
Mesure de l'activité des entreprises informelles à Mayotte : mise en œuvre et traitements post-collecte

Sylvain DAUBRÉE

Insee, Direction régionale La Réunion - Mayotte

`sylvain.daubree@insee.fr`

Mots-clés. collecte, entreprises, informel, Mayotte, pondération, imputation

Résumé

Mayotte est un département d'outre-mer depuis 2011. De ce fait, le territoire est contraint de fournir des statistiques fiables, notamment sur la contribution des entreprises à l'économie du territoire. Or, si l'économie se modernise rapidement, elle demeure pour une part jusque-là inconnue basée sur des entreprises informelles, c'est-à-dire qui ne sont pas connues des répertoires administratifs ou qui ne tiennent pas de comptabilité.

Un protocole de collecte innovant a donc été mis en place pour mesurer l'impact de ces entreprises informelles ainsi que pour mieux les connaître, d'abord en insérant dans une enquête ménages une question filtre pour repérer les chefs d'entreprises informelles, ensuite en leur proposant un questionnaire adapté au fait qu'ils ne tiennent pas de comptabilité formelle.

Ce papier commencera donc par rappeler le dispositif original mis en place avant de présenter en détail les traitements post-collectes mis en œuvre parmi lesquels une opération d'annualisation des données, le traitement des valeurs manquantes, le traitement des biais de collecte et le traitement de la non-réponse. La dernière partie de l'article dressera, à partir du retour d'expérience de cette première opération, quelques recommandations qui permettront d'améliorer la statistique structurelle d'entreprise à Mayotte.

Abstract

Mayotte has been an overseas region of European Union since 2011. Hence, Insee has to cover it with robust statistics, including business statistics, in order to calculate a reliable gross domestic product. Mayotte's economy is quickly modernizing at the moment. But the informal economy is still significant. A lot of enterprises are missing from administrative data or do not keep proper accounts. In order to address this issue, an innovative data collection process was designed in 2015 in order to assess the impact of informal enterprises. Some questions were added to a household survey allowing to identify informal enterprise's executives. These people subsequently

received a dedicated questionnaire which enabled to estimate their economic activity (turnover, employment, investment).

This paper will first recall the process implemented before detailing the different steps of the post-collect data processes : data annualization, partial and total non-response treatment. Finally, it will focus on possible ways of improvement for structural business statistics in Mayotte.

Introduction

Mayotte est devenu le 101^e département français en avril 2011, suite au résultat du référendum du 29 mars 2009 et a accédé au statut de « Région ultrapériphérique » (RUP) européenne le 1^{er} janvier 2014. Ce nouveau statut rend le territoire éligible au bénéfice des fonds structurels européens. Cependant, pour pérenniser ces fonds, Mayotte doit satisfaire à l'acquis communautaire qui rassemble 35 chapitres dont le chapitre 18 sur les statistiques. Un critère important d'éligibilité est basé sur le PIB moyen par habitant, lequel doit être calculé selon plusieurs approches et notamment via les comptes de branche.

Il a donc été nécessaire de recueillir des données structurelles sur les entreprises mahoraises. En métropole et dans les autres Dom, cela s'effectue à l'aide du dispositif Esane, qui utilise à la fois des données fiscales et des données d'enquêtes. Ce dispositif est basé sur la liste des entreprises inscrites dans les répertoires administratifs et qui donnent régulièrement signe de vie à l'administration. Il ne mesure donc pas l'activité des entreprises qui ne sont pas connues de l'administration : les entreprises informelles. Or, si jusque-là on n'en connaissait pas la part, ces entreprises ont une visibilité importante à Mayotte. Il n'était donc pas possible de se contenter d'étendre le dispositif Esane au territoire mahorais. Un dispositif d'enquête spécifique a donc été mis au point pour mesurer la contribution à l'économie mahoraise de l'ensemble des entreprises. Ce dispositif combine une enquête auprès des entreprises formelles assez classique avec une enquête plus innovante, inspirée d'une méthodologie mise au point pour les pays émergents, visant à mesurer la part dans l'économie des entreprises informelles.

À la suite d'une première présentation réalisée avant la mise en œuvre de l'enquête[5], ce document revient plus en détail, après la réalisation de l'enquête, sur le déroulement de l'enquête auprès des entreprises informelles et la méthodologie employée pour rendre les résultats utilisables.

Après avoir rappelé des éléments de contexte sur les spécificités du territoire mahorais, le chemin ayant mené au dispositif sera présenté. On détaillera ensuite la mise en œuvre de l'enquête sur les entreprises informelles, en s'attachant notamment à décrire les traitements post-collecte, pour conclure sur les enseignements tirés de cette expérience.

1 Mayotte : un département français, mais des caractéristiques singulières

Bien qu'étant un département français, Mayotte a des caractéristiques démographiques, économiques et sociales qui la différencient nettement des autres départements français, y compris ultramarins.

1.1 Une population jeune qui comporte 40 % d'étrangers

En 2017, Mayotte compte 256 500 habitants. Avec 690 habitants au km², il s'agit du département le plus dense après ceux d'Île-de-France. C'est par ailleurs le département le plus jeune, la moitié de la population y a moins de 17 ans contre 23 ans en Guyane et 39 ans en France métropolitaine. Le département est un territoire de destination pour de nombreux immigrés des pays voisins, notamment les Comores et Madagascar. Parmi la population ayant entre 18 et 79 ans, 4 résidents sur 10 sont ainsi de nationalité étrangère. Pour la moitié d'entre eux, leur situation administrative est irrégulière.

1.2 84 % de la population sous le seuil de pauvreté en 2011

En 2011, la moitié de la population mahoraise vit avec moins de 384 euros par mois et par unité de consommation. Le niveau de vie médian a augmenté de 62 % depuis 2005 hors inflation, soit + 8 % par an. Pourtant, 84 % de la population vit encore sous le seuil national de bas revenus.

1.3 Des conditions de logements éloignées des standards métropolitains

Fin 2013, Mayotte compte 53 200 résidences principales, dont 37 % sont des maisons en tôle. La majorité des ménages vit dans des conditions de logement difficiles, éloignées des standards des autres départements français. Ainsi, 28 % ne disposent pas d'eau courante dans leur logement et 3 400 vivent sans électricité.

1.4 Un taux de chômage à 26 % qui touche particulièrement les jeunes

Au 2^e trimestre 2017, 48 900 mahorais âgés de 15 à 64 ans occupent un emploi soit seulement 38 % des personnes en âge de travailler, contre 69 % en métropole. Le chômage s'établit à 26 %, en légère baisse par rapport à 2016. Comme dans les autres départements ultramarins, les jeunes sont particulièrement concernés (43,1 %).

Ce chômage important se développe dans un contexte de faible qualification de la population, puisque 7 mahorais sur 10 de 15 ans ou plus n'ont pas de diplôme qualifiant. Un tiers de la population n'a jamais été à l'école, et parmi les personnes en âge de travailler, 60 % ne maîtrisent pas les compétences de base à l'écrit en langue française.

Les caractéristiques de la société mahoraise synthétisées ci-dessus témoignent ainsi de spécificités fortes par rapport au reste du territoire français. La forte natalité, le fort taux d'illettrisme en français, le faible niveau scolaire, le faible taux d'emploi, les conditions de logements et les niveaux de revenus rappellent des constats faits dans beaucoup de pays africains.

Pourtant, le changement de statut de Mayotte, devenu département en 2011 mais aussi « Région Ultrapériphérique » (RUP) européenne depuis le 1er janvier 2014, imposent la construction d'un système statistique robuste, conforme aux standards nationaux. En tant que RUP, Mayotte bénéficie de fonds structurels européens. Pour pérenniser ces fonds, Mayotte doit satisfaire à l'acquis communautaire qui rassemble 35 chapitres dont le chapitre 18 sur les statistiques. Un critère important d'éligibilité est basé sur le PIB moyen par habitant, lequel doit être calculé

selon plusieurs approches et en intégrant des comptes de branches. En France et dans les quatre DOM historiques, la mesure du PIB repose sur l'existence de fichiers administratifs robustes complétés d'enquêtes auprès des entreprises.

Les spécificités décrites ci-dessus ne permettent donc malheureusement pas à Mayotte d'intégrer le système d'élaboration des statistiques structurelles d'entreprises (Esane) mis en place en France métropolitaine et dans les autres DOM, indispensable pour le calcul du PIB. Et ce principalement parce que :

- une base de sondage de qualité suffisante n'est pas encore disponible ;
- les données administratives ne sont pas de qualité suffisante ;
- l'économie informelle est plus prégnante que dans les autres départements et il convient donc d'en avoir une estimation.

Pour pallier ces problèmes il a donc été décidé de mettre en place le système d'enquêtes qui est décrit ci-dessous.

2 Une première enquête entreprise expérimentale en 2012 débouchant sur un dispositif d'enquête en deux parties

La construction du système statistique d'entreprises repose en France sur la construction d'un répertoire statistique de qualité (Sirus), ce dernier étant adossé au répertoire inter-administratif Sirene. Le répertoire Sirene est encore très récent à Mayotte et comporte des fragilités.

Sirene s'est construit à partir de 2008. Auparavant, il n'y avait pas de répertoire unifié, chaque administration gérait un répertoire d'entreprises qui lui était propre. Les fichiers administratifs sont encore partiels à Mayotte (fichier des impôts notamment), toutes les entreprises ne faisant pas les démarches obligatoires. Ils n'ont donc pas encore le niveau de qualité et d'exhaustivité nécessaire à l'établissement de statistiques robustes.

2.1 Mise en place d'une première expérimentation

Une première enquête expérimentale auprès des entreprises a été conduite en 2012 par la direction régionale de l'Insee de La Réunion - Mayotte (enquête papier). 8 600 entreprises potentiellement actives ont été identifiées à partir de différents fichiers administratifs : le répertoire Sirene mais aussi des fichiers administratifs et fiscaux. Parmi ces entreprises, 1 300 (soit 15 % d'entre elles) n'étaient alors pas présentes dans Sirene. Par ailleurs, en raison de l'absence d'un répertoire centralisé historique, certaines entreprises pouvaient avoir jusqu'à trois numéros Siren différents et être ainsi présentes plusieurs fois dans la base. Certains doublons pouvaient également être générés par des problèmes d'état civil (homonymes nés à la même date, favorisés par des dates de naissance approximatives qui s'accumulent au 1^{er} janvier de chaque année). De plus, beaucoup d'entreprises étaient probablement cessés sans que l'on ne puisse en estimer le nombre. Ce problème existe également en métropole et dans les autres Dom, mais il est encore plus flagrant à Mayotte : sur 2017, on a ainsi compté 5 cessations seulement pour 928 créations.

Cette enquête a permis de tirer les enseignements suivants :

1. il est difficile d'organiser une enquête papier à Mayotte en raison d'une part d'un adressage encore très partiel, et d'autre part d'un service de distribution du courrier défaillant. Ainsi, le taux de retour de courrier non distribué s'est élevé à 22 %, et on soupçonne le taux de non-distribution du courrier d'être encore plus important ;

2. Le taux de réponse a été faible : 14 %, et il était impossible de savoir si les non-répondantes étaient encore en activité. Autrement dit, il était impossible de faire la différence entre une entreprise non-répondante et une entreprise hors champ de l'enquête, rendant les traitements post-collecte particulièrement périlleux ;
3. Les entreprises déposant des liasses fiscales avaient un taux de réponse spontané correct et une qualité des réponses jugée bonne ;
4. Les entreprises ne déposant pas de liasses fiscales avaient beaucoup de difficulté à remplir le questionnaire qui leur était envoyé, car, ne tenant pas de comptabilité, elles ne disposaient pas des éléments demandés.

À l'issue de cette expérimentation, il est donc apparu qu'il n'était pas possible, avec des enquêtes entreprises classiques comme celles menées en métropole, d'évaluer le nombre d'entreprises actives à Mayotte et d'estimer des agrégats (total de la valeur ajoutée, du chiffre d'affaires ou des effectifs par grands secteurs par exemple) même en se restreignant au seul secteur formel de l'économie. Les spécificités du territoire de Mayotte avec un adressage encore très partiel et beaucoup d'entreprises informelles ont conduit à se tourner vers la méthodologie des enquêtes 1-2-3 mise en place par l'institut de recherche et de développement Dial (Développement, Institutions et Mondialisation) pour essayer de bâtir un système d'information adapté à la réalité du terrain.

Cette première expérience a également permis d'engager une nouvelle opération qualité sur Sirene afin de rendre le répertoire le plus exhaustif possible. Les fichiers des partenaires (chambre de commerce et d'industrie, chambre des métiers et de l'agriculture, DRFiP principalement) avec lesquels cette opération a été mise en place sont maintenant cohérents avec Sirene à 99,6 %.

2.2 Un dispositif en deux piliers hérité des conclusions de l'enquête expérimentale

Forts des constats et difficultés mesurés lors de la première enquête expérimentale, l'Insee a décidé de mettre en place un système d'enquêtes s'appuyant sur deux piliers, dont le deuxième s'inspire des enquêtes 1-2-3, utilisées dans les pays en voie de développement pour mesurer les activités informelles. L'objectif de ce dispositif est de mesurer l'activité économique des entreprises de Mayotte, formelles et informelles, et de permettre la construction de comptes économiques ventilés par branche. Les deux piliers mis en œuvre sont les suivants :

1. un premier pilier « classique » visant les entreprises formelles. Il s'agira d'interroger par courrier les entreprises ayant déposé une liasse fiscale au cours d'au moins un des trois exercices précédents, avec des relances téléphoniques. Comme ces entreprises ont déjà déposé une liasse fiscale, elles tiennent une comptabilité et sont donc en mesure de répondre à un questionnaire de type comptable. Cette première enquête a été réalisée sur les exercices fiscaux 2013, 2015 et 2016 et a vocation à être renouvelée tous les ans ;
2. un second pilier visant les entreprises informelles. Ici, deux difficultés : la première est de les identifier, car elles ne sont pas toutes présentes dans les répertoires administratifs, et la seconde est de les interroger. En effet, pour beaucoup, elles ne tiennent pas de comptabilité et ne sont donc pas en mesure de comprendre et encore moins de répondre à des questions telles que « quel est votre chiffre d'affaires sur l'année 2015 ? ». Des questions spécifiques du type « travaillez-vous à votre compte ? » ont donc été intégrées dans une enquête ménage s'adressant à l'ensemble de la population mahoraise afin de repérer un échantillon d'entreprises informelles, puis, des enquêteurs sont venus les aider à remplir un questionnaire adapté qui permet de reconstruire la comptabilité de l'entreprise à partir des éléments que l'enquêté est en mesure de fournir.

Initialement, un troisième pilier s'intéressant spécifiquement aux entreprises présentes dans le répertoire Sirene mais ne tenant pas de comptabilité (entreprises semi-formelles) était prévu, mais a finalement été abandonné compte tenu des difficultés de repérage rencontrées. Ces entreprises ont été enquêtées dans le second pilier, de la même manière que les entreprises absentes de Sirene.

L'économie informelle

La comptabilité nationale d'un territoire intègre l'ensemble de l'activité économique qui y est exercée. Or, en France, la comptabilité nationale reposant souvent sur des sources administratives (notamment fiscales), elle ne prend pas en compte *a priori* les activités économiques inconnues des sources administratives (fiscales ou d'emploi). Les données de ces sources sont donc corrigées d'une évaluation du montant des activités dites « informelles ». Au sens large, cette activité peut être de différentes natures :

- la production dissimulée réalisée par des entreprises sans existence légale ;
- la production dissimulée réalisée par des entreprises ayant une existence légale ;
- la production illégale ;
- la production manquante en raison des carences du dispositif de collecte.

Dans cette enquête, seule une partie de l'économie informelle est couverte. Les « entreprises informelles » correspondent aux entreprises ayant ou non une existence juridique matérialisée par l'inscription au répertoire Sirene, mais qui ne sont pas connues de l'administration fiscale au sens des déclarations fiscales.

Le champ de l'enquête ne couvre donc pas la production économique réalisée par des entreprises connues de l'administration fiscale, mais qui ne déclarent qu'une partie de leur activité, ni la production des activités illégales. Il ne couvre pas non plus la production des entreprises agricoles informelles.

2.3 Un pilier 1 robuste qui a produit des résultats nouveaux sur une partie du champ

Le pilier 1 a été mené pour la première fois en 2014 sur l'exercice 2013. L'échantillon a été constitué des entreprises ayant déposé au moins une liasse fiscale entre 2011 et 2013 et appartenant à l'ensemble des secteurs principalement marchands, non agricoles et non financiers. Ce champ est équivalent au champ des enquêtes structurelles d'entreprises menées en métropole et dans les autres Dom, enquêtes qui sont l'un des maillons essentiels pour l'élaboration des comptes nationaux.

1 705 unités légales ont ainsi été interrogées, 1 180 ont répondu dont 120 qui se sont révélées hors champ de l'enquête (entreprises cessées essentiellement). Cela représente un taux de réponse de 69 % en nombre d'entreprises et de 90 % en chiffre d'affaires déclarés aux impôts soit un taux de réponse équivalent à celui obtenu à La Réunion.

Cette première enquête a ainsi permis de mesurer pour la première fois la richesse créée par ces entreprises et la répartition de leurs activités par grands secteurs à Mayotte. Ces entreprises ont ainsi généré 448 millions d'euros de valeur ajoutée en 2013. L'exploitation de cette enquête, qui s'est centrée sur les entreprises de 1 à 500 salariés, a permis une comparaison avec la France métropolitaine. À Mayotte, la moitié de la richesse créée est issue du commerce et de la construction, ces deux secteurs ayant un poids supérieur à la métropole. C'est aussi le cas du secteur des transports qui représente 12 % de la valeur ajoutée des entreprises de 1 à 500 salariés contre 5 % en métropole.

Cette première photographie des entreprises privées mahoraises a représenté un pas important dans la connaissance de l'économie de ce territoire. Elle ne couvre cependant qu'une partie de l'activité privée : celle des entreprises les plus formelles, déposant une liasse fiscale aux services des impôts. Afin d'avoir une mesure complète de l'activité économique de Mayotte, cette mesure doit être complétée d'une mesure de l'activité informelle. C'est ce que prévoit le pilier 2 mis en œuvre en 2016.

3 Le pilier 2, une enquête proche des enquêtes 1-2-3 classiques

Après l'abandon du pilier concernant les entreprises ne déposant pas de liasses fiscales mais présentes dans Sirene, le pilier 2 est devenu essentiel pour mesurer l'activité économique de toutes les entreprises inconnues des services fiscaux. Ce sont des entreprises pour lesquelles on ne dispose d'aucune information comptable et que l'on ne peut repérer facilement sur le terrain. Mener une enquête de ce type doit permettre à l'Insee de mesurer la richesse créée par ces entreprises et d'en déduire des coefficients d'extrapolation par secteur que l'on pourra appliquer aux résultats des enquêtes annuelles du pilier 1, plusieurs années de suite. L'objectif de l'enquête était double :

- estimer le nombre d'unités de production informelles et leur poids économique ;
- caractériser au mieux le secteur informel.

Le principe des enquêtes 1-2-3 [1] consiste à utiliser une enquête auprès des ménages pour identifier des entrepreneurs individuels. Le plus souvent cette enquête auprès des ménages est une enquête sur l'emploi. A Mayotte, une enquête annuelle sur l'emploi est bien mise en place depuis 2013, mais elle est en partie panelisée, c'est-à-dire que les deux tiers des répondants sont réinterrogés l'année suivante. Compte-tenu de ce mode d'enquête, l'Insee n'a pas souhaité que l'on greffe à cette enquête emploi en cours de déploiement une autre enquête qui aurait pu détériorer les taux de réponses des individus réinterrogés l'année suivante. La seule possibilité aurait été de n'utiliser que le tiers sortant de l'enquête comme filtre mais cela réduisait beaucoup trop l'échantillon. D'autre part cela risquait également de faire baisser le taux de réponse à l'enquête « entreprise » par lassitude des répondants. C'est donc finalement l'enquête Migration Famille Vieillesse prévue en collaboration avec l'INED en 2015-2016 à Mayotte qui a été retenue comme enquête filtre pour repérer les entrepreneurs. Cette enquête avait le double avantage de s'appuyer sur un échantillon important (plus de 3 000 répondants) et d'aborder la thématique de l'emploi.

3.1 L'enquête filtre utilisée : enquête sur les migrations, la famille et le vieillissement

Migration, famille et vieillissement (MFV) est une enquête auprès de l'ensemble de la population mahoraise qui s'est déroulée entre septembre 2015 et janvier 2016 pour la première vague (4 600 logements), et février et avril 2016 (1 000 logements) pour la seconde. Le filtre permettant d'identifier les entreprises informelles était constitué de plusieurs questions qui concernaient chacun des habitants du logement. La première, pour les personnes entre 15 et 79 ans qui ont actuellement un emploi : « actuellement, <prenom> exerce cette profession... 1 – à son compte 2 comme salarié 3 comme aide ». Les personnes à leur compte étaient considérées comme éligibles. D'autres questions étaient posées pour rattraper les personnes n'exerçant pas actuellement une activité, mais en ayant exercé une dans les six derniers mois. Dans ce cas on listait jusqu'à deux de ces activités, en demandant à chaque fois si la personne était à son compte ou non, ainsi

que le secteur de cette activité. Les deux vagues ont permis d'établir une liste de 477 unités de production à interroger. À ce stade, des disparités importantes dans les taux d'éligibles, par enquêteur ainsi que par vague de MFV, ont fait craindre un biais de collecte important, qui a été étudié (voir plus bas).

3.2 Un questionnaire largement inspiré de la deuxième phase de l'enquête 1-2-3 malgache

Le questionnaire utilisé pour le pilier 2 est très largement repris du questionnaire mis au point pour la deuxième phase de l'enquête 1-2-3 mise en place à Madagascar [2]. Il a été raccourci pour limiter le temps de collecte. Les questions sur la concurrence, sur le financement, ainsi que le module sur les problèmes et les perspectives ont été retirés, car ils ont été jugés peu pertinents au regard soit des objectifs de l'enquête, soit des spécificités du territoire. L'enquête a été préparée en collaboration avec l'institut de statistiques de Madagascar (Instat), qui a aidé l'Insee à mettre en place le questionnaire, a fourni et adapté l'outil de saisie, et a formé les enquêteurs lors du test. Dial a également fourni un appui méthodologique tout au long de l'enquête, et a participé à différentes phases terrain.

Au final, le questionnaire tient sur une feuille A3 et comporte les modules suivants :

- un premier module sur les caractéristiques de l'entreprise (local, équipement, ancienneté...);
- un module main d'œuvre qui liste les personnes travaillant pour l'entreprise et récupère le détail de leur rémunération;
- un module production qui liste l'ensemble des produits et services vendus par l'entreprise;
- un module dépenses et charges qui liste les coûts supportés par l'entreprise pour produire les biens et services listés dans le module production. En effet, pour simplifier le questionnaire, les stocks ne sont pas pris en compte;
- un module rythme visant à passer des données mensuelles aux données annuelles;
- un module clients et fournisseurs;
- un module équipement qui permet notamment de calculer l'investissement.

Le principe du questionnaire est d'obtenir une comptabilité annuelle pour l'unité de production en partant du principe que l'enquêté ne tient pas de comptabilité, et qu'on doit donc se baser, au moins en partie, sur sa mémoire. Ainsi, pour chaque produit ou service vendu, l'enquêteur va demander à l'enquêté s'il est capable de dire combien ont été vendus sur un mois. S'il n'est pas capable de le dire on descend à la semaine. Et s'il ne peut estimer combien il en a vendus sur une semaine, on lui demande sur la journée précédant l'enquête, et ainsi de suite pour tous les produits. Pour les charges, on procède de même : certaines charges sont connues au mois, comme le loyer, d'autres à l'année, comme une assurance, d'autres encore au jour : à chaque fois, on part de ce que l'enquêté connaît et on reconstruit *a posteriori* un mois de comptabilité. Ensuite, le module rythme (détaillé plus bas) demande à l'enquêté de positionner l'activité du mois par rapport aux autres mois, afin de pouvoir déterminer si, pour obtenir l'année, on multiplie par 12 ou par un autre coefficient.

3.3 La collecte et son bilan

La collecte s'est déroulée au plus près de celle de MFV, pour faciliter le repérage et limiter l'attrition. Comme la collecte de MFV était scindée en deux vagues, et qu'une seule de ces vagues ne pouvait fournir un échantillon suffisant pour le pilier 2 (l'objectif fixé étant d'interroger environ 500 unités de production), la collecte du pilier 2 a elle-même été scindée en deux vagues de 416

et 61 enquêtes. Elle a été réalisée par l'ensemble des douze enquêteurs de l'Insee à Mayotte. Un effort important a été réalisé pour la formation et le suivi des enquêteurs. En effet, le test de 2014 avait permis de se rendre compte que l'enquête était déroutante pour eux à de nombreux titres. Tout d'abord, ils n'avaient jusque-là réalisé que des enquêtes ménages, et n'ont pas de notion de comptabilité. Ensuite, ils ne sont pas tous à l'aise avec les manipulations calculatoires que requiert la passation du questionnaire : règle de trois, conversion d'unités, addition rapide pour vérifier un ordre de grandeur. Enfin, alors que, dans les enquêtes ménages, la consigne est de poser la question aussi littéralement que possible pour ne pas introduire de biais, il fallait là, au contraire, bien comprendre à chaque fois le fonctionnement de l'entreprise pour ne pas risquer de rater un élément que l'enquêté oublierait de déclarer (par exemple une partie des charges, comme le loyer ou l'électricité, auxquels l'enquêté n'aurait pas spontanément pensés).

La période de collecte de la première vague étant très courte (un mois), il fallait s'assurer que les enquêteurs avançaient vite. Ce rythme nécessairement rapide imposait également de suivre de très près les enquêteurs pour être capable de corriger une consigne mal comprise, ou d'adapter un point du protocole. Par conséquent, deux mesures ont été mises en place : des accompagnements fréquents des enquêteurs, et une remontée très régulière des questionnaires comprenant une relecture avec l'enquêteur. Ces relectures permettaient de contrôler la bonne compréhension du protocole par l'enquêteur ainsi que la cohérence interne du questionnaire. En cas de problème, le questionnaire était corrigé en se basant sur la mémoire de l'enquêteur, les notes qu'ils étaient encouragés à prendre ou, si nécessaire, un retour chez l'enquêté. Il est à noter que, en dépit d'une montée en compétence certaine des enquêteurs, cette phase de relecture a été indispensable jusqu'à la fin de l'enquête : il n'a pas été possible de faire en sorte que les enquêteurs se relisent seuls, et jusqu'à la fin, des questionnaires inutilisables remontaient. Outre la cohérence du questionnaire, le relecteur codifiait les activités et mensuralisait tous les montants qui étaient renseignés. Les accompagnements, quant à eux, servaient d'une part à s'assurer que l'enquêteur avait bien intégré la formation, et d'autre part à assister celui-ci lors des enquêtes complexes, c'est-à-dire les magasins d'alimentation générale, où la masse de produits à lister rendait nécessaire le fait d'être deux pour ne rien oublier.

Comme l'enquête ciblait des personnes qui avaient été enquêtées quelques semaines auparavant, le repérage des entreprises n'a pas posé de grosses difficultés. Le risque évoqué en début de collecte d'avoir beaucoup de gens ayant leur entreprise loin de leur domicile n'était finalement pas fondé, la très grande majorité des enquêtés travaillant dans leur domicile ou à proximité immédiate.

Dès le début de la collecte, une difficulté inattendue a été identifiée : de nombreux petits magasins ne connaissaient pas le prix d'achat des marchandises qu'ils vendaient. Cela faisait peser un risque important sur l'enquête dans la mesure où sans prix d'achat, il devient impossible de calculer une marge commerciale, et donc l'ensemble des soldes comptables qui en dépendent (dont la valeur ajoutée qui est un des éléments de base de la comptabilité nationale). Il a donc été décidé d'identifier les questionnaires qui posaient problème, de récupérer le maximum d'informations sur les produits dont le prix d'achat était inconnu (magasin, conditionnement, marque) afin de permettre, après l'enquête, d'aller récupérer les prix de ces produits auprès des magasins dans lesquels ils ont été achetés (une particularité mahoraise est que les commerçants ne s'approvisionnent pas chez des grossistes mais dans des grandes surfaces ouvertes à tous). Ce processus a permis de récupérer 44 % des prix manquants et ainsi de réduire d'autant le volume de non-réponse partielle à traiter. Cette action de collecte complémentaire a donc eu un impact direct et fort sur la qualité finale de l'enquête. Les cas pour lesquels il n'a pas été possible d'obtenir un prix concernaient les produits pour lesquels les enquêteurs n'avaient pas notés suffisamment d'information pour permettre d'identifier le produit. Ces derniers ont fait l'objet d'une imputation statistique (voir partie suivante).

D6a. Comment a varié l'activité de votre établissement au cours des 12 derniers mois ?

RYTHME												
mois en clair	m-1	m-2	m-3	m-4	m-5	m-6	m-7	m-8	m-9	m-10	m-11	m-12
1. Maximum												
2. Moyen												
3. Minimum												
0 Pas d'activité												
Codage												
D6b.												
Recette mensuelle max. €											_ _ _ _ _ _ _	
Recette mensuelle min. €											_ _ _ _ _ _ _	
D7. Employez-vous de la main d'œuvre temporaire ?												
1. Oui 2. Non ----->											_	
Si Oui, Combien de temporaires au cours du mois d'activité maximale ?											_ _ _	

FIGURE 1 – Module rythme du questionnaire

Les taux de réponses ont été satisfaisants, sur les 477 questionnaires, on compte 58 % de réponse, 35 % de hors champ et 7 % de refus ou d'absence. Le taux élevé de hors champ s'explique par l'enquête filtre pas très adaptée qui a engendré une qualité insatisfaisante des réponses aux questions filtres. Le problème provenait des questions sur le secteur d'activité, qui n'ont donc pas pu être utilisées en amont de l'enquête pour exclure les secteurs hors champ, comme l'agriculture.

Parmi les répondants, 8 % ont déclaré tenir une comptabilité et il a été possible de les identifier dans la liste des déclarants fiscaux fournie par la DRFiP. Ceux-là n'ont donc pas dû remplir le questionnaire pilier 2 pour lequel ils ont été considérés comme hors champ mais ont été intégrés dans l'enquête courrier du pilier 1 sur les entreprises formelles qui s'est déroulée quelques mois plus tard.

4 La phase de traitement aval

4.1 Passage à l'annuel

Le questionnaire du pilier 2 s'attache essentiellement à reconstruire la comptabilité d'une entreprise sur le mois précédant l'enquête, afin de minimiser les biais de mémoire des enquêtés qui ne tiennent pas de comptabilité. Pour obtenir des résultats annuels, une technique brute serait de multiplier les résultats par 12. Cependant, une partie importante des activités de l'île ont une forte saisonnalité, liée soit au climat (moins d'activités pendant la saison humide), soit au calendrier scolaire et religieux (pic d'activité pendant le mois de ramadan lié à la forte hausse de la consommation de denrées alimentaires et d'habillement). Aussi, un module spécifique est présent dans le questionnaire pour décrire cette saisonnalité afin de pouvoir ensuite en rendre compte dans le calcul des données annuelles.

Le module de saisonnalité (voir figure 1) est composé de deux parties : d'abord, on demande pour chaque mois de l'année écoulée (y compris le mois de l'enquête) si l'activité a été maximum, moyenne, minimum ou nulle, puis on demande quel a été le chiffre d'affaires pour les meilleurs mois et pour les moins bons. Une dernière question concerne le recours à une main d'œuvre temporaire éventuellement employée pendant les meilleurs mois, mais les enquêtés ont systématiquement répondu non à cette question.

Il était indiqué aux enquêteurs qu'à la place des valeurs minimales et maximales, ils pouvaient

saisir un rapport si les enquêtés n'avaient pas idée des valeurs (par exemple, si l'enquêteur arrive à savoir que l'activité est deux fois plus importante les meilleurs mois que les moins bons, il remplit 1 dans la valeur min et 2 dans la valeur max). Le passage à l'année s'effectue en deux opérations :

- calcul d'un coefficient d'annualisation : par combien faut-il multiplier une valeur mensuelle pour obtenir une valeur annuelle ;
- détermination des variables concernées par ce coefficient, et de celles qui ne sont pas sujettes à la saisonnalité (par exemple, on va utiliser le coefficient d'annualisation pour les achats de matières premières, mais pas pour le loyer qu'on va simplement multiplier par 12).

Le calcul du coefficient d'annualisation sera différent selon que le mois de référence est un mois moyen, minimal ou maximal. Le mois de référence ne peut pas être un mois sans activité. Ce calcul va dans tous les cas se baser sur le rapport entre la valeur de la recette mensuelle minimale et la recette mensuelle maximale, $r = \frac{valMin}{valMax}$. En effet, comme les enquêteurs ont collectés dans certains cas de vraies valeurs des minimums et maximums, et dans d'autres directement leur rapport, seul ce dernier va être utilisé. Pour mener à bien le calcul, il est également nécessaire de faire une hypothèse sur la valeur du mois moyen par rapport aux valeurs minimale et maximale. Faute de mieux, on fait l'hypothèse que la valeur moyenne se situe entre les deux, c'est-à-dire que $valMoy = \frac{valMin+valMax}{2}$.

Sous ces hypothèses, on arrive aux résultats suivants :

- si le mois d'enquête est un mois moyen :

$$C_{annualisation} = nbMoisMoy + nbMoisMax \times \frac{2}{r+1} + nbMoisMin \times \frac{2}{1+1/r}$$

- si le mois d'enquête est un mois minimum :

$$C_{annualisation} = nbMoisMin + nbMoisMax \times \frac{1}{r} + nbMoisMoy \times \frac{1+1/r}{2}$$

- si le mois d'enquête est un mois maximum :

$$C_{annualisation} = nbMoisMax + nbMoisMin \times r + nbMoisMoy \times \frac{1+r}{2}$$

Tout ceci repose donc sur l'hypothèse forte sur la valeur de $valMoy$. La validité de cette hypothèse a été testée : cela n'est possible que quand le mois de l'enquête est un mois moyen, et que l'enquêteur a rempli des vraies valeurs pour les recettes mensuelles max et min, et non un rapport comme cela leur avait été proposé. Dans la majorité des cas, les valeurs estimées et réelles sont proches. Il y a néanmoins des cas où l'écart est important. Aussi, la méthode est infléchiée pour que, dans les cas où les valeurs minimales et maximales sont plausibles et que le mois de l'enquête est un mois moyen (10 questionnaires), elles soient utilisées directement plutôt qu'estimées. Ceci donne dans notre cas des résultats qui semblent plus juste, mais va à l'encontre du postulat, vérifié par ailleurs, que les entrepreneurs ne sont pas capables d'estimer de manière correcte leur chiffre d'affaires mensuel. La formule devient :

$$C_{annualisation} = nbMoisMoy + nbMoisMax \times \frac{valMoisMax}{valMoisMoy} + nbMoisMin \times \frac{valMoisMin}{valMoisMoy}$$

Le coefficient est ensuite appliqué à tout ce qui concerne la production, les achats de marchandises ou de matières premières. En ce qui concerne les charges, il est appliqué pour l'eau,

l'énergie, les communications, le transport. En revanche, pour le loyer, l'assurance, les rémunérations, les taxes et impôts (qui sont indépendants de l'activité : il s'agit par exemple d'un droit d'enregistrement ou d'une patente), le montant est simplement multiplié par 12.

Ceci doit néanmoins être réalisé avec précaution : l'expertise a fait ressortir quelques cas particuliers où le coefficient calculé ne doit pas être appliqué. C'est le cas par exemple de certains commerçants pour qui les frais de transports correspondent à l'acheminement d'un conteneur annuel. Dans ce cas, l'enquêteur a donné la valeur annuelle, qui a été divisé par 12 pour entrer dans le questionnaire, et qu'il faut donc à nouveau multiplier par 12 pour obtenir la valeur juste.

A posteriori, il est possible de tirer quelques enseignements concernant cette partie du questionnaire, enseignements qui pourraient être mise en œuvre lors d'une éventuelle réédition de l'enquête. Ainsi, ce traitement aurait été simplifié si, au lieu de ne saisir que les montants mensualisés de chacun des postes, il avait été décidé de saisir les valeurs « brutes » avec leur temporalité. En effet, dans le questionnaire papier on a l'information donnée par l'enquêteur avec sa temporalité, par exemple, 1 500 € de transport par an, ou 5 € par semaine. Ce chiffre est ensuite mensualisé, et c'est cette valeur qui est saisie. Bénéficiaire de toute l'information donnée par l'enquêteur aurait permis de différencier les traitements, par exemple, de dire que, pour une donnée avec une temporalité annuelle, on la prend directement alors que si la temporalité est mensuelle, ou infra-mensuelle, on la retravaille. Par exemple, pour le poste transport, on est face à une charge qui dépend *a priori* de l'activité : pour la majorité des commerçants, plus d'activité signifie plus de frais de transports pour aller chercher la marchandise. Mais pour les quelques-uns qui fonctionnent avec une livraison unique d'un conteneur pour toute l'année, on fait un contre-sens en faisant dépendre ces charges de l'activité, contre-sens qui a dû ici être traité à la main.

Le calcul du coefficient aurait par ailleurs gagné en précision si nous avions également demandé la recette du mois moyen, en plus de la recette du meilleur mois et du moins bon mois (par exemple : « entre le mois le moins bon et le meilleur, il y a un facteur 5. Mais entre le mois minimum et le mois moyen, a-t-on a un facteur 2 ? 3 ? »). Enfin, la formation aurait sans doute pu consacrer encore un peu plus de temps à ce module dont l'importance dans les résultats finaux est capitale, et qui n'a pas toujours été parfaitement compris par les enquêteurs.

4.2 Traitement des valeurs manquantes

L'ensemble du dispositif de collecte ayant été mis en place pour limiter autant que possible les données manquantes ou de mauvaise qualité, les questionnaires bruts étaient relativement complets. Les manques les plus gênants se situent au niveau des prix d'achats de marchandises (problème déjà évoqué ci-dessus), ou au niveau de certaines charges. Pour les prix des achats absents restant (un peu moins d'une centaine de produits, soit 6,5 %), l'imputation est réalisée de manière à ce que le taux de marge commerciale¹ sur les objets dont le prix d'achat est manquant soit une moyenne pondérée de la marge commerciale du magasin et de la marge commerciale du produit concerné. La pondération est choisie de manière à privilégier la méthode la plus solide, c'est-à-dire que si la proportion de produit d'un magasin à imputer est forte, on va privilégier la moyenne de la marge commerciale du type de produit, et inversement si elle est faible. Ainsi, le taux de marge imputée du produit P de l'entreprise E vaut, avec la proportion de produits P à imputer au total, β_E la proportion de produits à imputer pour l'entreprise E ,

1. La marge commerciale est la différence entre le prix de vente et le prix d'achat d'un produit. Le taux de marge commerciale (abrégé taux de marge par la suite) est le rapport entre la marge commerciale et le prix de vente.

$txMarge(P)$ le taux de marge du produit P sur l'ensemble de l'enquête, et $txMarge(E)$ le taux de marge de l'entreprise E , à chaque fois en se restreignant aux produits dont on connaît le prix d'achat :

$$txMargeImputé(P, E) = \frac{(1 - \alpha_P) \times txMarge(P) + (1 - \beta_E) \times txMarge(E)}{\alpha_P + \beta_E}$$

Par exemple, si on a besoin d'imputer le taux de marge commercial d'un paquet de biscuit, pour l'entreprise DukaTop, on sait que :

- ce produit n'est à imputer qu'une seule fois pour toute l'enquête, alors qu'il est présent à 10 reprises,
- l'entreprise qui le vend a 3 autres produits à imputer, sur un total de 10 produits vendus,
- le taux de marge commercial sur ce produit, pour l'ensemble de l'enquête est de 40 %,
- le taux de marge de l'entreprise, en dehors des produits à imputer, est de 20 %.

Alors on a :

$$txMargeImputé(biscuit, DukaTop) = \frac{0.9 \times 0.4 + 0.3 \times 0.2}{0.1 + 0.7}$$

Pour l'imputation des charges (impôts, électricité, transports, intérêts versés), il y a à chaque fois très peu de cas à imputer. Comme on a globalement peu d'individus, on considère arbitrairement que ces charges sont déterminées par le chiffre d'affaires et le domaine d'activité, et on impute chaque charge en conséquence. Pour l'électricité par exemple, on calcule la médiane du rapport électricité sur chiffre d'affaires pour chacun des répondants, on considère que ce rapport vaut pour les entreprises à imputer, et on en déduit les charges d'électricité de l'entreprise pour lesquelles ces charges sont manquantes.

4.3 Recherche d'un biais enquêteur

L'enquête filtre utilisée (MFV), comportait deux vagues successives qui ont été utilisées toutes les deux pour le pilier 2. Les enquêteurs qui en étaient chargés étaient les 12 enquêteurs du réseau pérenne de l'Insee, appuyés par 9 enquêteurs « occasionnels ». Les enquêteurs expérimentés n'ont pas travaillé sur la deuxième vague de MFV, car ils réalisaient la collecte du pilier 2. Deux phénomènes sont apparus rapidement qu'il a fallu étudier : d'une part, on note une dispersion très importante entre les enquêteurs sur le rapport entre le nombre d'individus ramenés par l'enquête filtre et l'échantillon affecté (de 2 % à 18 %). D'autre part, entre les deux vagues, ce taux d'éligibilité varie également fortement, passant de 6 % à 12 %, alors que le tirage des deux vagues a été réalisé de manière à ce que chacune soit indépendante et donc représentative.

Afin d'analyser d'où provenaient ces disparités, le taux d'éligibilité a été décomposé en plusieurs composantes. Tout d'abord, le taux auquel on s'intéresse est le taux d'éligibilité à raison, c'est-à-dire le nombre d'individus ramenés par l'enquête filtre et confirmés par l'enquête entreprise comme étant dans le champ rapporté au nombre d'enquêtes à réaliser par l'enquêteur. Plus précisément, on ne va ici considérer hors champs que les questionnaires hors champ pour des raisons détectables lors de l'enquête filtre. Ainsi défini, on peut réécrire le taux d'éligibilité comme le produit :

$$T_{\text{éligibilité}} = \frac{nbEligibles\grave{a}Raison}{nbEligibles} \times \frac{nbEligibles}{nbPersonnes} \times \frac{nbPersonnes}{nbRepondants} \times \frac{nbRepondants}{nbQuestionnaires}$$

On voit ainsi apparaître :

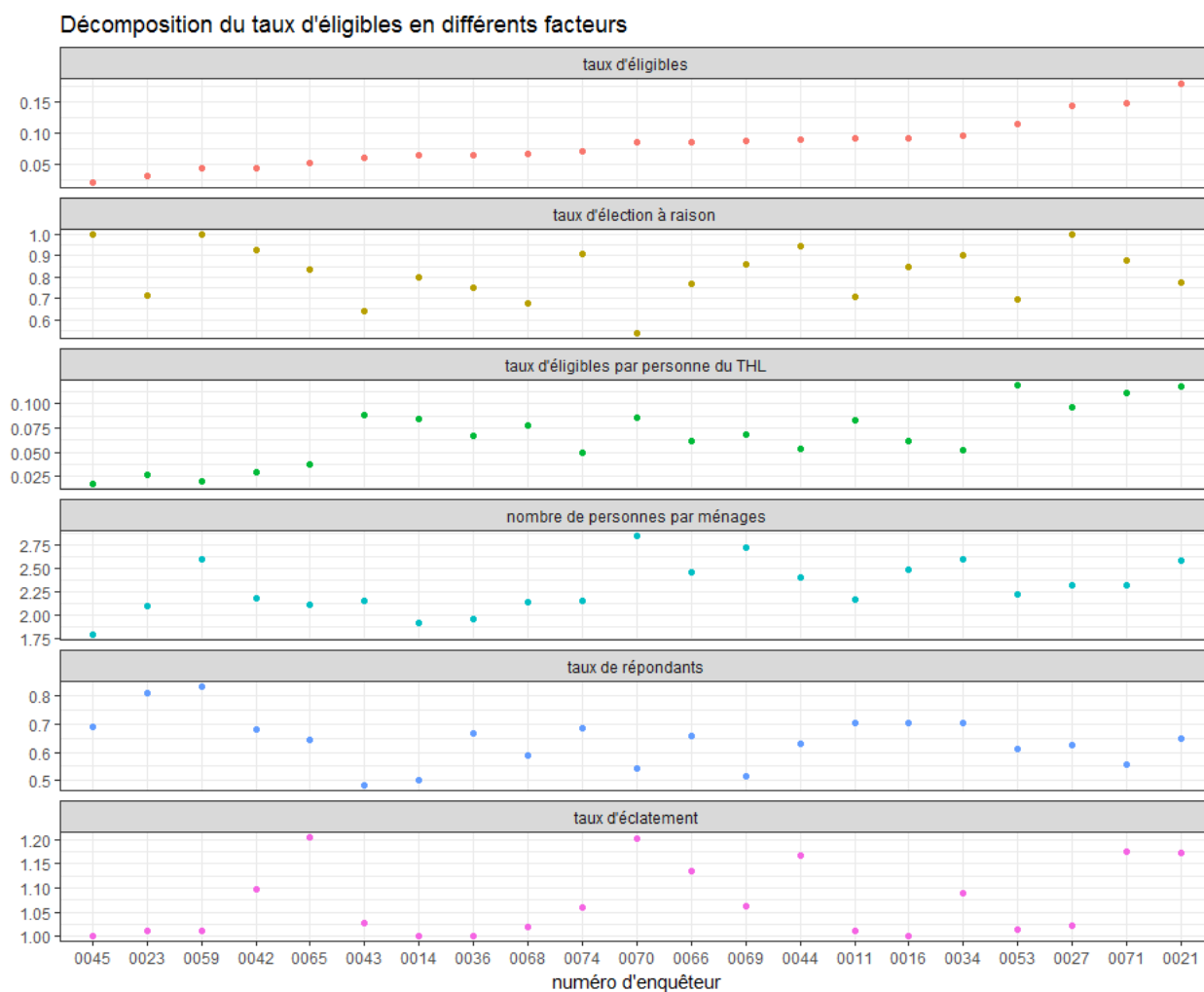


FIGURE 2 – Décomposition du taux d'éligibles en différents facteurs

- un taux d'élection à raison qui correspond au pourcentage d'entreprises identifiées par les enquêteurs qui sont effectivement des entreprises, et non des salariés ou des chômeurs : un taux bas signifie que l'enquêteur a signalé de nombreuses entreprises à tort ;
- un taux d'éligibles par membre du ménage qui mesure la part de personnes déclarées comme éligibles parmi l'ensemble des personnes qui auraient pu l'être ;
- le nombre de personne par ménage, très dépendant de la zone de l'enquêteur ;
- et enfin le taux de réponse des ménages à l'enquête.

Ces taux ont été calculés pour l'ensemble des enquêteurs (voir figure 2). Il ressort que le facteur le plus déterminant est le taux d'éligibles par personne du ménage. Ces disparités peuvent s'expliquer soit par un « effet enquêteur » qu'il faudrait alors corriger, soit par un « effet zone d'enquête » qui refléterait alors simplement la disparité de la localisation des chefs d'entreprises informelles. Les sources sont peu nombreuses sur le sujet à Mayotte. Néanmoins, en comparant des cartes issues du recensement de la population et de l'enquête emploi avec celles de ce taux d'éligibilité par personne (figure 3, à analyser avec prudence car les concepts ne sont pas parfaitement identiques entre ces trois sources), on observe une disparité géographique semblable qui plaide donc plutôt en faveur d'un effet zone d'enquête.

Il apparaît qu'en dehors d'un enquêteur qui avait par ailleurs donné des signes de fragilité, aucune aberration supplémentaire ne ressort. Aussi, aucun traitement n'a-t-il été entrepris pour

Taux d'éligibles par communes

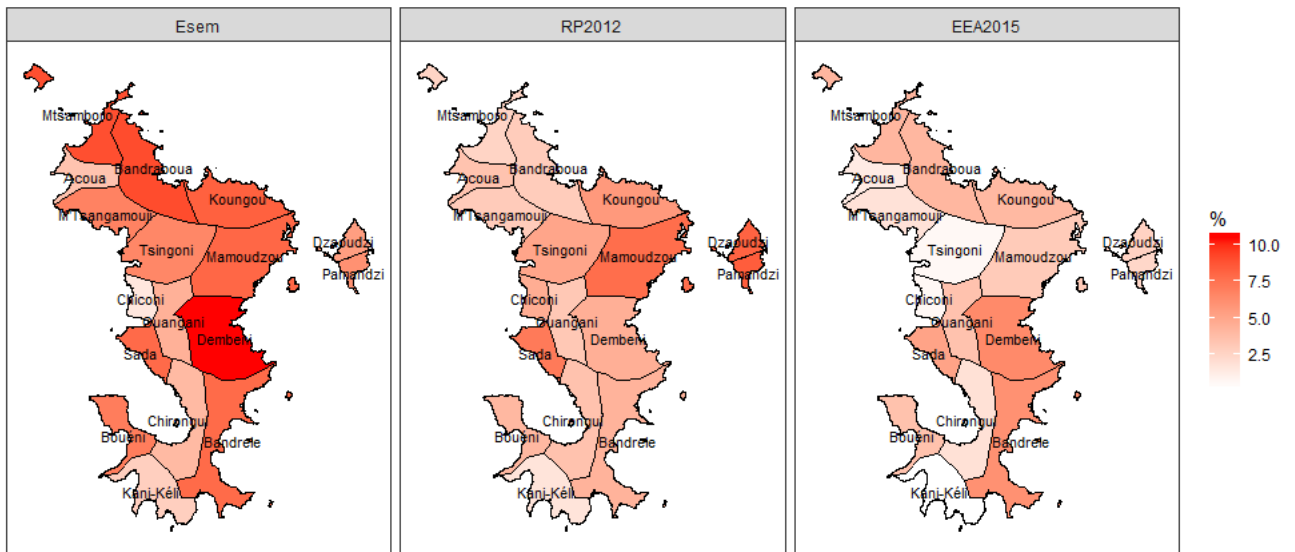


FIGURE 3 – Taux d'éligibles par commune

corriger la disparité inter-enquêteurs.

La comparaison entre les deux vagues fait par contre apparaître que, lors de la vague 2, le taux d'élection à raison est bien plus élevé que lors de la vague 1, générant un effet inverse à celui constaté sur le taux final. Pour la plupart des enquêteurs, ce taux est même de 100 % ce qui signifie qu'aucun éligible détecté n'a ensuite été mis hors champ lors de l'enquête entreprise. Et c'est encore une fois le taux d'éligibles par personne du ménage qui explique la baisse importante du taux résultant. Après des entretiens avec les enquêteurs et les gestionnaires de l'enquête, l'hypothèse retenue est que, entre les deux vagues, les enquêteurs ont changé de comportement sur la question filtre. Alors qu'ils avaient été plutôt lâches lors de la première vague, plaçant dans les éligibles beaucoup de personnes, certaines à tort, ils ont, lors de la seconde, été beaucoup plus strictes. Ils ont donc sélectionné moins d'éligibles à tort, mais ont probablement oubliés certains « vrais » éligibles. Cette hypothèse sera utilisée lors de la repondération pour redresser ce manque.

4.4 Expertise des données

L'expertise des données est effectuée en deux temps : tout d'abord, juste après la saisie, des contrôles de cohérences sont effectués pour repérer les erreurs de saisie ou de codification. Comme l'ensemble des questionnaires avaient préalablement été relus avec soin, cette phase n'a pas identifié d'autres problèmes que des fautes de frappes.

Dans un deuxième temps, et en l'absence de données auxquelles comparer le contenu des questionnaires, l'expertise a porté sur deux variables : le coefficient d'annualisation dont on a parlé

auparavant, et le taux de valeur ajoutée².

Pour le coefficient d'annualisation, les valeurs très en dessous de 12 ne sont pas alarmantes : la plupart du temps, c'est simplement le signe d'une interruption de l'activité pendant l'année, elles sont donc plausibles. Par contre, les valeurs très élevées qui correspondent aux cas de figure où le mois d'enquête est un mois jugé de faible activité par l'enquêté (supérieures à 24) sont à contrôler. Elles peuvent en effet donner un poids très important à une entreprise (le coefficient monte jusqu'à 400). Si la conjonction d'un mois enquêté labellisé mauvais et d'un rapport important entre le mois minimal et le mois maximal peuvent expliquer ça, il faut néanmoins prendre des précautions. Les 12 questionnaires concernés ont donc été étudiés afin de vérifier qu'ils sont plausibles, et corrigés dans le cas contraire. Dans deux cas, les données ne sont pas crédibles, et aucune correction n'a pu être envisagée. Comme en plus l'enquêteur a commenté en disant qu'il n'a pas confiance dans les réponses, ces données n'ont pas été utilisées.

Pour le taux de valeur ajoutée, les valeurs problématiques sont celles qui sont négatives (ce qui signifie que, avant de payer les travailleurs, l'entreprise travaille à perte) et les valeurs très proches de 1 (ce qui signifie que le chiffre d'affaires est égal à la valeur ajoutée, et donc que l'entreprise n'a pas de charges d'aucune sorte). Pour les cas concernés là encore, le questionnaire est décortiqué et corrigé si les données ne semblent pas plausibles. Par exemple, dans le cas d'un taxi fortement déficitaire, il s'avère que ses comptes sont fortement impactés par une dépense exceptionnelle de réparation déclarée le mois de la collecte. Si on fait l'hypothèse que cette réparation a certes été faite au cours du mois passé, mais ne doit pas être reproduite chaque mois, les comptes retrouvent un profil proche des autres taxis : la modification est donc appliquée.

4.5 Pondération, traitement de la non-réponse totale

Les individus interrogés dans l'enquête entreprise sont tous les individus détectés lors de l'enquête MFV, il n'y a pas de tirage réalisé. C'est donc la pondération finale de l'enquête MFV qui fait office de poids de tirage pour les individus de l'enquête entreprise.

Les individus définis comme non-retours sont ceux que l'on n'a pas retrouvé sur le terrain ou qui étaient absents au moment de l'enquête ainsi que ceux qui ont refusé l'enquête. Les individus hors-champ sont soit des salariés ou des chômeurs qu'on interroge suite à une erreur du filtre, soit des professions agricoles hors du champ de notre enquête. Les entreprises enquêtées qui ont déclaré effectuer une comptabilité ont été traitées tout à fait à part. Elles sont donc classées hors champ dans la suite des traitements et ont été rajoutées à la toute fin. Au final, on a 50 % de répondants, 7 % de non-retours et 43 % de hors-champ.

L'obtention de la pondération finale se fait en trois étapes : on détermine d'abord, parmi les non-retours, leur présence ou non dans le champ de l'enquête. Ensuite, on traite la non-réponse par repondération. Enfin, on corrige le problème d'omission d'individus éligibles dans la seconde vague. Aucun calage n'était en effet possible car aucune marge crédible n'était disponible sur cette population.

2. Le taux de valeur ajoutée est le rapport entre la valeur ajoutée et le chiffre d'affaires de l'entreprise.

Méthodologie de l'enquête Migration, famille et vieillissement

L'enquête MFV est la deuxième enquête auprès de ménages de grande ampleur conduite à Mayotte. 4 600 ménages ont été sélectionnés suivant un plan de sondage stratifié et équilibré. La base de sondage était constituée des adresses repérées dans les îlots cartographiés en 2014 pour la vague 1 et en 2015 pour la vague 2.

Dans chaque logement, l'ensemble des habitants étaient recensés puis une personne à enquêter était sélectionnée au hasard pour répondre à l'ensemble du questionnaire.

Les traitements post-collectes ont consisté en deux traitements successifs de la non-réponse, puis en un calage sur marge. Les deux traitements de la non-réponse, le premier au niveau logement et le second au niveau individu, ont été réalisés avec la méthode des groupes homogènes de réponses en constituant ces groupes avec une régression logistique.

4.5.1 Présence dans le champ des non-retours

Pour déterminer la présence dans le champ des non-retours³, on effectue une régression logistique sur les individus répondants et sur les individus hors champ à partir des variables de l'échantillon, c'est-à-dire des variables de l'enquête MFV, et on applique le modèle résultant aux non-retours.

Le modèle s'écrit : $\log\left(\frac{P_k}{1-P_k}\right) = z'_k\beta$ où P_k désigne la probabilité d'être dans le champ et les z_k sont les variables auxiliaires. Dans notre cas, il s'agit des variables du tronc commun des ménages. Après une sélection des variables les plus représentatives, il reste dans le modèle 4 variables : le type d'enquêteur (occasionnel ou permanent), le niveau d'étude, le secteur de profession et la localisation géographique.

Ainsi, on obtient une probabilité estimée d'être dans le champ de l'enquête pour les 33 non-répondants. Cette probabilité estimée va de 0,09 à 0,95 avec une moyenne de 0,53. La méthode d'imputation choisie est de conserver dans le champ de l'enquête les répondants qui ont une probabilité supérieure à 50 %, les autres étant hors champ. 17 questionnaires sont donc classés en non-répondants, auxquels on ajoute les deux questionnaires qui ont été jugés inexploitables lors de l'expertise. La suppression de 16 non-retours du champ de l'enquête revient à réduire la somme des poids de 330, ce qui réduit d'autant le nombre d'entreprises informelles par rapport à une situation où on considérerait tous les non-retours comme des non-répondants.

4.5.2 Traitement de la non-réponse totale

Lors du traitement de la non-réponse totale, on cherche à reporter le poids des non-répondants sur les répondants. La non-réponse est traitée par la méthode des groupes de réponse homogènes : on modélise la probabilité de répondre quand on est dans le champ pour l'ensemble des individus, de cette modélisation on constitue des groupes aussi homogènes que possible, et, à l'intérieur d'un groupe, on répartit le poids des non-répondants sur les répondants. Dans la modélisation de la probabilité de réponse, on prend garde de ne sélectionner que les variables corrélées avec les variables d'intérêt afin de ne pas induire de variance qui ne réduise pas de biais.

On a choisi comme variable d'intérêt le chiffre d'affaires, la première étape est donc de tester la corrélation au chiffre d'affaires des variables potentiellement explicatives, qui sont là encore les

3. On distingue les non-retours des non-répondants : les non-retours sont des individus qui n'ont pas répondu à l'enquête, mais pour lesquels on ne sait pas s'ils auraient été dans le champ, alors que les non-répondants sont les individus dont on considère qu'ils sont dans le champ de l'enquête.

variables du tronc commun des ménages. Cinq variables sont corrélées avec la variable d'intérêt : le lieu de naissance, la nationalité, le secteur de profession, le nombre de pièce du logement et enfin le statut d'occupation du logement. De ces cinq variables une seule sera retenue lors de la modélisation, le statut d'occupation du logement.

Cette variable a trois modalités, les propriétaires du sol et/ou de la maison, les locataires, et les logés à titre gratuit. Chaque modalité va donc correspondre à un groupe homogène de réponses. Au sein de chacun de ces groupes, on répartit le poids des non-répondants sur les répondants de manière égale.

Le traitement de la non-réponse totale fait passer le total du chiffre d'affaires de 100,5 millions d'euros à 107,8 millions d'euros.

4.5.3 Traitement de l'éligibilité

Comme vu plus haut dans le paragraphe sur la recherche d'un biais enquêteur, le taux d'éligibles est beaucoup plus faible lors de la vague 2 de MFV que lors de la vague 1 (1,2 % contre 2,3 %). Après analyse, l'hypothèse faite est qu'une auto-censure des enquêteurs lors de la vague 2 a conduit à un défaut de couverture, autrement dit qu'il manque des éligibles dans la vague 2, ce qui conduit à une sous-évaluation du nombre d'entreprises et des soldes comptables. Nous allons ici corriger cela, d'abord en mesurant la somme des poids manquante, puis en la répartissant sur les répondants.

Pour cela, on fait l'hypothèse que le taux d'éligibilité lors des deux vagues aurait du être identique (les vagues sont peu éloignées dans le temps, aucun phénomène particulier n'a été identifié qui pourrait impacter ce taux, les deux tirages sont équilibrés de la même manière de façon à ce que chaque vague soit également représentative du territoire), et que le taux mesuré lors de la vague 1 est fiable.

Plus précisément, c'est le taux d'éligibles répondants qui est corrigé et non le taux d'éligibles brut. En effet, s'il manque des éligibles dans la vague 2, la collecte a montré que, dans la vague 1, il y en a trop : il s'agit des éligibles dont on s'aperçoit, quand on les enquête, qu'ils ne sont en fait pas à leur compte. Sachant que la somme des poids après correction de la non-réponse est de 4 755 avec un taux d'éligibles global de 2,1 %, on obtient la somme des poids souhaitable par un produit en croix :

$$\sum W_{\text{aprèsCorrection}} = \sum W_{\text{avantCorrection}} \cdot \frac{\text{tauxEligiblesAprèsCorrection}}{\text{tauxEligiblesAvantCorrection}} = 5182$$

Ceci correspond à une augmentation de la somme des poids de 428, qu'il va falloir répartir sur les 238 répondants. En effet, il n'est pas opportun de répartir cette somme uniquement sur la vague 2 car cela reviendrait à sur-pondérer des individus uniquement du fait de leur appartenance à une vague. Cela induirait une distorsion des poids importantes et donc une variabilité plus forte des indicateurs. On cherche plutôt à répartir ces poids sur l'ensemble des individus qui sont proches de ceux qui manquent.

À cette fin, on recherche à l'aide d'une régression logistique quelles sont les variables du tronc commun qui, croisées avec la vague, expliquent l'éligibilité. En effet, on veut déterminer les caractéristiques des individus manquants de la vague 2. Quatre variables issues de MFV ressortent, par ordre d'importance :

- la seconde condition du filtre d'éligibilité croisée avec la vague (« Avez-vous, dans les 6 derniers mois, occupé un emploi dans lequel vous étiez à votre compte ? ») ;
- la première condition du filtre d'éligibilité croisée avec la vague (« Occupez-vous actuellement un emploi à votre compte ? ») ;

		Marge	Effectif dans l'échantillon	Variation
Filtre 1	Oui	2 968	2 784	6,2 %
	Non	2 214	1 971	11,0 %
Filtre 2	Oui	2 436	2 150	11,7 %
	Non	2 746	2 604	5,2 %
Profession	Commerce	1 672	1 631	2,4 %
	Autre	281	227	19,2 %
	Sans profession	3 292	3 123	5,1 %
Nationalité	Français	1 890	1 632	13,6 %
	Étrangers	3 292	3 123	5,1 %
Poids		5 182	4 755	8,2 %

FIGURE 4 – Évolution des poids avec le calage sur marge

- le secteur de la profession croisé avec la vague (commerce/autre/sans profession) ;
- la nationalité croisée avec la vague (français/étrangers).

Ce sont donc ces variables sur lesquelles le fait d'avoir omis des répondants en vague 2 a le plus joué. On va donc utiliser la répartition de ces variables sur la vague 1 de l'échantillon, et faire en sorte de corriger les poids pour qu'on la retrouve sur l'ensemble de l'échantillon par un système de calage sur marge (voir figure 4). Par exemple, on a sur la vague une 36,5 % de français et 63,5 % d'étrangers, nous allons donc chercher, au total, à obtenir $0,365 \times 5\,182 = 1\,890$ français sur l'ensemble de la population. L'exercice est répété pour chacune des variables, et ces marges vont alors être utilisées dans un calage sur marge, ainsi que la marge en volume sur la somme des poids.

4.5.4 Traitement des entreprises ayant déclarées tenir une comptabilité

Pour les entreprises qui déclarent tenir une comptabilité, le questionnaire du pilier 2 ne leur est pas passé. Elles sont appariées avec la base d'entreprises du pilier 1. Quand elles y sont aussi présentes, ce sont en fait des entreprises formelles, elles sont déclarées hors champ du pilier 2. Dans le cas contraire, elles sont enquêtées avec le questionnaire du pilier 1, mais sont conservées dans le champ du pilier 2. Ces entreprises sont très peu nombreuses (15, dont 4 répondantes). Elles sont traitées à part car atypiques par rapport aux autres entreprises du pilier 2 : elles possèdent en effet un chiffre d'affaires quatre fois plus important que celles ayant déclarées ne pas faire de comptabilité. Parmi les 11 non répondantes, 3 sont hors champ, 1 est non-répondante et le reste sont des non-retours. Comme pour la majorité des questionnaires, la première étape est de déterminer quels sont les non-retours qui sont dans le champ. Sur si peu d'unités, il n'est pas possible de réaliser une régression logistique, aussi on applique le modèle précédent à ces unités pour déterminer leur probabilité d'appartenance au champ. En ne conservant que les entreprises qui ont une probabilité d'être dans le champ supérieure à 50 %, 2 entreprises supplémentaires sont non-répondantes, le reste hors champ. Pour répartir le poids des 3 non-répondants sur les 4 répondants, là aussi le peu d'unités concernées interdit tout traitement un peu fin, aussi a-t-il été décidé de répartir les poids de manière à ce que la part des poids de chaque répondant soit constante :

$$W_{\text{aprèsTraitement}} = W_{\text{initial}} \cdot \frac{\sum W_{\text{aprèsTraitement}}}{\sum W_{\text{initiaux}}}$$

5 Enseignements

Le dispositif présenté ici a abouti en février 2018 sous la forme de deux publications présentant des statistiques structurelles sur les entreprises formelles et informelles à Mayotte [3][4]. Le message principal qu'il en ressort est qu'à Mayotte, les deux tiers des entreprises sont informelles alors qu'elles ne génèrent que 9 % de la valeur ajoutée. C'est la première fois qu'une estimation de la part de l'informel est réalisée, elle a donc permis d'apporter un cadre à un débat dans lequel des avis peu argumentés s'affrontaient jusqu'alors.

Comme les entreprises informelles ont finalement une contribution assez faible à l'économie et que l'opération a été relativement coûteuse, il n'a pas été jugé pertinent de la reconduire tous les ans. Par contre, le pilier 1 sur les entreprises formelles est entré lui dans un rythme annuel afin de pouvoir suivre précisément le développement de l'économie mahoraise. Pour le pilier 2, une réédition à l'échéance 2022 est envisagée sans que rien ne soit pour le moment décidé. D'ici là, une méthode pour faire vieillir la participation des entreprises informelles devra être mise au point.

L'expérience menée sur l'exercice 2015 est donc positive, la prochaine édition gagnera à s'appuyer sur l'expérience acquise notamment sur les points suivants :

- il est préférable de choisir une enquête filtre de type enquête emploi, afin de limiter le nombre de hors champ. Comme l'emploi n'était pas au cœur de la problématique de MFV, le filtre n'a pas permis de récupérer le secteur de manière satisfaisante, et il n'a donc pas été possible de trier en amont (en enlevant les agriculteurs notamment) ;
- l'importance du module rythme, à la base de l'annualisation, a sans doute été sous-estimé, avec pour conséquence une qualité parfois inégale selon les enquêteurs. L'enquête aurait sans doute gagné à ce que la formation insiste davantage dessus. Ce module pourrait par ailleurs être complété en y ajoutant le revenu moyen, en plus des revenus minimaux et maximaux ;
- le questionnaire devrait être toiletté : certaines parties n'ont en effet jamais été utilisées, notamment le module équipement et le détail des primes. Cela permettrait d'alléger un questionnaire souvent perçu comme très long pour les structures importantes ;
- l'organisation mise en place, avec une très forte présence des équipes en bureaux auprès des enquêteurs, était bien nécessaire. En effet, les enquêteurs ne sont pas du tout habitués à ce type d'enquête et il est essentiel de pouvoir repérer les malentendus au plus vite afin de les lever ;
- l'ensemble des informations, notamment celles sur la temporalité, devraient être saisies afin de simplifier l'expertise sans nécessiter de retourner au questionnaire papier.

Références

- [1] L'enquête 1-2-3 sur l'emploi et le secteur informel à yaoundé. *Stateco*, 78 (juin 1994).
- [2] Enquête nationale sur l'emploi et le secteur informel : méthode originale conçue pour appréhender l'économie informelle. *Rapport méthodologique de l'Instat Madagascar* (novembre 2013).
- [3] DAUBRÉE, S. Les entreprises formelles génèrent 525 millions d'euros. *Insee Flash Mayotte*, 62 (février 2018).
- [4] DAUBRÉE, S. Les entreprises informelles créent 9 % de la richesse. *Insee Analyses Mayotte*, 16 (février 2018).
- [5] QUENUM, S., MONTEIL, C., AND BRASSET, M. Les entreprises à mayotte : une population difficile à enquêter. *XIIIe Journées de Méthodologie Statistique de l'Insee* (juin 2015).