soit par une confirmation des données. (Si par exemple dans le cas présent il s'agit d'une entreprise qui n'emploie que des cadres à salaire élevé, il est normal que par le jeu du plafond des cotisations, le ratio en question soit très sensiblement inférieur à la limite inférieure de la fourchette fixée.)

Le fichier est considéré comme bon, lorsqu'il n'existe plus de « non-déclarés », lorsque toutes les égalités ou inégalités possibles sont vérifiées, lorsque tous les ratios demeurant suspects ont été confirmés pour une raison ou pour une autre.

D'une façon générale, tous ces « défauts » du fichier se traduisent par une liste de messages d'erreur qui est envoyée à la direction régionale de l'I. N. S. E. E. compétente. Les directions régionales ont pour tâche de reprendre les dossiers (1), de déceler la cause des erreurs signalées, et de modifier leur fichier en conséquence. La réponse à une liste de messages d'erreur se traduit donc par l'envoi à la division électronique d'un jeu de cartes de mise à jour qui, mélangé au fichier initial donne en principe un fichier parfait. Le fichier est parfait si lorsque l'on repasse le programme, aucun message d'erreur n'apparaît. L'expérience a en fait montré que la complexité de l'opération était telle que jamais un fichier parfait n'était obtenu du premier coup et qu'il était nécessaire de procéder en moyenne à deux ou trois mises à jour successives.

Pour des raisons assez évidentes, les différentes unités statistiques ont été traitées séparément, c'est-à-dire qu'un programme a été établi pour les unités « établissement », un autre pour les unités « entreprise », le cas des entreprises à établissement unique étant obtenu par réunion de ces deux programmes. Un dernier programme enfin a effectué la réconciliation entre l'entreprise à établissements multiples et ses différents établissements. L'opération s'est achevée en septembre 1966.

LA VÉRIFICATION DES QUESTIONNAIRES SPÉCIFIQUES

Ces questionnaires se caractérisent par le fait qu'à une unité statistique correspondent un nombre et des types de cartes aléatoires. Ainsi la carte 02 correspond à une production, et il existe une carte 02 par produit. Tel établissement fabriquera un seul produit, un autre en fabriquera une dizaine, un autre n'en fabriquera aucun. Ceci explique que si pour les questionnaires spécifiques il existe également un programme d'introduction dont nous

1. A l'époque déjà tardive où se sont effectués ces contrôles (1966) il n'était plus question pour des raisons évidentes, de retourner auprès des entreprises.

reparlerons ci-dessous, il ne peut en revanche exister de programme s'assurant que le nombre de cartes est correct.

Il ne s'agit plus ici d'un seul programme *d'introduction* mais de toute une série de contrôles des codes. Le programme d'introduction utilisé pour les questionnaires généraux a été repris, avec quelques légères modifications, et complété par d'autres programmes de contrôle. En effet, chaque questionnaire spécifique contenait un grand nombre d'éléments : produits fabriqués, livrés ou consommés, branches productrices, branches consommatrices, clients, zones d'exportation... Tout ceci s'est traduit dans les cartes par un nombre impressionnant de codes dont il fallait au moins s'assurer qu'ils étaient, sinon corrects, tout au moins plausibles, ceci bien avant de savoir si les livraisons ou consommations enregistrées étaient correctes. Plusieurs milliers de codes ont ainsi été mis en mémoire. Il a fallu prévoir plusieurs passages parce que la capacité de mémoire n'était pas suffisante pour traiter le tout en une seule fois.

Cette opération, lourde mais peu complexe, a naturellement pu être lancée alors que la bande des questionnaires généraux était encore en cours de traitement. En revanche, le programme suivant chargé *d'assurer la cohérence* entre les questionnaires spécifiques et les questionnaires généraux n'a évidemment pu être lancé qu'une fois ces derniers jugés corrects. Ce programme, excessivement lourd, fut certainement, et de loin, le plus difficile à concevoir et à rendre opérationnel, la structure même des questionnaires spécifiques rendant par sa complexité, presque impossible la réalisation d'un programme suffisamment général pour être d'une quelconque utilité et suffisamment « spécifique » pour tenir compte des particularités de chaque branche. Quoi qu'il en soit, pour chaque unité, la cohérence a été assurée au niveau de 4 variables fondamentales : la main-d'œuvre ouvrière, le montant des livraisons, les taxes payées, le montant des travaux à façon effectués.

L'enjeu était de taille : c'est la première fois qu'une telle opération, au niveau le plus fin, est tentée, et... réussie, facilitant ainsi grandement la réconciliation secteur-branche et l'établissement des comptes de la nation pour l'année de base 62.

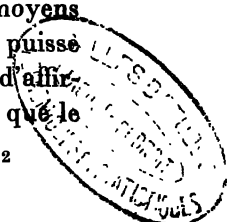
Un dernier programme s'est chargé d'harmoniser différentes données à l'intérieur d'un même questionnaire spécifique. Il n'est d'ailleurs peut-être pas tout à fait le dernier, des contrôles restant à effectuer sur les données en quantités physiques qui n'ont été qu'imparfaitement traitées.

EN CONCLUSION

Au total, cette opération, décidée en mars 1965, lancée en octobre 1965 et achevée en février 1967 aura coûté environ 4 années-travail de programmation et 1 500 heures de 7074 I. B. M.

Elle aurait pu, de loin, être moins complexe, non pas si les questionnaires avaient eux-mêmes été moins complexes, mais surtout s'ils avaient été moins disparates. Des questionnaires qui diffèrent les uns des autres, même légèrement, appellent des programmes différents, qu'il faut écrire, tester et lancer : il en résulte un gaspillage énorme et en programmeurs et en temps machine.

En revanche la complexité du questionnaire lui-même ne doit pas effrayer ; les moyens mis à la disposition du statisticien par l'électronique sont tels qu'aucun problème ne puisse être résolu d'une façon ou d'une autre. Nous irons même plus loin en ne craignant pas d'affirmer que c'est précisément dans le cas où le questionnaire est suffisamment complexe que le



contrôle en devient payant, puisque le nombre de relations exigées entre les différentes données devient tel qu'il est pratiquement exclu de laisser passer une erreur grossière. A la limite, c'est le questionnaire comportant une seule question qui est invérifiable puisque cette question unique ne peut être recoupée par aucune autre.

A l'heure actuelle, on peut dire que les résultats obtenus sont bons, parfois excellents, tant en ce qui concerne les statistiques par secteur que les statistiques par branche; ceci n'est pas douteux. Il nous est malheureusement impossible pour l'instant d'avoir une idée du gain obtenu en qualité après contrôle du fichier par l'électronique. Une telle étude, bien que difficile, n'est pas impossible à envisager.

Quoi qu'il en soit, le nombre d'erreurs redressées est tel que l'on ne peut que frémir à ce qui serait advenu du Recensement de l'industrie si des méthodes plus classiques avaient été conservées. Cette expérience, particulièrement enrichissante, a suffisamment montré qu'une opération de l'ampleur de celle-ci ne pouvait être menée à bien qu'en utilisant les techniques les plus modernes.

Que la division de l'électronique soit ici remerciée chaleureusement de l'énorme contribution qu'elle a apportée au succès de ce premier recensement.

J.-P. RIVET

DISCUSSION

M. M. DUMAS fait préciser qu'un établissement peut avoir des activités dans plusieurs branches et que son secteur correspond à son activité principale; c'est le repère à trois chiffres de cette dernière qui, sauf exception, figure en tête de son numéro I. N. S. E. E. Il signale qu'il connaît le cas de bien des firmes de la mécanique où ces trois chiffres correspondent mal aux activités principales réelles, mais il se déclare favorablement impressionné par l'indication de M. Rivet, comme quoi au cours du dépouillement les erreurs de ce genre, décelées, ont atteint seulement 5 %.

M. M. Dumas se préoccupe également de la façon dont se raccordent les productions déclarées et celles dont la déclaration était attendue d'après les statistiques industrielles, périodiques depuis 1952 : il convient d'attendre le résultat final du travail correspondant, actuellement en cours, pour se prononcer en connaissance de cause, mais M. Rivet manifeste sa confiance que le raccordement sera satisfaisant.

M. de VERDIÈRE demande au conférencier si l'exploitation des questionnaires remplis par les entreprises internationales — entendant par là celles qui possédaient des établissements à la fois en France et à l'Étranger — avait posé à l'I. N. S. E. E. des problèmes particuliers. Il lui est répondu que non.

M. NAHON. — La connaissance des erreurs constatées lors des dépouillements des réponses au recensement a-t-elle été utilisée pour effectuer des corrections dans le fichier des établissements industriels et commerciaux de l'I. N. S. E. E.? Il lui est répondu que la loi sur le secret statistique s'oppose à ce que les résultats individuels de l'enquête soient utilisés à des fins administratives : en particulier, ces résultats ne sauraient servir à effectuer des corrections dans le fichier, même lorsque l'enquête a montré que le numéro d'identité était erroné.